ABSTRACT

Title of dissertation: DECADAL TO CENTENNIAL SCALE CLIMATE DYNAMICS IN MODELS OF VARYING COMPLEXITY

Adria K. Schwarber, Doctor of Philosophy, 2019

Dissertation directed by: Professor Fernando Miralles-Wilhelm
Chair, AOSC
Interim Director, ESSIC

Though concerted climate action by the world's governments intends to limit long-term (e.g. 100 years) global average temperature rise, attention has recently focused on reducing climate impacts in our lifetime by reducing emissions of short-lived climate forcers (SLCFs). SLCFs are pollutants that remain in the atmosphere for a short time (e.g. methane or black carbon) and have the potential to impact the climate in the near-term by increasing or decreasing temperature, depending on the species emitted. There is a more limited set of literature, however, that robustly characterizes short-term climate dynamics in the 20-30 year time horizon within models or observations that can be used to inform scientific and policy work.

In this dissertation, we seek to clarify climate dynamics on shorter time scales using models of varying complexity—from complex models, which take several months to simulate 100 years of climate on a supercomputer, to simple climate models (SCMs) that can simulate the same period on a personal computer in less than a minute, in addition to using several observational datasets.

We first characterize the basic climate processes within several SCMs, finding that some comprehensive SCMs fail to capture response timescales of more complex models, for example under BC forcing perturbations. These results suggest where improvements should be made to SCMs, which affect numerous scientific endeavors and illustrates the necessity of integrating fundamental tests into SCM development.

We then robustly determine how realistic complex model variability is compared to observations across all time scales using power spectra of temperature-time series. We investigate model variability at the regional level, using the continental-scale regions defined by PAGES2k. We find that compared to observations the suite of CMIP5 models investigated have lower variability in certain regions (e.g. Antarctica) and higher variability in others (e.g., Australasia), with some consistency across timescales. Our approach allows for a more robust assessment of complex model variability at time periods and regional levels important to human systems.

From this, we analyze the range of temperature responses over time in complex model results from phase 5 of the Coupled Model Intercomparison Project (CMIP5) at the hemispheric scale to create a realistic range of possible temperature changes. We find that the range of responses of land/ocean varied less than the range of hemispheric responses. Our results are a first step of better quantifying the short-term climate responses to changes in SLCFs.

DECADAL TO CENTENNIAL SCALE CLIMATE DYNAMICS IN MODELS OF VARYING COMPLEXITY

by

Adria K. Schwarber

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Fernando Miralles-Wilhelm, Chair
Dr. Steven J. Smith, Co-Advisor
Dr. Corinne Hartin, Co-Advisor
Professor Ross Salawitch, Academic Advisor
Professor Akua Asa-Awuku, Dean's Representative

Dedication

I want to dedicate this dissertation to all the young girls and women aspiring to a career in science. Pursuing science is challenging in many ways. Though you will expand your understanding of the world around you, you will also face many adversities that are unfortunately ingrained in the STEM education system. Without the mentors I have had throughout my career, including my 6th-grade science teacher, Ms. Katy Drinkhouse, and my high-school science teachers, Sr. Mary-Ethel Parrott and Sr. Judith Averbeck, I would not have pursued a STEM degree.

In addition, without the strong support of my partner, Vikram–to whom I also dedicate this dissertation–I could not have persevered through many obstacles to complete this thesis.

I hope one day you will not face questions of the space you occupy or your abilities as a scientist. Until then, I leave you with this inspirational poem:

And Still I Rise

You may write me down in history

With your bitter, twisted lies,

You may trod me in the very dirt

But still, like dust, I'll rise.

Does my sassiness upset you?

Why are you beset with gloom?

'Cause I walk like I've got oil wells

Pumping in my living room.

Just like moons and like suns,

With the certainty of tides,

Just like hopes springing high,

Still I'll rise.

Did you want to see me broken?

Bowed head and lowered eyes?

Shoulders falling down like teardrops,

Weakened by my soulful cries?

Does my haughtiness offend you?

Don't you take it awful hard

'Cause I laugh like I've got gold mines

Diggin' in my own backyard.

You may shoot me with your words,

You may cut me with your eyes,

You may kill me with your hatefulness,

But still, like air, I'll rise.

Does my sexiness upset you?

Does it come as a surprise

That I dance like I've got diamonds

At the meeting of my thighs?

Out of the huts of history's shame

I rise

Up from a past that's rooted in pain

I rise

I'm a black ocean, leaping and wide,

Welling and swelling I bear in the tide.

Leaving behind nights of terror and fear

I rise

Into a daybreak that's wondrously clear

I rise

Bringing the gifts that my ancestors gave,

I am the dream and the hope of the slave.

I rise

I rise

I rise.

Maya Angelou

Acknowledgments

This dissertation would not have been possible without the help, support, and encouragement of countless individuals. I cannot name every teacher, advisor, or friend in this section, but I do want to take time to expound on several that deserve all the praise possible for getting me across the finish line.

In particular, I would like to thank my advisors, Drs. Steven J. Smith and Corinne Hartin. Without the generosity of their time and effort, I would never have completed this work. Working with both of them, and all the staff and scientists at the Joint Global Change Research Institute, was the greatest of opportunities and has taught me more about myself and my goals than anyone may have realized. Steve and Corinne served not only as my primary research advisors, but were always available for career advice and the occasional work-life balance discussion. The funding agencies that supported my work while at JGCRI also deserve recognition including the U.S. Department of Energy Office of Science as part of research in the Multi-Sector Dynamics, Earth and Environmental Systems Modeling Program, the US Environmental Protection Agency, and the Battelle Memorial Institute.

I would also like to thank the other members of my dissertation committee. Professor Fernando Mirrelles-Wilhelm agreed to serve as the Chair of my dissertation committee and very generously provided funding for my final full semester of research. Professor Ross Salawitch offered great insight on my work and helped me hold my written work to a high standard. I would also like to thank Professor Asa-Awuku for agreeing to serve as the Dean's Representative. Finally, though

Professor Elisabeth Gilmore was not able to be present at my defense, she supported my Master's degree work and provided insights into the policy implications of my dissertation during its formulation.

I also want to thank the entire Atmospheric and Oceanic Science Department at the University of Maryland for the support provided over my long Ph.D. career. I have felt cared for from the moment I joined the program, including through several generous acknowledgements of my departmental service and the flexibility of the staff at every step along my Ph.D. path. Because of an early conversation with Tammy Hendershot on my interests in science policy and her dedication, I was connected with the Joint Global Change Research Institute where I was able to complete my work. I also want to thank June Sherer and Bernadette Gatewood who both helped me understand financial and benefit information and made it very easy to do my research without worry. Every student in the Atmospheric and Oceanic Science Department knows they are very lucky to have such excellent and dedicated staff.

I would like to thank the Graduate School at the University of Maryland for supporting my work through the Ann Wiley Dissertation Fellowship. In particular, I would like to thank Assistant Dean of the College of Behavioral and Social Sciences, Jeffrey Franke, who previously served as the Associate Dean of the Graduate School. Not only did Jeff support the Graduate Student Government, where I found a home for 5 years, but he also took the time to personally support my own career efforts. Thank you for all that you do, Jeff!

On a personal note, I want to thank my family. My parents, Bill and Louise

Schwarber, never gave up on me and always pushed me to pursue my dreams, whether that be digging in the dirt on the baseball field or collecting dust for my 6th-grade science fair project. I have cultivated a high level of perseverance because of them. Remember, as they always say, "This too shall pass."

Finally, and certainly not least, I owe gratitude to my partner, Vikram, for pushing me to finish my Ph.D., despite my own self doubt. Thank you for keeping me honest and driving me to fight for myself. I could not have done this without you.

To everyone I did not explicitly name above, including my friends and family in Kentucky, and my friends and peers in Maryland, please know that I would have been lost my without your support and friendship. Thank you!

Previously Published Work

Some of the research leading to this thesis has appeared previously in the following publications.

Parts of Chapter 2 have appeared in:

Schwarber, A. K., Smith, S. J., Hartin, C. A., Vega-Westhoff, B. A., and Sriver, R.: Evaluating climate emulation: fundamental impulse testing of simple climate models, Earth Syst. Dynam., 10, 729–739, https://doi.org/10.5194/esd-10-729-2019, 2019.

Contents

Dedication . . . ii

Acknowledgements . . . v

Previously Published Work . . . viii

List of Abbreviations . . . xv

1. Introduction . . . 1
1.1 Overview of climate change . . . 1
1.1.1 Radiative forcing concept . . . 2
1.1.2 Temperature response . . . 5
1.2 Climate models of varying complexity . . . 8
1.2.1 Energy balance models . . . 8
1.2.2 Simple climate models . . . 8
1.2.3 Earth System Models of Intermediate Complexity . . . 11
1.2.4 Complex Models: Coupled Atmosphere-Ocean General Circulation Models and Earth System Models . . . 11
1.3 Thesis objectives and outline . . . 13

2. Simple Climate Model Evaluation Using Fundamental Unit Tests (published as Schwarber et al., 2019) . . . 16
2.1 Introduction . . . 16
2.2 Methods . . . 19
2.2.1 Fundamental impulse tests. . . . 19
2.2.2 Background concentrations. . . . 22
2.2.3 Models. . . . 23
2.2.4 Parameter selection. . . . 25
2.3 Results . . . 27
2.3.1 Responses to CO_2 Concentration Impulse. . . . 27
2.3.2 Responses to Emissions Impulses. . . . 28
2.3.3 Responses to 4xCO_2 Concentration Step. . . . 38
2.4 Conclusions . . . 40

3. Observations . . . 45
3.1 Introduction . . . 45

3.2 Datasets . . . 46
3.2.1 Historical datasets . . . 47
3.2.2 PAGES2k Paleoclimate Reconstruction . . . 53
3.3 Processing Choices . . . 54
3.3.1 Impact of regridding method . . . 58
3.4 Spectral Density Estimates Using Base Method . . . 59
3.4.1 Region size sensitivity tests . . . 61
3.5 Conclusions . . . 64

4. Characterization of Unforced Model Variability in CMIP5 (in prep for submission to Journal of Geophysical Research - Atmospheres) . . . 66
4.1 Introduction . . . 66
4.2 Data and Methods . . . 70
4.3 Results and Discussion . . . 77
4.3.1 Control Run Variability . . . 77
4.3.2 Model Noise and Forced Variability . . . 80
4.3.3 Model-to-Observations Comparison . . . 86
4.4 Regional Analysis . . . 86
4.4.1 Australasia . . . 89
4.4.2 Antarctica . . . 93
4.4.3 Arctic . . . 96
4.4.4 Asia . . . 100
4.4.5 Europe . . . 102
4.4.6 North America . . . 104
4.4.7 South America . . . 106
4.5 Conclusions . . . 107

5. Climate Response Over Time (in prep for submission to Journal of Climate) 110
5.1 Motivation . . . 110
5.2 Background . . . 111
5.3 Methods . . . 112
5.4 Preliminary Results . . . 114
5.5 Conclusions and Work Plans . . . 126

6. Concluding Remarks and Future Work . . . 127
6.1 Concluding Remarks . . . 127
6.1.1 Fundamental Impulse response tests in SCMs . . . 127
6.1.2 Characterizing unforced variability in CMIP5 models . . . 128
6.1.3 Changes in Climate Response Over Time . . . 129
6.2 Future Work . . . 130
6.2.1 CMIP5 Emulation in MAGICC 5.3 . . . 132
6.2.2 Understanding Climate Response to SLCFs . . . 135

Appendix 137

7. Appendix 138
7.1 Additional Results for Chapter 4 138
7.1.1 Additional CMIP5 Data Description 138
7.1.2 Control Run Variability 138
7.2 Additional Results for Chapter 5 138
7.2.1 Appendix B: Correlation Analysis 141
7.2.2 Appendix C: Decomposing the Spatial and Temporal Responses 144

List of Tables

2.1 Main carbon cycle and climate characteristics of SCMs and IRFs . . . 26
2.2 Time-integrated Airborne Fraction from a 100 GtC CO_2 Emissions Impulse in SCMs Compared to Results from Table 4 in Joos et al. (2013) . . . 33
2.3 Time-integrated temperature response from a 100 GtC CO_2 Emissions Impulse in SCMs Compared to Results from Table 7 in Joos et al. (2013) . . . 35
2.4 Integrated temperature response differences . . . 44

3.1 Summary of Global Surface Temperature Datasets . . . 55
3.2 PAGES2k region definitions . . . 57
3.3 Region Definitions for Sensitivity Analysis . . . 63

4.1 CMIP5 Data and Experiment Description . . . 71

7.1 Additional CMIP5 Past1000 Data Description . . . 139
7.2 Robustness of Average Spectral Density and Halfwidths for Each Model, Region, and Period Bin at L=500 . . . 140

List of Figures

1.1 Schematic of Aerosol Forcing Mechanisms 4
1.2 Global Temperature Projections Updated from IPCC AR5 Fig. 11.25 6
1.3 Diagram of climate models of varying complexity 9

2.1 Global mean temperature response (a) and integrated global mean temperature response (b) from a CO_2 concentration perturbation in SCMs 29
2.2 Total forcing response from CO_2 (dashed) and CH_4 (solid) emissions perturbations in SCMs 30
2.3 Time-integrated airborne fraction from a 100GtC CO_2 emissions impulse in SCMs compared to Joos et al. (2013) 34
2.4 Time-integrated temperature response from a 100GtC CO_2 emissions impulse in SCMs compared to Joos et al. 36
2.5 Global mean temperature response from CO_2 and CH_4 emissions perturbations (a) and BC emissions perturbation (b) in SCMs 37
2.6 Global mean temperature response from a 4xBC emissions step in the SCMs 39
2.7 Global mean temperature response from 4xCO_2 concentration step in CMIP5 models (grey) and SCMs 40

3.1 Annual mean temperature change between 1901-1920 and 1991-2010 from HadCRUT 48
3.2 Annual mean temperature change between 1901-1920 and 1991-2010 from NOAA MLOST 50
3.3 Annual mean land temperature change between 1901-1920 and 1991-2010 from GISTEMP 51
3.4 Annual mean land temperature change between 1901-1920 and 1991-2010 from Berkeley Earth 52
3.5 Graphical representation of the PAGES2k regions with land-ocean masking based on the continental scale regions defined by PAGES2 . 57
3.6 Comparison of temperature-time series produced using various regridding techniques for each in-situ dataset over Australasia 60
3.7 Spectral density estimates of observational data over the PAGES2k regions 62
3.8 Exploring the impact of region size on spectral density 65

4.1 Schematic of multitaper method . . . 73
4.2 Spectral Density of Models Across all Regions . . . 75
4.3 Schematic of Method to Find Normalized Spectral Difference . . . 76
4.4 Bootstrap Estimation of PiControl Internal Variability at L=150 years 79
4.5 Bootstrap Estimation of PiControl Internal Variability at L=500 years 81
4.6 Normalized Model-to-Model Control Run Spectral Difference With Values . . . 83
4.7 Normalized Model-to-Model Control Run Spectral Difference . . . 84
4.8 Ratio of Short-term to Long-term Normalized Average Spectral Difference Between Models . . . 85
4.9 Normalized Model-to-Observation Spectral Difference With Values . . 87
4.10 Normalized Model-to-Observation Spectral Difference . . . 88

5.1 Ratio of land to ocean warming produced from CMIP5 1pctCO_2 RT_t series . . . 115
5.2 Ratio of NH to SH warming produced from CMIP5 1pctCO_2 RT_t series 116
5.3 CMIP5 model comparison of the regional (NH/SH, land/ocean) RT_t ratios in order of increasing transient climate response values (TCR in °C). . . . 117
5.4 Mean and range of the hemispheric (NH/SH, land/ocean) RT_t ratios 118
5.5 Ratio of Arctic hemispheric-level warming produced from CMIP5 1pctCO_2 RT_t series . . . 119
5.6 Ratio of Antarctic hemispheric-level warming produced from CMIP5 1pctCO_2 RT_t series . . . 120
5.7 Signal to noise analysis, with the red dashed line showing the ToE for each model . . . 122
5.8 Time of Emergence (in years) values plotted for the global-level RT_t series from each of the CMIP5 model used in this study . . . 123
5.9 IPCC Transient Climate Response values compared to the $RT_t(70)$ multiplied by $t = 70$ for each CMIP5 model . . . 124
5.10 Global ratio of near-term, $\overline{RT_t(20)}$, to longer-term, $\overline{RT_t(70)}$, response for each CMIP5 model . . . 125

6.1 Response ration for each CMIP5 model and MAGICC5.3 for different climate sensitivities . . . 134

7.1 ToE values for each model plotted against the half-width at year 20 . 141
7.2 $\overline{RT_t(20)}/\overline{RT_t(70)}$ values for each model plotted against the hemispheric values of ϕ (left) and NH/SH (right) . . . 142
7.3 $\overline{RT_t(20)}$ (top) and $\overline{RT_t(70)}$ (bottom) values for each model plotted against the Time of Emergence . . . 143
7.4 Northern Hemispheric ratio of near-term, $\overline{RT_t(20)}$, to longer-term, $\overline{RT_t(70)}$, response for each CMIP5 model . . . 144
7.5 Southern Hemispheric ratio of near-term, $\overline{RT_t(20)}$, to longer-term, $\overline{RT_t(70)}$, response for each CMIP5 model . . . 146

7.6 Summary of the ratios of near-term, $\overline{RT_t(20)}$, to longer-term, $\overline{RT_t(70)}$, response for each CMIP5 model for each hemispheric level 147

List of Abbreviations

1pctCO_2	stylized CMIP5 experiment whereby CO_2 concentration is increase 1% yr^{-1} until doubling
4xCO_2	stylized CMIP5 experiment whereby CO_2 concentration is instantaneously quadrupled
AOGCM	Coupled Atmosphere-Ocean General Circulation Models
BC	Black Carbon
CO_2	Carbon Dioxide
CH_4	Methane
CS	Climate Sensitivity
CMIP5	Phase 5 of the Coupled Model Intercomparison Project
EBM	Energy Balance Model
ECS	Equilibrium Climate Sensitivity
EMICs	Earth System Models of Intermediate Complexity
ESMs	Earth System Models
FAIR	Finite Amplitude Impulse Response Model
GHG	Greenhouse Gas
GMST	Global Mean Surface Temperature
GTP	Global Temperature Potential
GWP	Global Warming Potential
IAMs	Integrated Assessment Models
IPCC AR5	Intergovernmental Panel on Climate Change Fifth Assessment Report
MAGICC	Model for the Assessment of Greenhouse-gas Induced Climate Change
N_2O	Nitrous Oxide
NAS	National Academy of Sciences
NH	Northern Hemisphere
past1000	CMIP5 Paleoclimate Runs
PiControl	CMIP5 Pre-Industrial Control Runs
RCPs	Representative Concentration Pathways
RF	Radiative Forcing
RWF	Realized Warming Fraction
SCMs	Simple Climate Models
SH	Southern Hemisphere
SLCFs	Short-Lived Climate Forcers
TCR	Transient Climate Response
TOA	Top of the atmosphere
ToE	Time of Emergence
WMGHG	Well-mixed greenhouse gas

Chapter 1: Introduction

1.1 Overview of climate change

Following the Industrial Revolution in 1750, rapid economic growth driven by rising population and standards of living continues to increase fossil fuel usage (Abram et al. 2016). Various industrialization processes, such as electricity generation, transportation, cement production, and land-use changes from agricultural activities, have increased the atmospheric burden of infrared active gases, called greenhouse gases (GHG; Hoffert et al. 2002).

Anthropogenic climate change is primarily attributable to increasing GHG concentrations since pre-industrial time, enhancing the atmospheric absorption of outgoing terrestrial radiation (Manabe and Strickler 1964). Of particular concern has been the increase in atmospheric carbon dioxide (CO_2) concentrations from approximately 280 parts per million (ppm) to currently over 412 ppm (Figure SPM 4 see in Stocker et al. 2013 and Keeling et al. 2005 as of September 2019), a 47% increase since pre-industrial. Further, CO_2 concentrations are expected to continue rising as economic growth, particularly of developing nations, increases energy demands and fossil fuel consumption (Stocker et al. 2013).

1.1.1 Radiative forcing concept

The modulation of climate is often characterized as a change in top-of-the-atmosphere (TOA) radiative forcing (RF), a measurement of the capacity of a gas or other forcing agent to affect Earth's energy balance. The RF of a gas is defined as the difference between incoming solar radiation and outgoing radiation caused by the increased concentration of that gas, expressed in Wm^{-2} (Stocker et al. 2013). For CO_2, a well-mixed GHG, the relationship between concentration and RF follows a logarithmic relationship (see Equation 1.1) given by:

$$\Delta RF_{CO_2} = 5.35x \ln \frac{C}{C_0} \tag{1.1}$$

where C is the CO_2 concentration in parts per million by volume (ppmv) and C_0 is the reference concentration.

The relationship derives from the absorption spectrum of CO_2, where the major absorption bands decay exponentially with the distance to band center (Pierrehumbert 2011). The spectrally averaged absorptivity, therefore, grows logarithmically with the absorber amount (Huang and Bani Shahabadi 2014). We exploit this relationship in Chapter 5 of this dissertation.

Temporal changes in climate response to perturbations vary depending on the species emitted and are important for understanding near-term climate change addressed in this study (Pierrehumbert 2014). Global attention has been focused on long-term climate change responses from increased CO_2 concentrations, the most abundant anthropogenic GHG. Of increasing importance are near-term climate change responses and the components that drive this change (Anenberg et al. 2012;

Shindell et al. 2012).

Short-lived climate forcers (SLCFs) constitute a class of atmospheric species with lifetimes shorter than CO_2 and an outsized near-term climate impact, including both methane (CH_4) and aerosols, such as black carbon (BC). Methane, a well-mixed GHG (WMGHG), has a relatively short atmospheric lifetime compared to CO_2 and is often included as a SLCF. In addition, CH_4 is a precursor to other SLCFs, such as tropospheric ozone (Lelieveld et al. 1993; Shine 2010). CH_4 accounts for 17% of the net global increase in RF (Rogelj et al. 2014) since 1750, and its individual RF is given by:

$$\mathrm{RF}_{CH_4} = 0.036\,(\sqrt{M} - \sqrt{M_0}) - (f(M,\ N_0) - f(M_0,\ N_0)) \tag{1.2}$$

where M is CH_4 in ppbv, N is nitrous oxide (N_2O) in ppbv, and f represents a function of the overlapping absorption bands with N_2O and water vapor (Huang and Bani Shahabadi 2014).

Methane is of increasing importance as both natural (melting of Arctic permafrost or emissions from tropical wetlands (e.g., Christensen et al. 2004; Walter et al. 2006) and anthropogenic (e.g., hydraulic fracturing) sources are causing its atmospheric concentration to rise (e.g., Howarth 2015; Howarth et al. 2011). This increase is of acute concern because methane is a stronger GHG than CO_2—it has a 28 times larger global warning potential (GWP) than CO_2 (Myhre et al. 2014), a measure of the radiative forcing of 1 ton of emissions of a given gas over a time horizon, relative to the emissions of 1 ton of CO_2 on a 100-year time horizon.

Aerosols, particles in the atmosphere formed through physical and chemical

processes, have lifetimes generally many orders of magnitude shorter than CH_4. Though the term 'aerosol' encompasses a myriad of chemical species from both natural (e.g., volcanic eruptions) and anthropogenic sources (e.g., biomass burning) with differing climate effects, the net anthropogenic RF on climate is negative and partially offsets the WMGHG contributions (Myhre et al. 2014).

The impacts of anthropogenic aerosols are usually taken as the sum of the direct and indirect effects (Mahowald et al. 2011). Figure 1.1 shows the aerosol impacts on Earth's radiative budget and climate (Solomon et al. 2007). The direct effect refers to the scattering and absorption of radiation by the particles, while the semi-direct and indirect effects refer to the influence of the aerosols on cloud properties and lifetime (Charlson et al. 1992). Taken as the sum, the global mean RF of aerosols and their precursors, including mineral dust, sulfate, nitrate, organic carbon, and BC as –0.35 Wm^{-2} (Myhre et al. 2014). (Note: The radiative forcing is for the period 1750-2011 and includes an uncertainty range of –0.85 to 0.15 Wm^{-2}. See Table 8.6 in Myhre et al. 2014.)

The radiative forcing of BC is a function of its emissions (EBC) given by:

$$RF_{BC} = 0.0743\ Wm^{-2}\ TG^{-1} \times EBC \tag{1.3}$$

where the coefficient includes both indirect and direct forcings of BC (fossil fuel and biomass; see Table C1 in Bond et al. 2013).

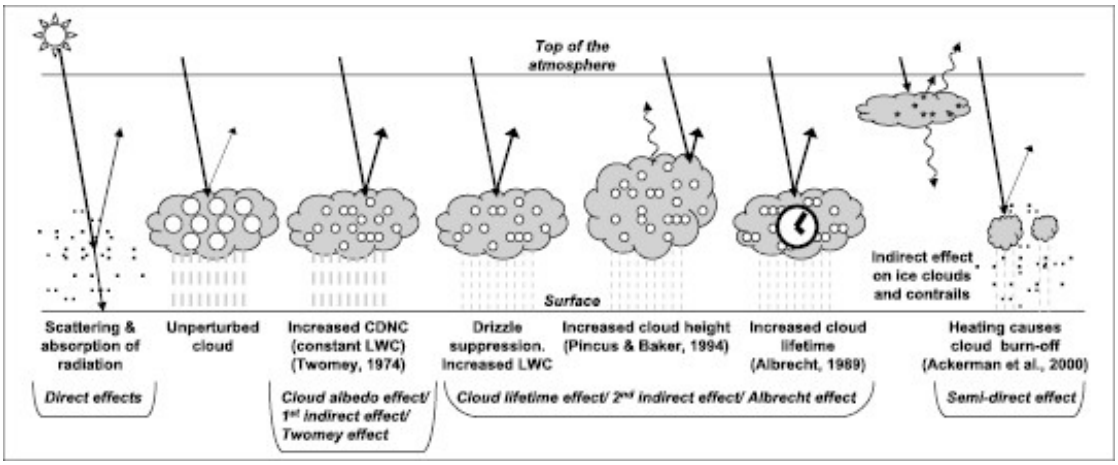


Fig. 1.1: Schematic diagram of various direct and indirect effects from aerosols (Solomon et al. 2007).

1.1.2 Temperature response

The response of the climate system to perturbations is commonly described as changes in global-mean surface temperature (GMST). At equilibrium, total TOA RF (RF_{total}) is related through a linear relationship to the change in GMST (ΔT_s), given by:

$$\Delta T_s = \lambda \times \ RF_{total}(t) \tag{1.4}$$

where λ is the climate sensitivity parameter, the inverse of the feedback parameter, in units of °C/Wm^{-2} (Forster et al. 2007).

In 2017, human-induced warming reached approximately 1°C (*likely*— >66% probability — between 0.81-2°C) above pre-industrial levels (Allen et al. 2018a), increasing at 0.2°C (*likely* between 0.1-0.3°C) per decade, an unprecedented rate in modern time (Smith et al. 2015). Recent international agreements, such as the Paris Climate Accord, are attempting to limit end-of-century global mean temperature

rise to 1.5°C (Allen et al. 2018a). Even central estimates of future global mean temperature, however, estimate the rise in temperature to be 1.8-3.2°C by 2100, relative to pre-industrial levels (Fricko et al. 2017; Riahi et al. 2017; see Figure 1.2), and calls into question the ability of Paris pledges to reach the temperature target (Allen et al. 2018a).

Uncertainty in long-term projections of future warming derives from several sources extrinsic (i.e., assumptions about future GHG emissions or policy decisions) and intrinsic (i.e. internal variability or climate response to GHG concentrations) to the climate system. Examples of the intrinsic modes of variability that influences GMST on seasonal to multi-decadal to centennial timescales includes El Niño–Southern Oscillation or the Atlantic Meridional Overturning Circulation, among others (Canty et al. 2013; Easterling and Wehner 2009; Hope et al. 2017; Meehl et al. 2007; Stocker et al. 2013). Further, for a given future scenario of anthropogenic emissions, the carbon cycle adds uncertainty in the relationship between emissions and concentrations, and physical processes and feedbacks contribute to the uncertainty in the climate response to a given GHG concentration or aerosol forcing (Friedlingstein et al. 2014). Thus, there is increasing uncertainty at each step in the cause-effect chain (Myhre et al. 2014). We make note of this in Chapter 2.

It is the uncertainty in climate sensitivity (CS) values, λ in Equation 1.4, which is a measure of the temperature increase for a unit increase in forcing, that provides the main source of uncertainty in long-term projections of future warming (Hawkins and Sutton 2009). The range in CS derives from our lack of understanding

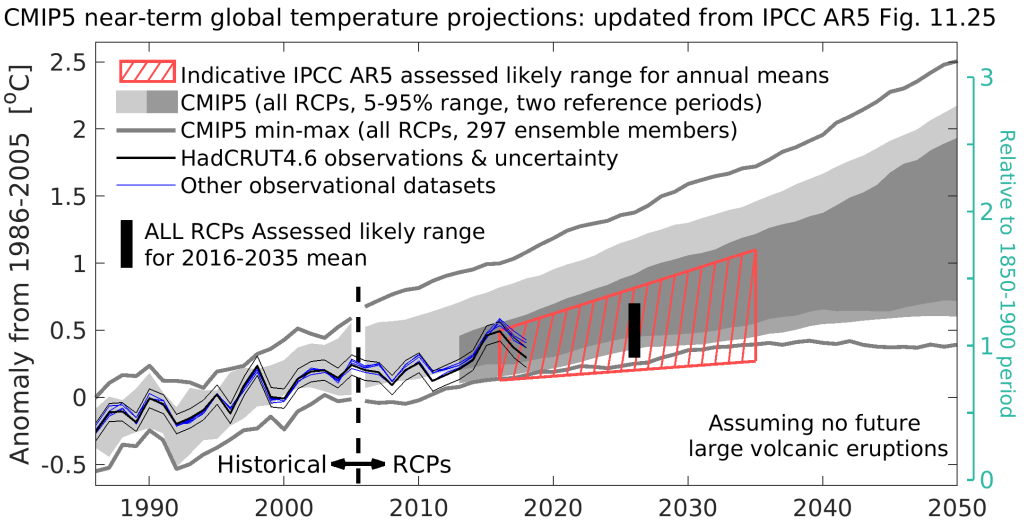


Fig. 1.2: Global temperature projections showing an updated version of Figure 11.25b from IPCC AR5. CMIP5 model projections relative to 1986-2005 are shown in light grey, and those relative to 2006-2012 are shown in dark grey. The IPCC AR5 "likely range" for global temperatures in the 2016-2035 period is shown in red hatching, and the black bar is the 2016-2035 average. The HadCRUT4.6 global temperature time-series and uncertainty are shown in the black line, while the other observational datasets are shown in the blue lines (Cowtan and Way, NASA GISTEMP, NOAA GlobalTemp, BEST). The green axis (right) shows the temperatures relative to 1850-1900. (Image credt: Ed Hawkins and the Climate Lab Notebook.)

and observations of physical feedbacks, such as the large range in cloud feedbacks, +0.6 (−0.2 to +2.0 $Wm^{-2}/°C$) (Boucher et al. 2013) and the uncertainty in aerosol forcing (Shindell 2014).

Model and instrumental records indicate that the transient climate response, the temperature at CO_2 doubling, is *likely* to be in the range of 1 - 2.5°C (Stocker et al. 2013). By contrast, equilibrium climate sensitivity, defined as the global temperature rise following a doubling of CO_2 concentration in the atmosphere compared to pre-industrial levels, ranges in values from 2–4.5°C with the best estimate of 3°C in the IPCC AR4 (see Figure 1 in Ceppi et al. 2017). This has been subsequently revised to a *likely* range of 1.5–4.5°C in the IPCC AR5 (Stocker et al. 2013). In Chapter 5, we explore the range of climate response over time.

1.2 Climate models of varying complexity

Climate models are one of the primary tools used to understand changes in Earth's climate. Models are generally classified by their complexity and comprehensiveness, ranging from energy balance models (EBMs) to Earth System Models (ESMs) (see Figure 1.3). Here, we provide an overview of the classifications of climate models and place them into the context of the literature and this dissertation.

1.2.1 *Energy balance models*

Numerical models were originally developed to address fundamental climate questions before the advent of readily-available modern computational power (Bryan et al. 1982). During this period, simple models were generally used to understand Earth's energy balance (EBMs, e.g., Budyko 1969; Sellers 1969) and the climate and carbon-cycle response to anthropogenic perturbations (Hansen et al. 1984; Hoffert et al. 1980). EBMs are still employed to answer fundamental climate questions and disentangle climate responses, though the community notes their abilities are limited (Dommenget and Flöter 2011; Geoffroy et al. 2013).

1.2.2 *Simple climate models*

Simple climate models are one step in complexity above EBMs and have representations of the most fundamental equations driving the Earth system, such the carbon cycle. SCMs have several advantageous features (e.g., low computational intensity, ease of use, etc.), and are commonly used to addresses computationally intensive questions, such as probabilistic studies (van Vuuren et al. 2008), or within interdisciplinary models, which typically represent the dynamics within the human system such as energy systems and land-use changes (Strassmann and Joos 2018).

We follow the established literature and categorize simple climate models (SCMs) into two broad categories based on their structure: comprehensive SCMs, physically based SCMs, and idealized SCMs, often utilizing sums of exponentials. We provide a brief introduction to these classes of models below, while noting the

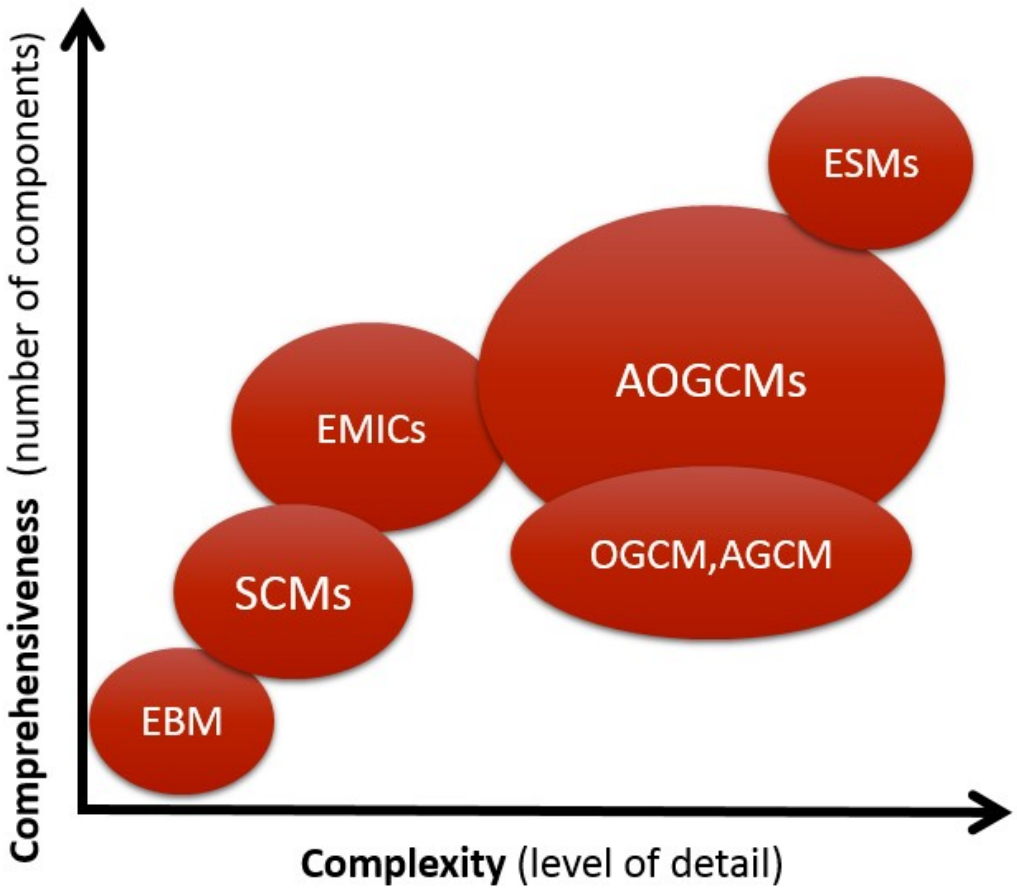


Fig. 1.3: Diagram of climate models of varying complexity reproduced from Provenzale 2014.

more details are available in Chapter 2, part of (Schwarber et al. 2019).

Idealized SCMs. Idealized SCMs and sums of exponentials are commonly used to investigate a number of climate indicators, such as transient climate response, and other metrics (Allen et al. 2018b; Millar et al. 2017). In addition, sums of exponentials are also commonly used to calculate other climate metrics, such as global warming potential (GWP) and global temperature potential (Aamaas et al. 2013; Berntsen and Fuglestvedt 2008; Fuglestvedt et al. 2010; Peters et al.

2011). Individual idealized SCM developers explore the ability of impulse-response functions to simulate the climate or carbon-cycles (Hooß et al. 2001; Millar et al. 2017; Strassmann and Joos 2018), often comparing to more complex models (Joos and Bruno 1996).

In Chapter 2, we characterize the response of two idealized SCMs to fundamental impulse tests—the Finite Amplitude Impulse Response Model (FAIR; Millar et al. 2015) and the impulse response model used in the IPCC AR5 (Flato et al. 2014).

Comprehensive SCMs. Comprehensive SCMs are commonly used in applications requiring a physically-based climate component, without the of disadvantages of more complex models—for example within human-Earth system models (Hartin et al. 2015; Ortiz and Markandya 2010; Schneider et al. 2000). Other studies evaluate SCMs based on their abilities to simulate the climate or carbon cycle (Friedlingstein et al. 2014; Joos et al. 1999; Knutti et al. 2008; Meinshausen et al. 2009; Rogelj et al. 2014; Schimel 1998; van Vuuren et al. 2011b) or to emulate the behavior of more complex models (Hope 2006; Kleidon and Renner 2017; Knutti et al. 2008; Meinshausen et al. 2008, 2011; Monckton et al. 2015). Often, this SCM analysis includes complex model emulation to understand the behavior of anthropogenic perturbations, and to address model spread in the various model intercomparison projects (MIPs) (Knutti and Sedláček 2013).

In Chapter 2, we characterize the response of three comprehensive SCMs to fundamental impulse tests—MAGICC versions 5.3 and 6.0 (Meinshausen et al. 2009, 2011), and Hectorv1.1 (Hartin et al. 2015). The wide and varied uses of such SCMs

illustrates the necessity of evaluating their behavior using fundamental impulse tests, which we explore in Chapter 2.

1.2.3 Earth System Models of Intermediate Complexity

Earth System Models of Intermediate Complexity (EMICs) fill the gap between SCMs and complex model, such as Earth System Models (ESMs). With generally simpler representation of one or more components of the Earth System, EMICs produce low internal variability with a faster integration time than fully coupled complex climate models (see below). This allows EMICs to potentially contain more subcomponent models than their ESM counterparts, and serve as a test bed for new model subcomponents (e.g., ice sheet models). Further, EMICs are often used to investigate system processes over inherently long time scales, making them ideal for testing paleoclimate reconstructions and understanding the forced components of the climate system (Claussen et al. 2002; Eby et al. 2013). Further, EMICs are commonly involved in intercomparison projects where they are compared with more complex models (Weber 2010), and citations therein) (see below). In our case, we use results from several EMICs to compare to our SCMs (see Chapter 2; Joos et al. 2013).

1.2.4 Complex Models: Coupled Atmosphere-Ocean General Circulation Models and Earth System Models

Coupled Atmosphere-Ocean General Circulation Models (AOGCMs) are general circulation models that solve the fundamental equations of motion with both

complex and comprehensive representations of the climate system, often requiring supercomputers to run. Slightly simplified versions of AOGCMs focus on one portion of the Earth system and are referred to as ocean GCMs (OGCMs), atmospheric GCMs (AGCMs), or terrestrial/land models. Often the terms AOGCMs and Earth System Models (ESMs) are used interchangeably, however ESMs also include interactive biogeochemical processes that interact with the climate system (Flato 2011).

Complex models, such as ESMs, are one of the most comprehensive tools available to understand changes in climate. Increasingly sophisticated computational abilities have allowed the comparison of ESMs to observational records and other complex models (Olivié et al. 2012) through coordinated international efforts (e.g., Friedlingstein et al. 2014; Gillett et al. 2013; Good et al. 2013; Knutti and Sedláček 2013; Schwinger et al. 2014, etc.)

CMIP5. A commonly used database of complex model results, Phase 5 of the Coupled Model Intercomparison Project (CMIP5), is a product of such efforts and serves as a significant scientific resource supporting many studies, including the IPCC reports.

We use complex model results from CMIP5 in subsequent chapters, forming one of the primary model data sources for this dissertation. In Chapter 2, we use results from the 4xCO_2 experiment, a stylized CMIP5 experiment whereby CO_2 concentration is instantaneously quadrupled. In Chapter 4, we use results from several CMIP5 experiments: PiControl (pre-industrial control, run for hundreds of years), past1000 (paleoclimate run from 850 to 1850), and historical (historical

simulation initiated from the past1000 and run for hundreds of years from ∼1850 to 2012). Finally, in Chapter 5, we return to astylized experiment and use results from the 1pctCO_2 experiment, whereby CO_2 concentration is increase 1% yr^{-1} until doubling. We generally refer to the complex models as ESMs throughout for simplicity, while noting the differences here.

Advances in Earth System modelling is not without its criticisms (Annan and Hargreaves 2017). As ESMs become more complex (e.g., ice sheet dynamics, cloud microphysics schemes, biogeochemical cycles) they can also become "black boxes" and scientists may have greater difficulty extracting meaning from model responses. Scientists often rely on the suite of model types and observations to provide insights into complex model behavior, therefore. We discuss more about the observations used in this thesis in Chapter 3.

1.3 Thesis objectives and outline

The literature above uses climate models to address a variety of scientific and policy questions. Many studies, however, do not fully characterize the models which underpin their analysis. Our work herein focuses on robustly characterizing climate models of varying complexity, from SCMs to ESMs, and seeks to take advantage of the full range of climate information available, including historical observations and paleoclimate reconstructions, to answer several important questions (numbered) and discuss the methods employed to address these questions (bullets):

1. **How do SCM responses compare under fundamental unit tests and**

how well do they compare to more complex models?

- Conduct fundamental impulse response tests in SCMs ranging in complexity from comprehensive SCMs to idealized SCMs using three main forcing and emission impulse tests to understand the response of the climate system and gas cycles in the models: (a) a concentration impulse of CO_2, (b) emissions impulses of BC, CH_4, or CO_2, (c) a 4xCO_2 step increase in CO_2.
- Compare response tests to more complex models by relying on previous work and results from stylized CMIP5 experiments.
- Provide a set of tests that we recommend as a standard validation suite for any SCM.

2. **How realistic is complex climate model variability compared to observations?**

   - Find the average spectral density of simulation and observation data over PAGES2k regions and for select time bins, focussing on periods of interest.
   - Understand the range of unforced variability ("noise") in complex models to infer reliability of spectral analysis.
   - Compare average spectral density from the past1000, PiControl, and Historical CMIP5 experiments with the relevant historical observational datasets and the PAGES2k paleoclimate reconstruction.

3. **How does the climate response in CMIP5 models vary over time?**

   - Employ a quantitative approach to investigate the climate responses from two CMIP5 experiments: piControl (pre-industrial control, run for hundreds of years) and 1pctCO_2 (1% annual increase in $[CO_2]$; Taylor et al. 2012), using the rate of temperature response over time (RT_t) and land-ocean warming ratios (ϕ).
   - Produced results showing the change in response over time at a sub-global scale.
   - Use linear CO_2 forcing regime without relying on both temperature and limited forcing data.

The dissertation is divided into six chapters. The first (and present) chapter presents an introduction to climate change and the wide use of climate models of varying complexity to address questions of scientific and policy import. Chapter 2 presents the results of fundamental unit tests in several SCMs. This work is published in Earth System Dynamics (Schwarber et al. 2019). Chapter 3 thoroughly describes the observational data used in Chapters 4 and 5, including sensitivity analysis for choices made when analyzing the observational data. Chapter 4 characterizes forced and unforced variability in complex models and compared the variability to observations over spatial and temporal scales important for human systems. This work will be submitted to the Journal of Geophysical Research - Atmospheres (Schwarber et al. 2020a*in prep*). Chapter 5 describes the changing climate sensitivity in complex models and will be submitted to the Journal of

Climate (Schwarber et al. 2020b). The summary and future directions are provided in Chapter 6.

Chapter 2: Simple Climate Model Evaluation Using Fundamental Unit Tests (published as Schwarber et al., 2019)

2.1 Introduction

Simple climate models (SCMs) are widely used in scientific and decision-making contexts largely because of their advantageous features, including their ease of use and low computational intensiveness. In particular, SCMs are traditionally used within human-Earth system models. These models couple the climate system with representations of the dynamics within the human system (e.g., energy systems and land-use changes; Hartin et al. 2015; Ortiz and Markandya 2010; Schneider et al. 2000; Strassmann and Joos 2018) and are used to assess global forcing or temperature targets (e.g., Representative Concentration Pathways van Vuuren et al. 2011a, Shared Socioeconomic Pathways Moss et al. 2010). Several studies investigated potential sources of human-Earth system model uncertainty by exploring the climate components driving the models (Calel and Stainforth 2017; Harmsen et al. 2015; van Vuuren et al. 2011b, 2008). Van Vuuren et al. (2011b) concluded that in most cases the results from human-Earth system models and SCMs were similar to the more complex, coupled Earth System Models (ESMs; van

Vuuren et al. 2011b). The authors further noted that differences in SCM results can have implications for decision makers informed by such results, illustrating the need for improvements in uncertainty analysis (e.g., carbon cycle feedbacks). Harmsen et al. (2015) extended van Vuuren's analysis to investigate emission reduction scenarios by including non-CO_2 radiative forcing. The authors concluded that many models may underestimate forcing differences after applying emission reduction scenarios, due to the omission of important short-lived climate forcers, such as black carbon (BC).

Few studies utilize idealized SCMs in human-Earth system models because of their inability to represent nonlinear forcings, such as air-sea exchanges (Khodayari et al. 2013) or ocean chemistry (Hooß et al. 2001; Tanaka and Kriegler 2007). With simple extensions of the carbon cycle (e.g., ocean carbonate chemistry), both Hoos et al. (2001) and Tanaka and Kriegler (2007) found improved responses from their respective impulse response models, applicable when coupling to human-Earth system models (Tanaka and Kriegler 2007).

Comprehensive SCMs are also used to simulate the climate or carbon cycle (Friedlingstein et al. 2014; Joos et al. 1999; Knutti et al. 2008), explore responses to anthropogenic perturbations (Geoffroy et al. 2013; Hope 2006; Meinshausen et al. 2009; Rogelj et al. 2014), or address model spread in the various model intercomparison projects (MIPs; Knutti and Sedláček 2013; Monckton et al. 2015; Rogelj et al. 2012). These analyses often include comparisons to more complex models (Meinshausen et al. 2008, 2011). One comprehensive SCM in particular, MAGICC 6.0, is used as a reference in many studies because of its well-documented

ability to emulate complex models (e.g., van Vuuren et al. 2011b).

Similarly, individual idealized SCM developers also explore the ability of impulse-response functions to simulate climate or carbon-cycle responses to perturbations (Hooß et al. 2001; Millar et al. 2017; Sausen and Schumann 2000; Strassmann and Joos 2018; Thompson and Randerson 1999), often comparing to more complex models (Joos and Bruno 1996). Sand et al. (2016), for example, employed an idealized SCM using sums of exponentials (AR5-IR) to find the Arctic temperature response to regional short-lived climate forcer emissions (e.g., BC) and compared these responses to more complex models (Sand et al. 2016).

Climate indicators, such as transient climate response (TCR; Allen et al. 2018b; Millar et al. 2017), can also be informed using SCMs. TCR is the measure of the climate response to a 1% yr^{-1} increase in CO_2 concentration until doubling of CO_2 relative to pre-industrial level. TCR is useful for understanding the climate response on shorter time scales, as CO_2 concentration doubling takes place in 70 years, a time-frame relevant for many planning decisions (Flato et al. 2014; Millar et al. 2015). TCR and the equilibrium climate sensitivity (ECS) can be combined to estimate the realized warming fraction (RWF), the fraction of total warming manifested up to a given time. Millar et al. (2015) investigated TCR and ECS within an impulse-response model to show the implications of these values on future climate projections by specifically looking at the RWF (Millar et al. 2015).

Sums of exponentials are also commonly used to calculate other climate metrics, such as the global warming potential (GWP) and global temperature potential (GTP; Aamaas et al. 2013; Berntsen and Fuglestvedt 2008; Fuglestvedt

et al. 2010; Peters et al. 2011; Sarofim and Giordano 2018). Idealized SCMs, however, often do not account for carbon cycle feedbacks, important for more realistic representations of climate. Both Millar et al. (2017) and Gasser et al. (2017) investigated the effects of adding carbon cycle feedbacks on these metrics produced with idealized SCMs, and found that accounting for feedbacks improved model responses (Gasser et al. 2017; Millar et al. 2017).

Despite their importance and wide use, the fundamental responses of SCMs have not been fully characterized (Thompson 2018). In this paper, we use impulse-response tests to address this gap.

2.2 Methods

2.2.1 Fundamental impulse tests.

Impulse-response tests characterize the SCMs' climate and gas-cycle response to a forcing or emission impulse (Good et al. 2011; Joos et al. 2013). Though fundamental impulse tests have been used in the literature (e.g., Joos et al. 2013), we employ these existing techniques to evaluate several SCMs. In fact, the U.S. National Academies of Science (National Academies of Sciences and Medicine 2016) specifically suggested that SCMs be "assessed on the basis of the response to a pulse of emissions," which we do here.

We use three main forcing and emission impulse tests to understand the response of the climate system and gas cycles in the models: (a) a concentration impulse of CO_2, (b) emissions impulses of BC, CH_4, or CO_2, (c) a $4\text{x}CO_2$ step

increase in CO_2 *concentration*. We carry out these experiments by instantaneously increasing emissions or forcing values in 2015 to avoid the model base years of our SCMs.

We begin with CO_2 because this well-mixed greenhouse gas is the largest contributor to anthropogenic forcing changes (Myhre et al. 2014). Methane is also of interest because it is a shorter-lived greenhouse gas, with its major sink from chemical interactions with the hydroxyl radical and other species (Cicerone and Oremland 1988). Finally, we use BC perturbations to represent aerosols more generally because we are interested in model responses to short-lived climate forcers (Bond et al. 2013; Harmsen et al. 2015). SCM representations of other aerosols species are similar so we do not conduct impulse tests of other species.

In many SCMs, forcing over historical periods is explicitly calibrated to a model base year, so it is not possible to conduct perturbations during these time periods. Therefore, our perturbations are conducted for 2015 to avoid the model base years of our SCMs. We show some model responses out to 2300, the end of the MAGICC model runs, equal to 285 years after the perturbation.

We run reference scenarios in the SCMs, followed by each perturbation case described below. For each experiment (see below) we report the response, which is obtained by subtracting the reference from the perturbation results. For instance, the CO_2 concentration response is obtained as follows:

$$CO_{2\ concentration\ response}(t) = CO_{2\ concentration\ perturabtion}(t) - CO_{2\ concentration\ reference}(t) \tag{2.1}$$

We conducted the following impulse tests:

Concentration impulse (CO_2). These SCMs can be used in a mode where CO_2 concentrations are exogenously specified. We carry out this experiment by instantaneously increasing CO_2 concentration by 200 ppm in 2015. After 2015, CO_2 concentrations return to the baseline levels following the published RCP4.5 scenario. Note, we do not conduct separate forcing impulse experiments because this is functionally equivalent to a concentration impulse. In this experiment, we are only interested in the dynamics of the models' temperature response. This experiment eliminates the added uncertainty in the emissions to concentrations calculation and complicating factors from carbon cycle feedbacks.

Emissions impulse (BC, CH_4, CO_2). For this experiment all models were run with an emissions input. We carry out this experiment by increasing individual emissions (BC, CH_4, or CO_2) in one year. Following that year, the emissions return to the RCP4.5 pathway for all subsequent years. In this experiment CO_2 and other GHG concentrations are allowed to vary as determined by each model. We find our perturbation values by doubling the 2015 value for each chemical species equal to a 9.2 PgC pulse of CO_2, a 329 Tg pulse of CH_4, and a 7981 Gg pulse of BC. We also perturb CO_2 emissions in 2010, 2020, 2030, 2040, 2050 to understand changes in model responses over time and see a very small difference in the model response (S4). We compare results from three comprehensive SCMs to two IR models, AR5-IR and FAIR model (Millar et al. 2017; Myhre et al. 2014).

We also compared results to several ESMs and EMICs by carrying out a 100 GtC CO_2 impulse, following Joos et al. (2013). This is approximately 10x the CO_2

perturbation pulse described above (Joos et al. 2013).

Step increase in CO_2 concentration.

Similar to comparison (a), in this experiment, CO_2 concentrations are prescribed. We have CO_2 concentrations follow a pre-industrial pathway (278.0516 ppmv in 1765) until 2014. The CO_2 concentration is quadrupled (4x) in 2015 and maintained at this level until 2300. This follows the experimental protocol used in the CMIP5 experimental design (Taylor et al. 2012).

We compare these results to drift-corrected (Gupta et al. 2013) global mean temperature results from 20 complex climate models from the CMIP5 archive. We drift-correct the CMIP5 global mean temperature time series by subtracting the slope of the linear fit from the full-time series of the corresponding pre-industrial experiment for each individual model (see S3 of Schwarber et al. 2019 for details).

We note that impulse-response tests can be considered a type of unit test. Unit testing in software refers to a specific method of comparing output from the smallest portion of code, called a unit (i.e., function), to known outputs (Clune and Rood 2011). Here, we use this term in a similar way as van Vuuren et al. (2011b), where MAGICC 6.0 was used as the reference output to compare several human-Earth system models (van Vuuren et al. 2011b).

2.2.2 Background concentrations.

Our impulse response tests are conducted against a time-changing greenhouse gas (GHG) concentration background using emissions from the Representative

Concentration Pathway (RCP) 4.5 scenario (Thomson et al. 2011). For each test, therefore, we run a reference scenario in the SCMs, followed by each perturbation case. We report the response, which is obtained by subtracting the reference from the perturbation results for each model. A changing GHG background concentration is a more realistic scenario overall and also reveals biases not otherwise apparent under constant concentration conditions, for example, in SCMs insensitive to changing background concentrations. Further, for emissions impulses this methodology is more readily implemented as a standard impulse test, as we recommend below. Conducting tests against a constant concentration background in any but the most idealized SCM requires an inversion calculation to determine the emissions pathway that results in a constant concentration. This is an unnecessary barrier to conducting routine impulse response tests.

2.2.3 Models.

Three comprehensive SCMs–Hector v2.0 (Hartin et al. 2015; Kriegler 2005), MAGICC 5.3 BC-OC (Raper and Cubasch 1996; Smith and Bond 2014; Wigley and Raper 2002), and MAGICC 6.0 (Meinshausen et al. 2011)–are used in this study (S2). The models were selected based on their availability, use in the literature, and their applicability to decision making. We also include two idealized SCMs which employ sums of exponentials to represent the climate or gas-cycle responses, a general approach often used in the literature (Aamaas et al. 2013; Fuglestvedt et al. 2003), referred to as impulse response functions (IRFs). IRFs linearly approximate the response of a system to a given forcing (Hooß et al. 2001). A widely used version

tested here is the impulse response (IR) model used in the Intergovernmental Panel on Climate Change Fifth Assessment Report (Myhre et al. 2014; See Section 8.7.1.2 - 8.7.1.3; See Section 8.SM.11 for model equations), referred to here as AR5-IR. Additionally, we test version 1.0 of the Finite Amplitude IR (FAIR) model, an extension of AR5-IR including a representation of carbon cycle feedbacks and non-linear forcing (Millar et al. 2017).

IPCC AR5. The IPCC AR5 (See caption under Figure 8.28 in Myhre et al. 2014) describes the underlying multi-gas impulse response model used to quantify the multi-gas equivalence metric, Absolute Global Temperature Potential (AGTP), to compare temperature changes at a chosen time in response to a unit pulse of emissions i. AGTP is found via a convolution of the fraction of the species i remaining in the atmosphere after an emissions pulse and the climate response to a unit forcing (see See 8.7.13 and Equation 8.1 in Myhre et al. 2014). We refer to this model as AR5-IR. For this work, AR5-IR was recoded in R and is available for download with the Supplementary Materials of the published paper (Schwarber et al. 2019).

FAIR. The FAIR v1.0 model is a modified version of the AR5-IR carbon cycle component to include the state-dependence of the CO_2 airborne fraction to reproduce the relationship between CO_2-only emissions, concentrations, and temperature over the historical period. Millar et al. (2017) began with the impulse response functions used for calculation of multi-gas equivalence metrics in IPCC-AR5 (Myhre et al. 2014) and extended the CO_2 IRF by coupling the carbon-cycle to the thermal response and to cumulative carbon uptake by terrestrial and marine

sinks. FAIR is available for download at https://github.com/OMS-NetZero/FAIR (Millar et al. 2017.

FAIR calculates the global mean temperature response as the sum of the temperature response from the fast and slow timescale components, which represent the upper and deep ocean. The model does not report the internally-calculated forcing response, so this is not included in Figure 2.2 in the main paper.

Here, we use the first iteration of FAIR, but we note that two new versions have recently been published, FAIR v1.1 and FAIR v1.3. FAIR v1.3 extends the original version to, "calculate non-CO_2 greenhouse gas concentrations from emissions, aerosol forcing from aerosol precursor emissions, tropospheric and stratospheric ozone forcing from the emissions of precursors, and forcings from black carbon on snow, stratospheric methane oxidation to water vapour, contrails and land use change (Smith et al. 2018)." Additionally, a brief discussion of the parameter choices for this model is available in S2.4.

MAGICC 5.3 BC-OC.MAGICC 5.3 BC-OC is a version of MAGICC 5.3 developed in conjunction with the Global Change Assessment Model (GCAM). MAGICC 5.3 used here is available in GCAM version 4.4, available for download at https://github.com/JGCRI/gcam-core/releases. The major change in this version of MAGICC was the addition of explicit BC and OC (Smith and Bond 2014). We ran this model with all its default configuration settings unless otherwise noted in the text.

MAGICC 6.0. MAGICC 6.0 was run with all the default settings. For the main experiments, the climate sensitivity was set to 3.0° C to match

the default setting of MAGICC 5.3 BC-OC and Hector v2.0, unless otherwise noted. The MAGICC 6.0 executable is available for free download here: http://www.magicc.org/.

Hector v2.0 Settings. In the version we use here, Hector (v2.0), is coupled to a 1-D diffusive heat and energy balance model (DOECLIM: Diffusion Ocean Energy balance CLIMate model). We are using the 1-D diffusive ocean heat component of DOECLIM. DOECLIM is well documented and has been widely used in climate uncertainty studies (Bakker et al. 2017; Kriegler 2005; Urban et al. 2014). Using default Hector parameter values for climate sensitivity and heat diffusivity, we find that the new coupled model (Hector v2.0) exhibits improved vertical ocean structure and heat uptake, as well as surface temperature response to radiative forcing, compared to earlier versions of Hector.

2.2.4 Parameter selection.

We are testing the model responses as they would be 'out of the box' and only make modifications if required for the models to run. A model's ability to emulate an ESM or the multi-model ESM mean is generally explored by the individual SCM development teams, as noted in the references for the Hector, MAGICC, and FAIR models. While emulation is outside the scope of this paper, we conduct sensitivity tests by relying on parameters derived from ESM emulation experiments using MAGICC 6.0. We note that due to structural differences in the SCMs it is, in general, not possible to operate the models with identical parameter values. This reinforces the importance of conducting fundamental impulse response tests to

Model	Model description	Carbon cycle	Climate component
Hector v2.0 (Hartin et al., 2015, 2016; Kriegler, 2005)	mechanistic climate carbon-cycle model	One-pool atmosphere, three-pool land, and four-pool ocean	Global Energy balance model, with ocean heat diffusion
MAGICC 5.3 BC-OC (Raper and Cubasch, 1996; Smith and Bond, 2014; Wigley and Raper, 1992)	mechanistic climate carbon-cycle model	One-pool atmosphere, three-pool land, and one-pool ocean	4-box Energy balance model, with ocean heat upwelling diffusion
MAGICC 6.0 (Meinshausen et al., 2011)	mechanistic climate-carbon cycle model	One-pool atmosphere, three-pool land, and one-pool ocean	4-box Energy balance model, with ocean heat upwelling diffusion
AR5-IR (Myhre et al., 2013)	Impulse-response function	Impulse-response function	Equilibrium temperature as a function of RF
FAIR v1.0 (Millar et al., 2017)	Impulse-response function	Four timescale impulse-response function with state-dependence of the CO_2 airborne fraction	Equilibrium temperature as a function of RF; IRF with two timescales

Tab. 2.1: Main carbon cycle and climate characteristics of SCMs and IRFs.

quantify the behavior of the SCMs.

2.3 Results

In our paper, we evaluated the SCMs by comparing the models to each other and also, in the limited cases where this is possible, to more complex models. We compare against the suite of complex model results because it has been shown that the multi-model mean behavior of the complex models replicates well a broad suite of observations (e.g., Figure 9.7, Flato et al. 2014). We highlight differences in model responses to a suite of impulse tests to support an informed model selection.

We begin by testing the fundamental dynamics of the temperature response to a well-mixed greenhouse gas forcing impulse by perturbing CO_2 concentrations Figure 2.1, bypassing the carbon cycle (if present). We report both time-series responses (Figure 2.1a) and time-integrated responses (Figure 2.1b). Integrated responses form the basis of commonly used metrics, such as GWP and GTP (Fuglestvedt et al. 2010).

2.3.1 Responses to CO_2 Concentration Impulse.

First, we consider the comprehensive SCMs. Both versions of MAGICC show shifted responses in the first few years following the perturbation due to the way this model treats sub-annual integration of forcing. The shifted responses do not significantly impact integrated results. MAGICC 6.0 initially responds more strongly to the perturbation, with a 6% larger integrated temperature response 20

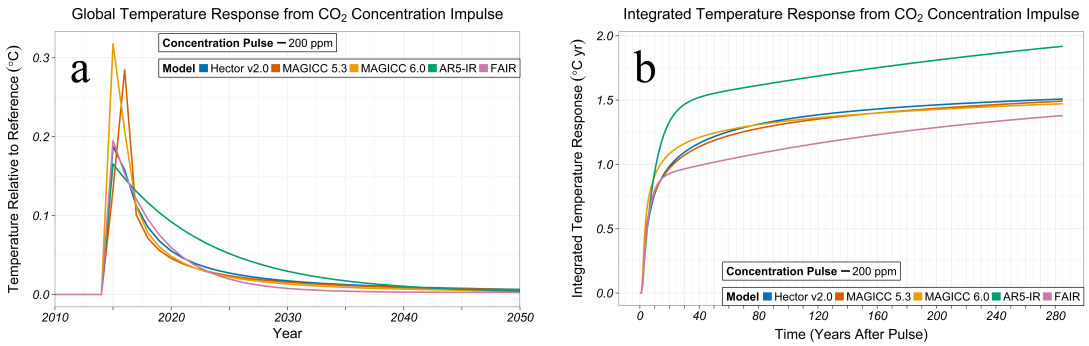


Fig. 2.1: Global mean temperature response (a) and integrated global mean temperature response (b) from a CO_2 concentration perturbation in SCMs (MAGICC 6.0 – yellow, MAGICC 5.3 BC–OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR – pink). The perturbations are conducted in 2015 against the background of the Representative Concentration Pathway (RCP) 4.5 scenario (see Methods). The time-integrated response, analogous to the absolute global temperature potential, is reported as 0–285 years after the perturbation.

years after the impulse compared to the comprehensive SCM average (S9). After 30 years, the comprehensive SCMs are within 2% of each other.

The idealized SCMs show varied responses to a CO_2 concentration impulse. Differences in the AR5-IR and FAIR responses are due to a nonlinearity also present in FAIR. According to Equation 8 in Millar et al. (2017) FAIR will have a differential response to changing background CO_2 concentrations. By contrast, AR5-IR parameterizes the climate response to a unit forcing using a sum of exponentials as given by Equation 8.SM.13 in Myhre et al. (2014).

AR5-IR has a much stronger response compared to the comprehensive SCMs; the integrated response is 6% larger than the comprehensive SCMs 20 years after the

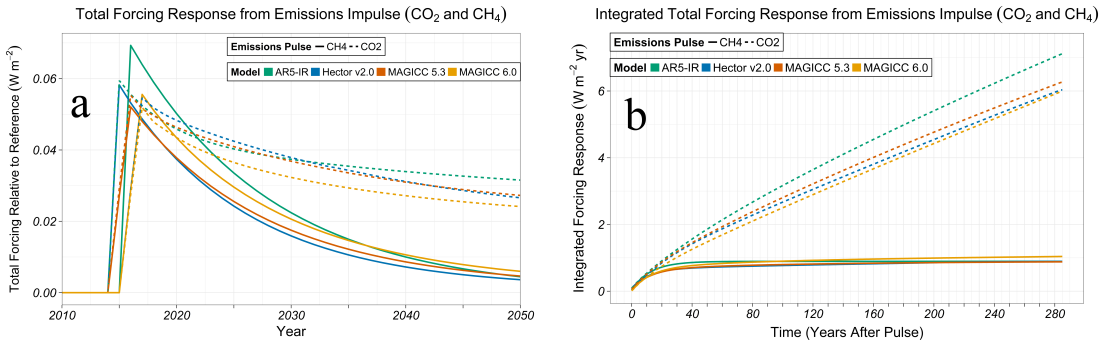


Fig. 2.2: Total forcing response from CO_2 (dashed) and CH_4 (solid) emissions perturbations in SCMs (MAGICC 6.0 – yellow, MAGICC 5.3 BC–OC – red, Hector v2.0 – blue, AR5-IR – green). FAIR does not report forcing. We report the total forcing response, which has slight differences from the gas-only forcing response. The perturbations are conducted in 2015 against the background of the Representative Concentration Pathway (RCP) 4.5 scenario (see Methods). The time-integrated response, analogous to the absolute global warming potential, is reported as 0–285 years after the perturbation.

pulse, increasing to 30% by the end of the model runs. This large difference is due to the absence of feedbacks and nonlinearities in the AR-IR model. FAIR contains an approximate representation of these nonlinearities, responding similarly to the comprehensive SCMs in the near-term, but has a 7% weaker integrated response 285 years after the impulse. The approximations used to represent the carbon cycle and non-linear forcing might account for this, but it is unclear from these results.

2.3.2 *Responses to Emissions Impulses.*

We now test the model response to an emissions impulse. Compared to forcing-only experiments, emissions perturbation experiments have additional levels of uncertainty from the conversion of emissions to concentrations, as well as carbon cycle feedbacks. As a diagnostic we examine the forcing response, functionally equivalent to examining the concentration response. The three comprehensive SCMs have small differences (<10%) in the integrated forcing response (Figure 2.2) from CO_2 (dashed) emission impulses for all time horizons. AR5-IR, an idealized SCM, responds 11% stronger than the comprehensive SCMs average 20 years after the pulse, increasing to a 17% difference in the integrated response 285 years after the impulse. FAIR does not calculate concentration or forcing, so cannot be included in these comparisons.

We complete the model response sequence by examining the temperature response from emissions perturbations, which is conceptually the combination of the temperature response from a concentration impulse (Figure 2.1) and the forcing response from an emissions impulse (Figure 2.2). Similarities in the comprehensive SCM responses in (Figure 2.1) and (Figure 2.2) are reflected in the <5% difference in the temperature response from a CO_2 emissions perturbation 20 years after the impulse (Figure 2.3a). AR5-IR responds 30% stronger and FAIR <10% weaker compared to the comprehensive SCM average 20 years after the perturbation (Figure 2.3a). FAIR introduces a state-dependent carbon cycle representation (Millar et al. 2017) and is, in general, an improvement over AR5-IR, but shows a systematic

difference with the comprehensive SCMs.

We also indirectly compare the time-integrated airborne fraction in our SCMs to three comprehensive ESMs and seven Earth System Models of Intermediate Complexity (EMICs) using results from the Joos et al. (2013) 100 GtC CO_2 pulse experiment, henceforth referred to as Joos et al. Unlike Joos et al., we conduct this experiment with a changing background concentration. We use the RCP 4.5 scenario and add a 100GtC CO_2 pulse for 2015. The airborne fraction, which is defined as the ratio of the annual increase of atmospheric CO_2 to total emissions from anthropogenic sources, is therefore higher in our results. Despite the difference in methodology, comparing the MAGICC 6.0 results here and in Joos et al. allows us to use transitive logic to draw broader conclusions about the other comprehensive SCMs. We note that the Joos et al. MAGICC 6.0 ensemble mean airborne fraction is similar to their multi-model mean at each time horizon. Because Hector and MAGICC 5.3 have a similar response to MAGICC 6.0 in our results, we conclude that the comprehensive SCM carbon cycle representations generally capture ESM and EMIC responses to the extent this can be evaluated for indirect comparison.

Table 2.2 shows the time-integrated airborne fraction at chosen time horizons from the 100 GtC pulse of CO_2 emissions. The Table 2.2 results are graphically represented in Figure 2.3. These results are largely discussed in the main paper.

Similarly, we compare the temperature response of the comprehensive SCMs to Joos et al. MAGICC 6.0 was used both here and by Joos et al., and we find similar responses with $\leqslant 1$°C yr difference from Joos et al. at each reported period. Though the other two comprehensive SCMs were not used by Joos et al., their

Time Horizon	20 yr	50 yr	100 yr
NCAR CSM1.4	13.8	27.8	46.6
HadGEM2-ES	14.7	30.9	53.3
MPI-ESM	14.5	29.2	48.8
Bern3D-LPJ (reference)	15.4	34.3	61.9
Bern3D-LPJ ensemble	15.1 (14.0-16.0)	32.7 (28.9-36.0)	57.6 (48.9-65.6)
Bern2.5D-LPJ	13.9	29.7	51.1
CLIMBER2-LPJ	13.0	26.8	49.2
DCESS	14.6	31.8	56.3
GENIE ensemble	13.6 (10.9-17.6)	28.9 (21.7-41.4)	50.5 (38.3-77.9)
LOVECLIM	13.5	27.9	45.3
MESMO	15.1	33.6	61.1
UVic2.9	13.7	29.5	53.0
ACC2	13.7	27.9	46.5
Bern-SAR	14.0	29.0	48.9
TOTEM2	16.9	38.3	66.6
MAGICC 6.0 ensemble	14.0 (12.0-16.1)	29.6 (23.6-35.7)	51.8 (40.0-64.2)
Multi-model mean	14.3 ± 1.8	30.2 ± 5.7	52.4 ± 11.3
Hector v2.0	16.2	34.0	58.3
MAGICC 5.3	16.0	33.4	58.3
MAGICC 6.0	15.3	32.2	57.9
AR5-IR	15.0	31.0	53.1
FAIR	14.6	32.6	61.6

Tab. 2.2: Time-integrated Airborne Fraction from a 100 GtC CO_2 Emissions Impulse in SCMs Compared to Results from Table 4 in Joos et al. (2013).

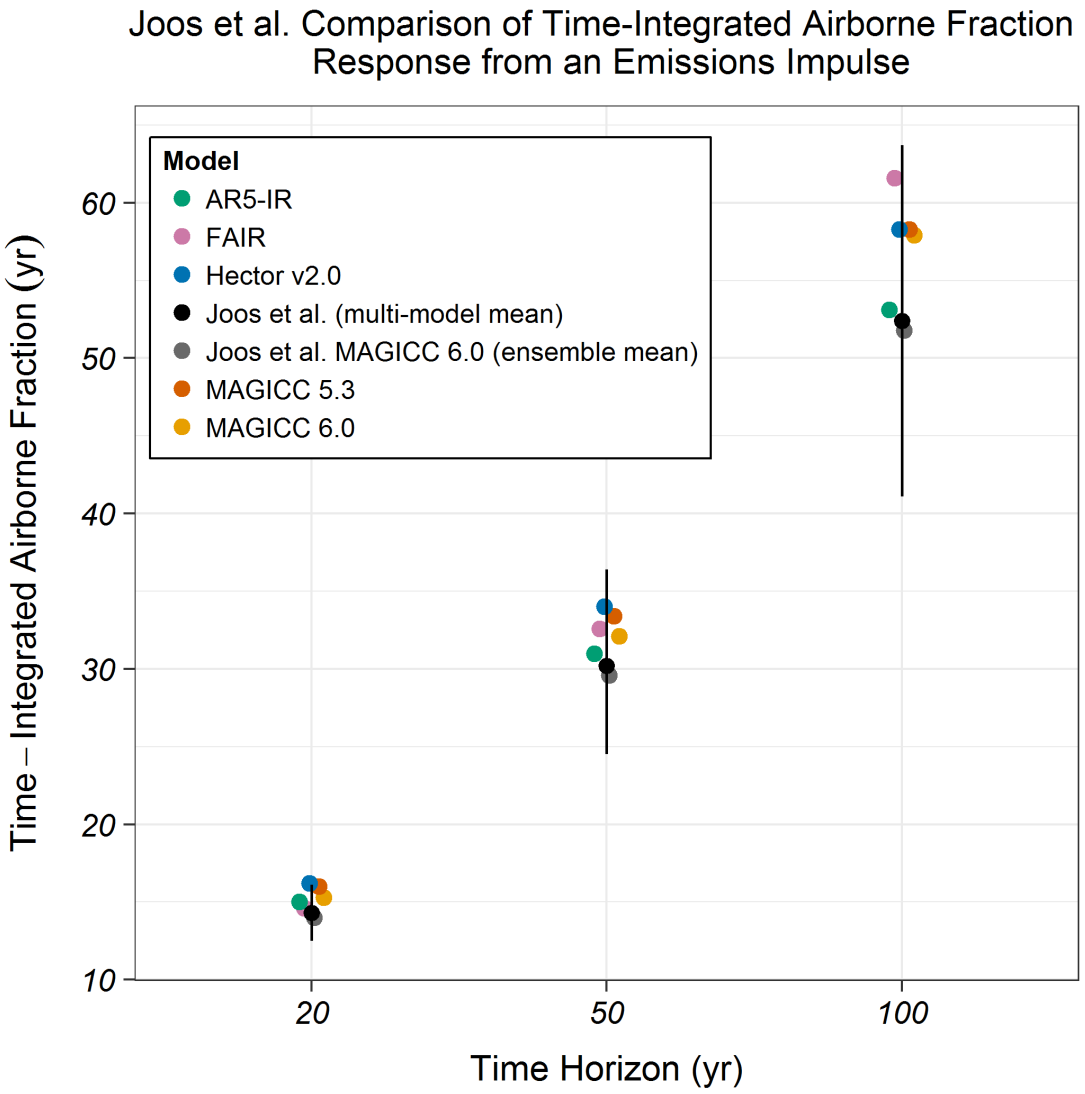


Fig. 2.3: Time-integrated airborne fraction from a 100GtC CO_2 emissions impulse in SCMs compared to Joos et al. This is not a direct comparison because we did not perform this experiment with a constant CO_2 concentration background, as done by Joos et al. The colored points represent the time-integrated airborne fraction in the SCMs used in this study, following Joos et al., and the Joos et al. MAGICC 6.0 ensemble mean. The black point is the Joos et al. multi-model mean and the vertical black line represents the range of the Joos et al. model results. (Joos et al. MAGICC 6.0 ensemble mean –grey, MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR –pink).

Time Horizon	20 yr	50 yr	100 yr
NCAR CSM1.4	2.53	7.36	10.6
HadGEM2-ES	4.24	12.4	30.3
MPI-ESM	3.83	8.84	19.1
Bern3D-LPJ (reference)	4.11	12.1	24.5
Bern3D-LPJ ensemble	3.20 (2.1-4.6)	8.61 (5.1-13.5)	17.3 (9.5-29.3)
Bern2.5D-LPJ	3.15	8.40	17.1
CLIMBER2-LPJ	3.05	7.96	16.5
DCESS	3.38	9.96	20.6
GENIE ensemble	3.77	10.54	21.6
LOVECLIM	0.22	3.46	7.83
MESMO	4.41	12.5	26.0
UVic2.9	3.40	9.17	18.5
ACC2	3.99	10.55	20.0
Bern-SAR	n/a	n/a	n/a
TOTEM2	n/a	n/a	n/a
MAGICC 6.0 ensemble	3.64 (2.7-4.7)	8.96 (6.6-12.7)	17.2 (12-26)
Multi-model mean	3.29 ± 2.03	9.13 ± 4.45	18.7 ± 11.1
Hector v2.0	3.05	8.20	15.54
MAGICC 5.3	3.13	8.19	15.73
MAGICC 6.0	3.39	8.28	15.54

Tab. 2.3: Time-integrated temperature response from a 100 GtC CO_2 Emissions Impulse in SCMs Compared to Results from Table 7 in Joos et al. (2013).

similar responses to our MAGICC 6.0 allow us to make a larger conclusion, as done in the main paper. Using this logic, we are able to validate our SCM responses from a finite pulse. We find that the comprehensive SCM responses are generally less varied, close to the Joos et al. ensemble mean 20 years after the pulse, and below most Joos et al. model responses 50 and 100 years after the pulse. We find that the comprehensive SCMs capture ESM and EMIC responses in the near-term, with expected differences in response over longer time horizons due to rising background concentrations.

For idealized SCMs, we find that under changing background conditions, FAIR underestimates the airborne fraction compared to the Joos et al. multi-model mean at each time horizon. Without a physical processes-based carbon cycle, AR5-IR is

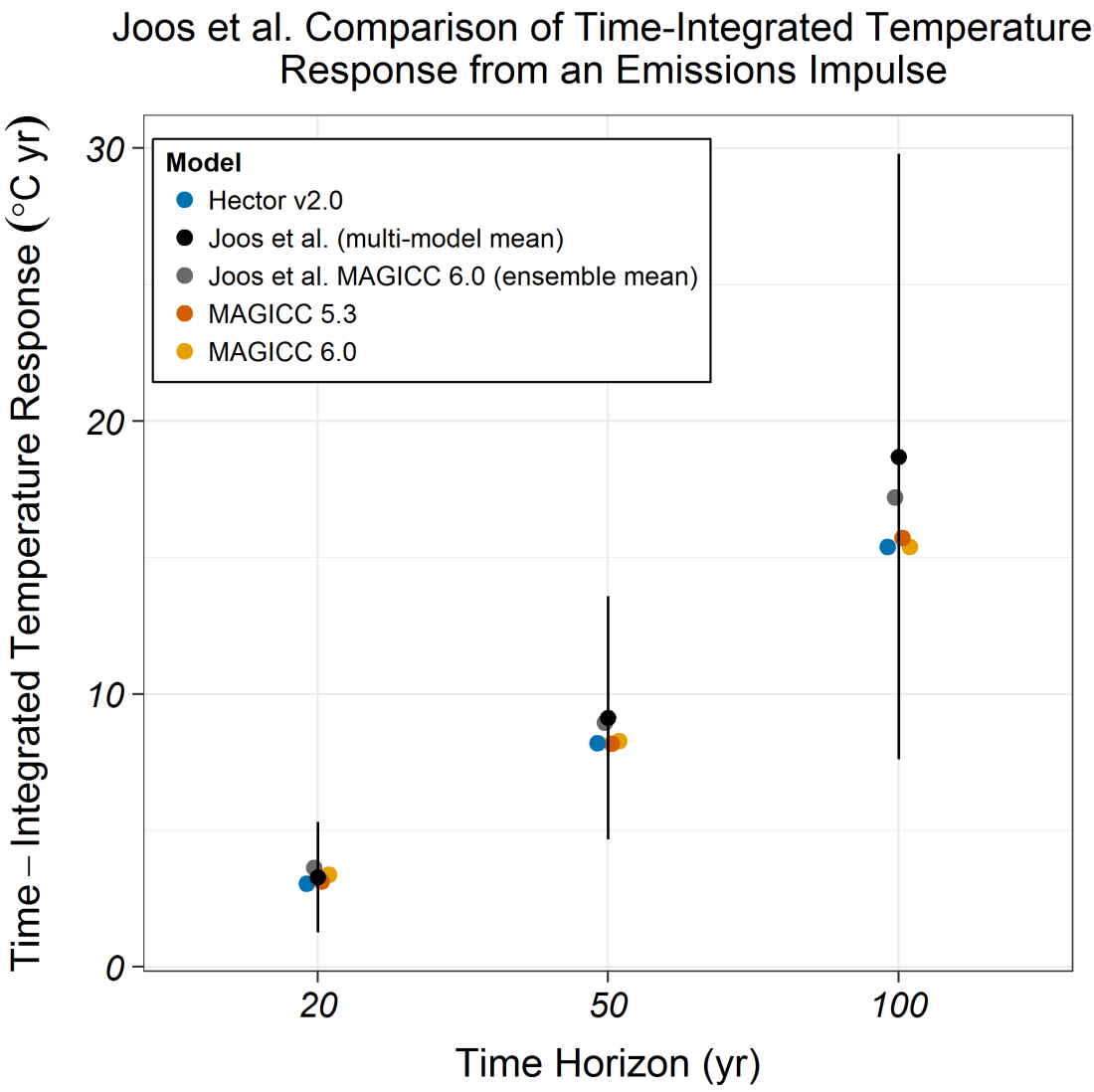


Fig. 2.4: Time-integrated temperature response from a 100GtC CO_2 emissions impulse in SCMs compared to Joos et al. This is not a direct comparison because we did not perform this experiment with a constant CO_2 concentration background, as done by Joos et al. The colored points represent the time-integrated temperature response in the SCMs used in this study, following Joos et al., and the Joos et al. MAGICC 6.0 ensemble mean. The black point is the Joos et al. multi-model mean and the vertical black line represents the range of the Joos et al. model results. (Joos et al. MAGICC 6.0 ensemble mean –grey, MAGICC 6.0 – yellow, MAGICC 5.3 BC-OC – red, Hector v2.0 – blue).

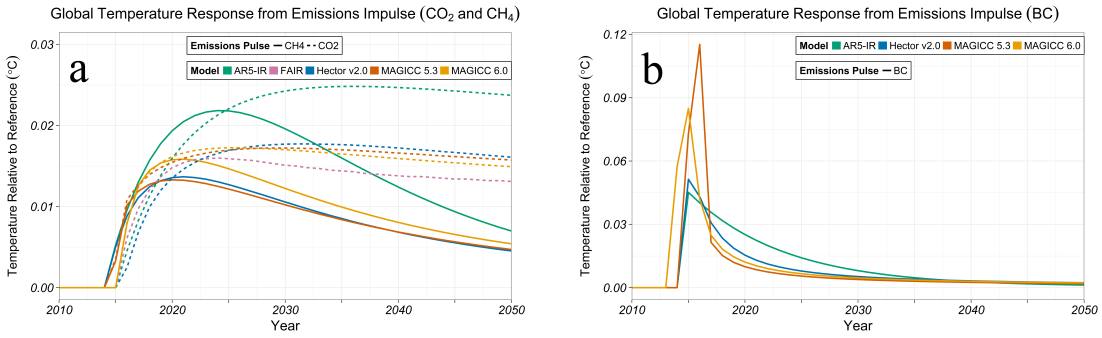


Fig. 2.5: Global mean temperature response from CO_2 and CH_4 emissions perturbations (a) and BC emissions perturbation (b) in SCMs(MAGICC 6.0 – yellow, MAGICC 5.3 BC–OC – red, Hector v2.0 – blue, AR5-IR – green, FAIR – pink).

insensitive to pulse size and background concentration (Millar et al. 2017), which results in a similar time-integrated airborne fraction compared to the Joos et al. multi-model mean at each time horizon. The comprehensive SCMs and to a lesser extent, FAIR, offer an improved response compared to AR5-IR (Millar et al. 2017).

We next consider model responses to methane (CH_4) emissions perturbations, a shorter-lived greenhouse gas with a dynamic atmospheric lifetime. The integrated forcing responses of Hector and MAGICC 5.3 are similar, as expected (see S9.3 of Schwarber et al. 2019). The MAGICC 6.0 integrated forcing response difference from the comprehensive SCM average is 9% larger 100 years after the pulse, however (Figure 2.2). As in the CO_2 emissions perturbations, AR5-IR has a much stronger response to a CH_4 emissions perturbation–22% larger 20 years after the pulse–with no meaningful increase 50 years after the pulse.

Finally, we look at the models' temperature responses to aerosols by perturbing black carbon (BC) forcing (Figure 2.5). The BC response increases quickly in both

MAGICC models compared to the other SCMs. Differences in these responses to a BC perturbation derive from model design. Both versions of MAGICC have differential and faster forcing responses over land, where most BC is located, compared to oceans, termed the geometrical effect (Meinshausen et al. 2011). This results in MAGICC responding faster than Hector v2.0, which does not differentiate forcing over land and ocean. Because AR5-IR represents the aerosol forcing as an exponential decay, the integrated temperature response is 20% stronger 20 years after the pulse compared to the other SCMs.

Due to the geometrical effect, we presume that the faster response in MAGICC is more realistic. However, models vary in the representations of aerosol effects. The greenhouse gas-like representation of aerosols in AR5-IR, for example, results in the unrealistically long response time scale found in this test. We do not explicitly conduct other aerosol perturbations (e.g., sulfate), but we would expect results showing similar responses.

BC has a unique set of atmospheric interactions as an absorbing aerosol, causing warming within the atmosphere, but potentially also surface cooling (Stjern et al. 2017; Yang et al. 2019). The response to a step change in BC emissions in two coupled model experiments has been found to have a flat long-term temperature response (Sand et al. 2016; Yang et al. 2019). In contrast, the comprehensive simple models continue to respond over a much longer time scale (Figure 2.6). This is an indication that SCM responses to BC, in particular, should be reevaluated.

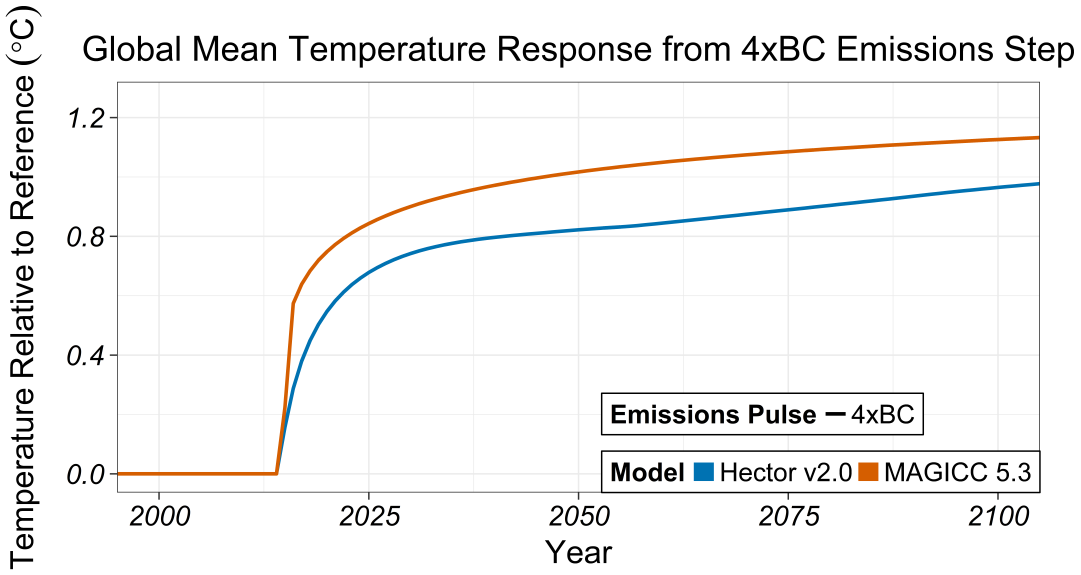


Fig. 2.6: Global mean temperature response from a 4xBC emissions step in the SCMs (MAGICC 5.3 BC-OC – red, Hector v2.0 – blue).

2.3.3 Responses to 4xCO_2 Concentration Step.

Finally, we compare our SCMs with complex models using the abrupt 4xCO_2 concentration experiment from Phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012). We find that Hector, MAGICC 5.3, and FAIR have initially quicker responses to an abrupt 4xCO_2 concentration increase compared to more complex models (Figure 2.7). The quicker responses are also reflected in their long term RWF, which is also larger than most of the complex models (see S10 in Schwarber et al. 2019). Compared to the other SCMs, AR5-IR has a faster response to an abrupt 4xCO_2 concentration increase and is consistent with the stronger response to a forcing impulse. Differences between the model responses to a finite pulse Figure 2.1 and a large concentration step Figure 2.7 demonstrates the

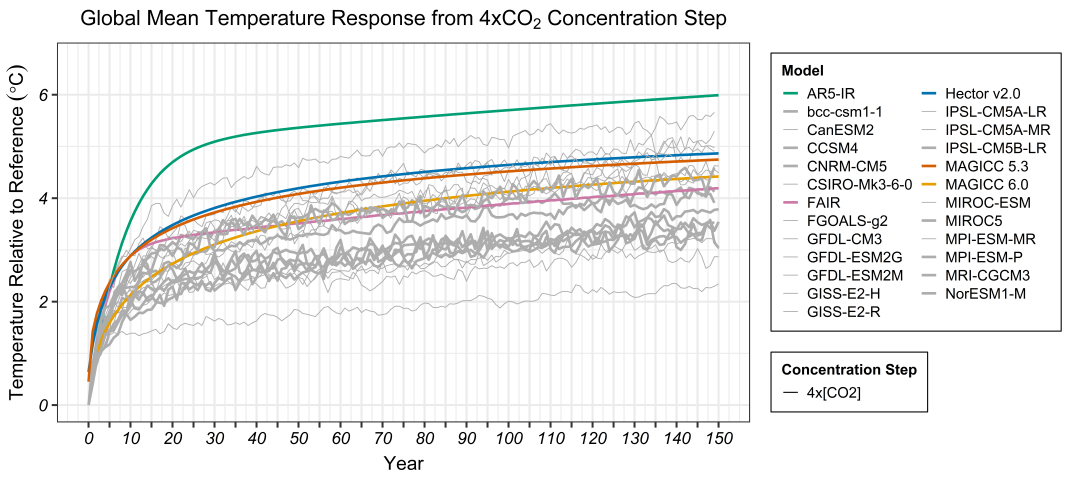


Fig. 2.7: Global mean temperature response from $4xCO_2$ concentration step in CMIP5 models (grey) and SCMs (MAGICC 6.0 – yellow, MAGICC 5.3 BC–OC – red, Hector v2.0 – blue, FAIR – pink, AR5-IR – green). A climate sensitivity value of 3 °C was used in the comprehensive SCMs, while in the idealized SCMs the parameter is not adjustable. The thick lines represent CMIP5 models with an ECS between 2.5 and 3.5 °C.

expected bias in AR5-IR under larger perturbations. The insensitivity of idealized SCMs to changing background concentrations will bias results if used under realistic future pathways (Millar et al. 2017).

Compared to the other comprehensive SCMs, MAGICC 6.0 initially responds more strongly under a CO_2 concentration impulse 2.1. In the non-linear abrupt $4xCO_2$ concentration regime, however, MAGICC 6.0 responds more slowly, similar to the complex model responses, especially in the first 20 years after the pulse. MAGICC 6.0 appears to respond more reasonably under stronger forcing conditions than the other SCMs.

2.4 Conclusions

The impulse response tests conducted here enable us to uncover differences in model behavior that are not apparent when running standard, multi-emission scenarios. Indeed, one of the important uses of SCMs is to conduct model experiments where there may be relatively small changes in emissions between two scenarios. Because SCMs do not exhibit internal variability, such experiments can be used to quantify such changes. Impulse response tests also allow us to understand, on a more fundamental level, differences between SCMs that have been found comparing simulations of more conventional scenarios (e.g., van Vuuren et al. 2011a).

By using fundamental impulse tests, we found that idealized SCMs using sums of exponentials often fail to capture the responses of more complex models. SCMs that include representations of non-linear processes, such as FAIR, show improved responses, though these models still do not perform as well as comprehensive SCMs with physically-based representations. Fundamental tests, such as a 4xCO_2 concentration step, show that most of the SCMs used here have a faster warming rate in this strong forcing regime compared to more complex models. However, comprehensive SCM responses are similar to more complex models under smaller, more realistic perturbations (Joos et al. 2013).

It is not possible to compare these fundamental responses with observations, and it is even more difficult to compare SCMs with the more complex models at decadal time horizons due to internal variability (e.g., Figure 2a from Joos et al.

2013). However, it is common in the climate modeling literature to use the multi-model mean as a base comparison. In fact, the CMIP5 multi-model mean has been shown to replicate observations better than any individual complex model (Flato et al. 2014).

Thus, we use the comprehensive SCM multi-model mean to compare to the individual model responses for most of our experiments. We also use the CMIP5 multi-model mean, developed using only those complex models with comparable climate sensitivity values to the SCMs, to compare the SCM responses from a 4xCO_2 forcing step. We posit that the responses closer to the multi-model mean are likely to more accurately represent that particular response pattern. We illustrate this assumption by reporting the time-integrated temperature response percent difference from the relevant multi-model mean in Table 2.4.

We note that the comprehensive SCM responses to a CO_2 concentration impulse are within 2% of the comprehensive SCM average, while the idealized SCMs, FAIR v1.0 and AR5-IR, have greater differences 100 years after the pulse.

Under the 4xCO_2 concentration step experiment, we can compare the SCM responses to more complex models from CMIP5. MAGICC 6.0 appears to respond more reasonably under stronger forcing conditions than the other SCMs 100 years after the pulse, though only marginally better than FAIR. Hector v2.0, MAGICC 5.3, and FAIR have initially quicker responses to an abrupt 4xCO_2 concentration increase compared to the ESMs. AR5-IR has too strong a response to an abrupt 4xCO_2 concentration increase and is insensitive to changing background concentrations.

For CH_4 emissions impulses, we use the difference from the comprehensive

SCM average to rate the responses. CH_4 is a well-mixed GHG and, therefore, we expect that the climate system response to CH_4 concentration perturbations will be similar to that for CO_2. However, it would be useful to evaluate in more complex models to determine if the simple representation of chemistry in the comprehensive SCMs adequately represents the time evolution of CH_4 concentrations in response to a change in emissions.

Finally, we do not have a definitive reference for the time-dependent response to BC forcing perturbations. Instead, we compare the SCMs using the difference from the average of both MAGICC models, which both differentiate aerosol forcing between land and ocean, resulting in a faster overall climate response to aerosols as compared to greenhouse gases (Sand et al. 2016; Shindell 2014; Yang et al. 2019). And in the case of BC, we note that the SCM responses should be taken critically because they do not accurately represent the temporal response to a BC step found in ESMs. A more definitive evaluation of climate system responses to aerosol perturbations would be useful. This would require additional complex model simulations of step emission changes for various aerosol species and/or forcing mechanisms. There are currently two studies that have conducted this test, one study specifically investigated NorESM's response to black carbon (BC) perturbations (Sand et al. 2016) and a more recent study that conducted similar BC perturbations in CESM (Yang et al. 2019).

There are numerous benefits to using simplified models, but the selection of the model should be rooted in a clear understanding of the model responses (see Table 2.4). Our work illustrates the necessity of using fundamental impulse tests

Species	Impulse	Time after pulse	Percent integrated temperature response differences for each simple climate model (%)				
			Hector v2.0	MAGICC 5.3	MAGICC 6.0	FAIR v1.0	AR5-IR
CO_2	4× forcing step	$H = 100$ yr	38 %	35 %	15 %	18 %	73 %
	Forcing impulse	$H = 100$ yr	1.0 %	−1.2 %	0.25 %	−16 %	23 %
	GHG emissions	$H = 100$ yr	−0.57 %	2.2 %	−1.6 %	−14 %	31 %
CH_4	GHG emissions	$H = 20$ yr	−3.1 %	−5.6 %	8.7 %	–	47 %
BC*	Aerosol emissions	$H = 20$ yr	−9.3 %	1.1 %	8.1 %	–	19 %

Tab. 2.4: Integrated temperature response differences. The values are the percent difference in time-integrated temperature response compared to the relevant reference (generally comprehensive SCM average; see Sect. S9). For BC specifically, we note that none of the SCMs reflect the temporal response for BC seen in two complex models (Sand et al. 2016; Yang et al. 2019).

to evaluate SCMs and we recommend that modeling communities adopt them as a standard validation suite for any SCM. Given that idealized SCMs are biased in their temporal responses, more comprehensive SCMs could be used for many applications without compromising on accessibility or computational requirements.

Chapter 3: Observations

3.1 Introduction

Historical surface temperature observations are often limited in spatial and temporal scale, but those gaps have largely been filled with the increase in computing resources and capabilities and with the onset of the satellite era, excepting paleoclimate reconstructions (Broccoli et al. 2001; Guo et al. 2015; Hartmann et al. 2013; Santer et al. 2011). Now, observations of the climate system include ground-based and remotely sensed datasets of surface climate variables (e.g., temperature (Morice et al. 2012; Tett et al. 1999), atmospheric composition (Hartmann et al. 2013), carbon and heat fluxes (Byrne et al. 2017; Byrne and O'Gorman 2013), teleconnection patterns (Graham 1994; Jones et al. 2001), etc.) over global and regional scales (Abram et al. 2016; Stott et al. 2010).

Numerous other efforts have been made to expand historical data back in time via paleoclimate reconstructions. For example, paleoclimate reconstructions produced by Neukom et al. (2014), Deser et al. (2012), and Friedmann et al. (2013) explore trends at inter-annual timescales (Deser et al. 2012; Friedman et al. 2013; Neukom et al. 2014), while those produced by Mann et al. (2007) and Ahmed et al. (2013) explore climate responses at multi-millennial time spans (Ahmed et al.

2013; Mann et al. 2007). The long time series produced through reconstruction efforts allow for comparisons of models with past climatic variations. For instance, reconstructions and simulations can also be used together to evaluate estimates of climate sensitivity to external forcings (e.g. Braconnot et al. 2012; Hegerl et al. 2006; Rohling et al. 2012).

Observational datasets, including those used in Chapters 4 and 5, are often used to test the fidelity of climate model results and improve climate model projections (Rahmstorf et al. 2007). Observations contain both the anthropogenic and natural influences on climate and maybe contain biases due to observational technique or post-processing, while simulations are parametric representations of the climate system driven by our understanding of the system. Thus, observations and simulations represent two different representations of the behavior of the actual climate system, and so comparisons often have large uncertainties (Ahmed et al. 2013).

Our work focusses on *in-situ* observations and paleoclimate reconstructions of surface temperature for regions that offer overlapping datasets in time and space to provide a complete view of the climate response (see Section 3.2). In Chapter 4 we compared variability in several observational datasets to the variability in complex models. In Chapter 5 we addressed changes in climate sensitivity over time and compare to trends in observations. Here, we provide more information on those observational datasets and some of the choices made in our comparisons.

3.2 Datasets

In this chapter, we use four well-known observational datasets to compare with variability in the CMIP5 models: HadCRUTv4.6 (Morice et al. 2012), BEST (Rohde et al. 2013b), NOAAGlobalTemp (Vose et al. 2012), GISTEMP (Hansen et al. 2010; see Table 3.1), which provides a more complete observational picture.

3.2.1 Historical datasets

In Chapter 4 and 5, we only use HadCRUT and GISTEMP because, for reasons that will become apparent below, we found that similarities in the observational records did not warrant repeating the analysis with all datasets (Figure 2.14 in Hartmann et al. 2013) and we aim to avoid the limitations of NOAAGlobalTemp and BEST. In addition, we use a multi-centennial reconstruction from seven continental-scale regions developed by the PAGES2k Consortium (Ahmed et al. 2013).

Hadley Centre Global Temperature (HadCRUTv4.6). The HadCRUTv4.6 (called HadCRUT from hereafter) dataset, produced by the Met Office Hadley Centre and the Climatic Research Unit at the University of East Anglia, provides gridded, global historical surface temperature anomalies relative to a 1961-1990 reference period. The gridded data are a blend of land-surface air temperature data from CRUTEM4 and sea-surface temperature data from HadSST3, and is available at a monthly temporal resolution since January 1850 on a 5x5 degree latitude-longitude grid. The dataset is presented as an ensemble of 100 realizations produced by sampling the uncertainty in non-climatic factors affecting

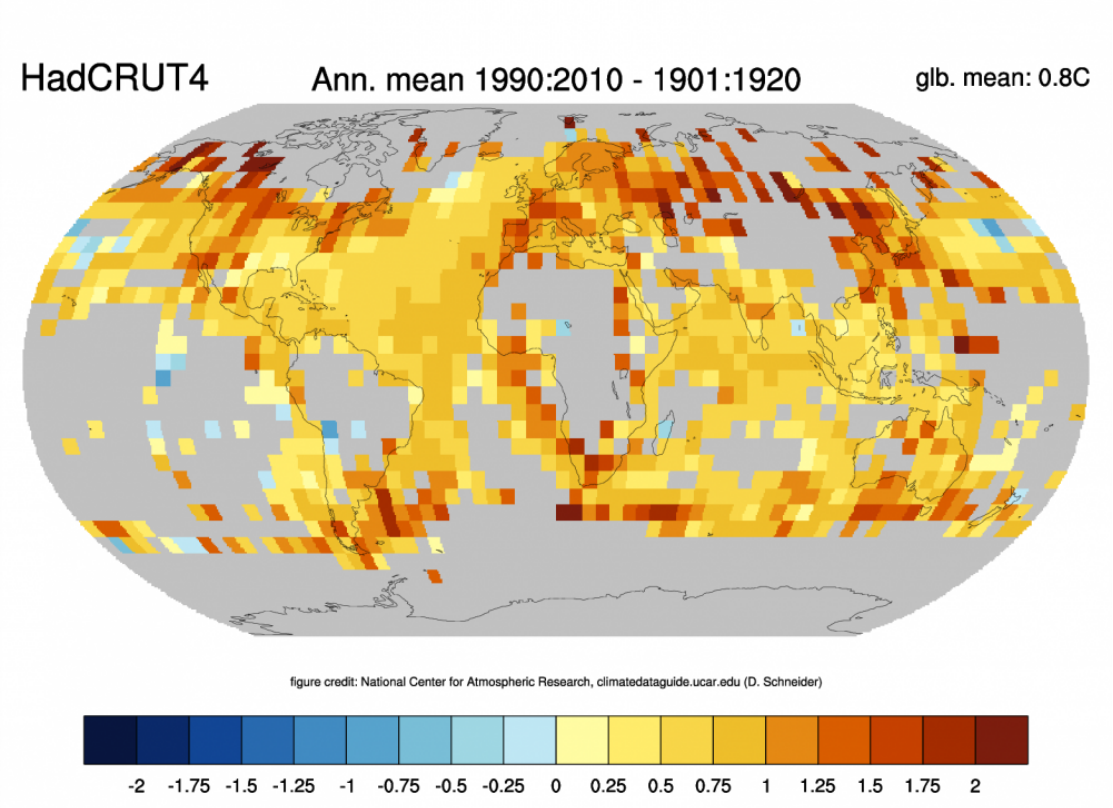


Fig. 3.1: Annual mean temperature change between 1901-1920 and 1991-2010 from HadCRUT. Credit: ClimateDataGuide, NCAR

near-surface temperature observations, such as assumptions about components of the error (see discussion in (Morice et al. 2012). Though all 100 realizations are available for use, our analyses rely on the ensemble median, which is produced for each grid box for each time step from the 100 members.

NOAAGlobalTemp (aka MLOST). The National Oceanic and Atmospheric Administration's (NOAA) Merged Land Ocean Global Surface Temperature Analysis Dataset (NOAAGlobalTemp), formally known as the Merged Land-Ocean Surface Temperature Analysis (MLOST), combines Global Historical

Climatology Network, Monthly (GHCN-M) version 3 land surface air temperatures with ERSSTv3b SSTs, providing a blended global surface temperature dataset (Vose et al. 2012). Data are provided at a monthly resolution from 1880 to the present on a 5x5 degree latitude-longitude grid. Station records are interpolated to provide broad spatial coverage, and the land and ocean domains are treated separately. Using empirical orthogonal functions, the high and low-frequency variations in each domain are reconstructed separately, then combined and blended with the original records. In addition, care is taken to handle data-sparse areas via masking to reduce the over expression of highly smoothed reconstructions because of their sampling frequency. In general, the NOAAGlobalTemp dataset has poor coverage in the polar regions because sea ice values are set to missing (NCAR 2014).

GISS Surface Temperature Analysis (GISTEMPv4). NASA Goddard's Global Surface Temperature Analysis (GISTEMPv4; hereafter GISTEMP) is a global surface temperature data set produced by blending land surface air temperatures primarily from meteorological stations operated through the NOAA GHCN v4 dataset, and SSTs from the ERSST v5 dataset (Hansen et al. 2010). The gridded data span 1880 to the present at a monthly resolution, on a 2x2 degree latitude-longitude grid. GISTEMP has greater polar coverage than MLOST or HadCRUT, in part because it includes Antarctic 'READER' station data, in addition to the interpolation method used, which relies on linear inverse distance weighting to infill gridboxes with records from stations up to 1200 km away (NCAR 2014). Another variation of the data provides data up to a 250 km limit, which we

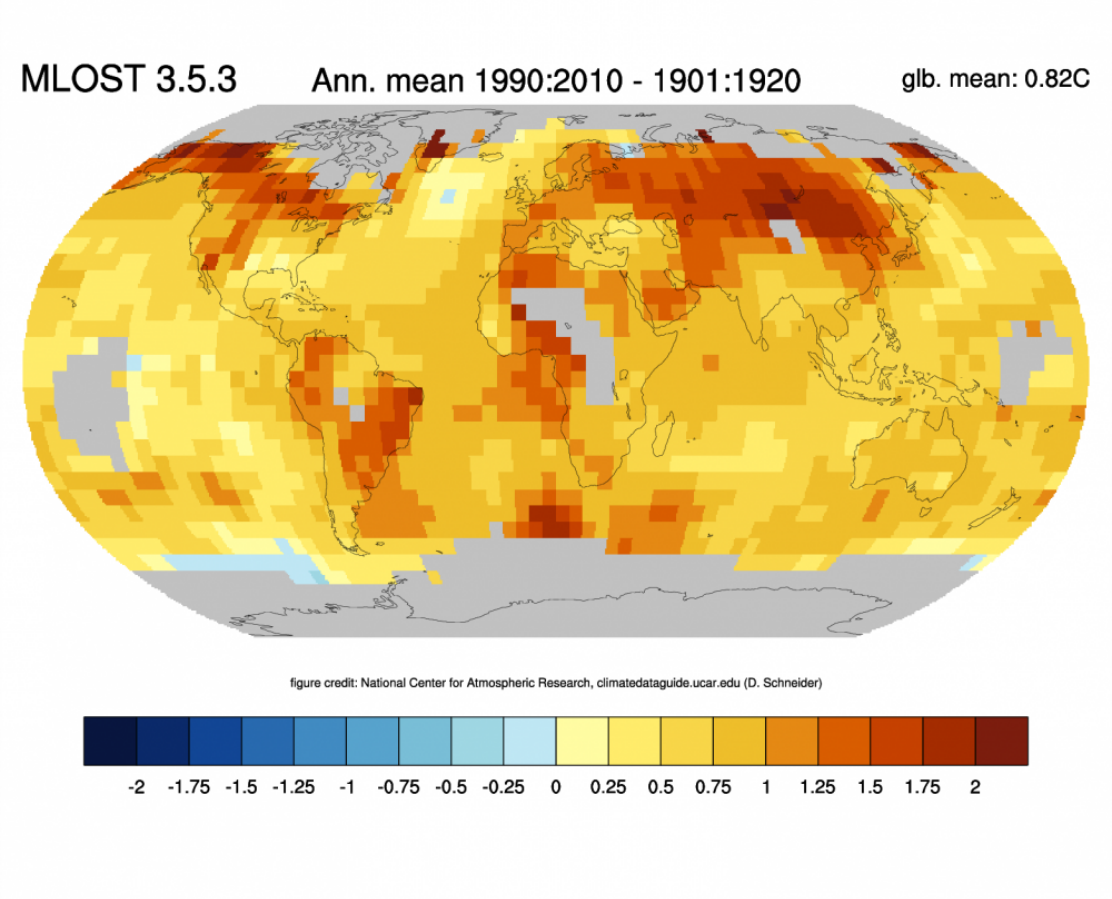


Fig. 3.2: Annual mean temperature change between 1901-1920 and 1991-2010 from NOAA MLOST. Credit: ClimateDataGuide, NCAR

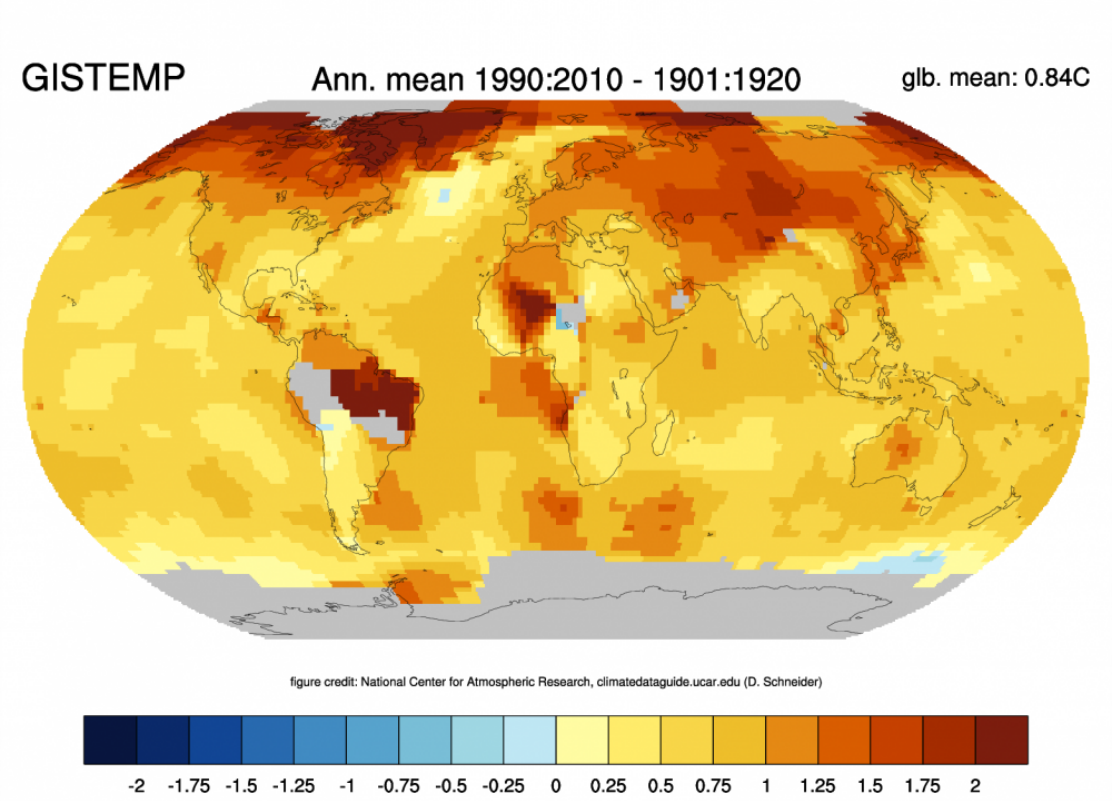


Fig. 3.3: Annual mean land temperature change between 1901-1920 and 1991-2010 from GISTEMP (with ERSSTv3b SSTs). Credit: ClimateDataGuide, NCAR

use in our analyses.

Berkeley Earth Surface Temperatures (BEST). Among other data products, the Berkeley Earth Surface Temperatures (BEST) provides a gridded reconstruction of land surface air temperature records from 1701 to present, which we used in our analyses. The Berkeley Earth project was developed to provide an alternative, independent assessment of global temperature change, incorporating approximately 39,000 records, in comparison to the other datasets listed above which rely on roughly 5000-7000 land stations. The large number of records BEST is able

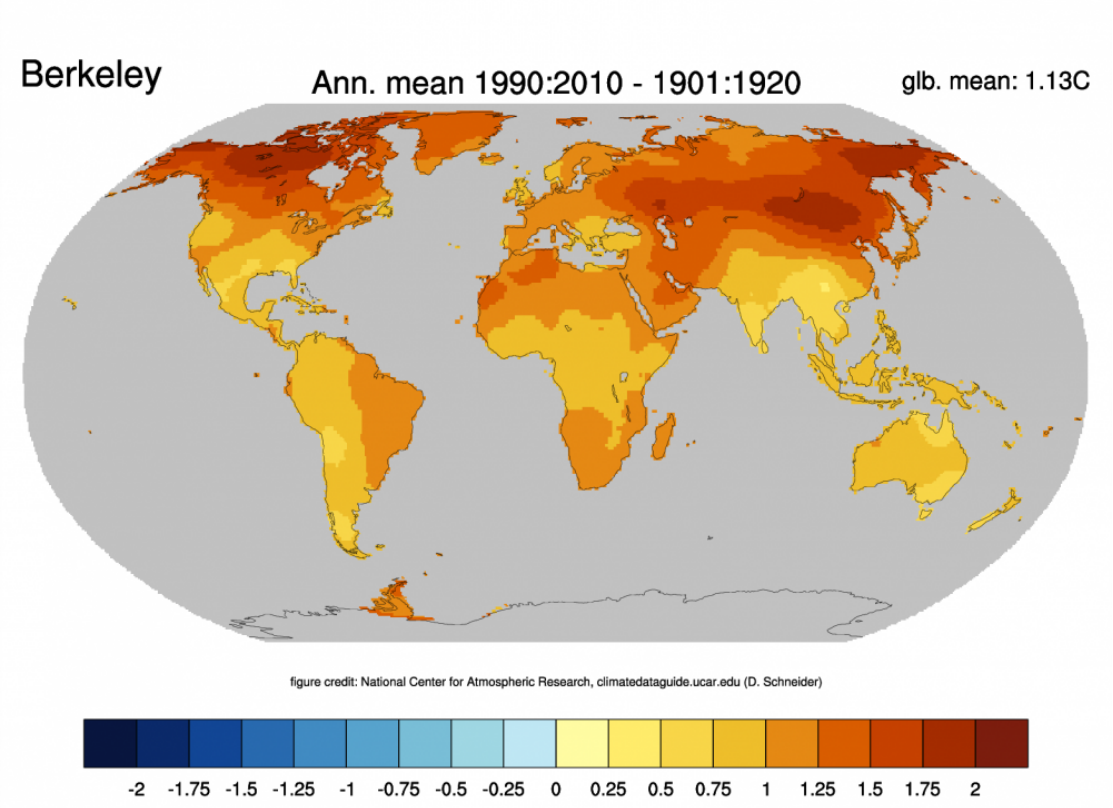


Fig. 3.4: Annual mean land temperature change between 1901-1920 and 1991-2010 from Berkeley Earth. Credit: ClimateDataGuide, NCAR

to use beyond GHCN derives from its use of a statistical model that allows short, fragmented timeseries to be incorporated into the complete dataset. The Berkeley group offers several products: land records available at higher spatial and temporal resolutions than the main reconstruction, one merged land-ocean dataset that infers air temperatures over sea-ice-covered regions, and an alternative merged land-ocean dataset that uses the nearest available SST observations (Rohde et al. 2013a). In this chapter, we use the main reconstruction.

General advantages and disadvantages. Each dataset has advantages and

disadvantages. Using all four datasets will provide a more complete observational picture. NOAA MLOST has nearly global coverage, but only provides anomalies. The HadCRUT dataset separates land and ocean data sets, with less analysis and smoothing/interpolation than other data sets for land and ocean domains. However, HadCRUT has less spatial coverage, resulting in more gaps as compared with in-filled datasets like GISTEMP or MLOST. Still, GISTEMP has limited coverage of polar regions. Finally, BEST uses a large number of stations with long temporal lengths, but uses a Kriging procedure to produce highly smoothed data (NCAR 2014).

3.2.2 PAGES2k Paleoclimate Reconstruction

The PAGES2k dataset provides a multi-centennial temperature reconstruction from seven continental-scale regions consisting of 511 time series of tree rings, pollen, corals, lake and marine sediments, glacier ice, speleothems and historical documents. The PAGES2k Consortium, the international collaborative that developed the dataset, allowed each of the international groups assigned to a given region to identify the proxy climate records "best suited" for reconstructing annual or warm-season temperature within their region based on *apriori* criteria (see Supplementary Database S1 of Ahmed et al. 2013). The temperature reconstructions over Arctic, Europe and Antarctica span approximately 2,000 years, Asia, South America and Australasia temperature reconstructions span the past 1,000-1,200 years, and North America consists to two datasets: a shorter decadally resolved tree-ring-based reconstruction back to 1200 and a longer 30-year-resolved pollen-based

reconstruction back to 360 (Ahmed et al. 2013). The temperature reconstructions are provided at an annual resolution, excepting North America, where we use the 30-year-resolved pollen-based reconstruction throughout our analyses because there is evidence of spectral biases in tree-ring proxy data (Franke et al. 2013). For all time series we limit the temperature-time series to data before 1850 to allow for a consistent comparison with model data in Chapter 4, thus analyzing our temperature-time series over the same temporal scale.

3.3 Processing Choices

To process the model and observational data we use the NCAR command operators, or CDO commands, in R (CDO 2018: Climate Data Operators. Available at: $http://www.mpimet.mpg.de/cdo$). Both model and observational data were remapped from the source grid (see Table 3.1) to a 360x180 longitude-latitude grid using a first-order conservation remapping tool (Jones 1998). Remapping or regridding between spherical grids is essentially required for earth-system modeling (Hill et al. 2004) and for this, CDO employs the Spherical Coordinate Remapping and Interpolation Package (SCRIP).

From Jones (1998), first-order conservative remapping can be explained by considering the remapping of a flux f on a source grid cell k to a destination grid F given by:

$$\bar{F}_k = \frac{1}{A_k} \sum_{n=1}^{N} \iint_{A_{nk}} f_n dA \tag{3.1}$$

Abbreviated Name	Long Name	Source	Period of Record	Available Timestep	Available Resolution	Input Data
GISTEMP	NASA Goddard Institute for Space Studies (GISS) Surface Temperature Analysis	Hansen et al. 2010	1880/01 to 2016/11	Monthly	2x2 degree	GHCN v3 (~6300 land stations); ERSSTv3b (SSTs); SCAR READER (for Antarctic stations)
HadCRUT4	HadCRUT4 and CRUTEM4	Morice et al. 2012	1850/01 to 2017/06	Climatology, Monthly	5x5 degrees	~5500 land station records from GHCN and other sources combined with HadSST3
NOAAGlobal Temp (aka MLOST)	NOAA Merged Land Ocean Global Surface Temperature Analysis Dataset	Vose et al. 2012	1871/01 to 2016/11	Monthly	5x5 degrees	GHCN v3.3 (~7000 land stations); ERSSTv3b (SSTs)
BEST	Berkeley Earth Surface Temperatures	Rohde et al. 2013	1701/01 to 2013/03	Climatology, Daily, Monthly	1x1 degree or equal-area grid, .25 degree available for USA and Europe	combination of 14 databases of station data; ~37,000 individual records used; merged land-ocean version uses HadSST3
PAGES2K	'2K Network' of the IGBP Past Global Change (PAGES) project	Ahmed, M et al. 2013	Varies for region	Annual, except Northern Hemisphere	Point source	692 records from 648 locations, with data sources covering all continents and oceans

Tab. 3.1: Summary of Global Surface Temperature Datasets (NCAR 2014).

where $\overline{F}$ is the area-averaged flux and A_k is the area of a cell k (Jones 1998). Unlike the case above, which calculates $\bar{F}_k$ using only those cell grids which are at least partially covered by the destination grid, one can calculate $\bar{F}_k$ when the cell k overlaps N cells on the source grid as:

$$\bar{F}_k = \frac{1}{A_k} \iint_{A_{nk}} f dA \tag{3.2}$$

Assuming the flux is constant across a source grid cell, f_n, the previous equation would lead to the first-order area-weighted scheme.

The model data are also aggregated from the original time steps to monthly averages using the "fldmean" CDO command and the monthly observational data was pulled from its source. Further, both model and observational data were processed to correspond to the PAGES2k temperature reconstructions (Bothe et al. 2013a) for each region (Table 3.2) by averaging over consistent temporal and area-weighted spatial scales, unless otherwise noted.

Each region was also masked using three different partitions: land-only, ocean-only, and land-and-ocean surface temperature. For the models, the masking was carried out using the corresponding fractional land masking in the "areacella" file from CMIP5, which provides information on the area of the grid for a given model. For the observations, we used the area file provided with the individual dataset.

For the remainder of this chapter, we further investigate the impacts of our methodological choices described above.

Region	Time scale	Min lat	Max lat	Min lon	Max lon
Antarctica	Annual	-90	-60	0	360
Arctic	Annual	60	90	0	360
Asia	JJA	23.5	55	60	160
Australasia	SONDJF	-55	0	110	180
Europe	JJA	35	70	40	350
North America	Annual	30	55	230	285
South America	DJF	-65	-20	285	330

Tab. 3.2: PAGES2k region definitions from Goosse 2015.

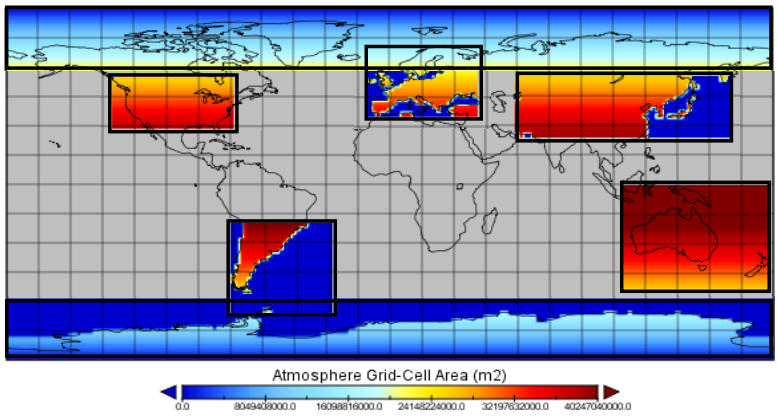


Fig. 3.5: Graphical representation of the PAGES2k regions with masking (Goosse 2015).

3.3.1 Impact of regridding method

When processing the historical observational datasets to match the PAGES2k regions, as described above, we remap the original grid to a 180x360 grid using the first-order remapping tool. However, there are numerous grid interpolation methods and the appropriate method should be selected based on the intended task. Here, we compare the temperature-time series results from the four observational datasets produced using the first-order conservative remapping tool, compared to three others: second-order remapping (Jones 1998), bilinear remapping (Press et al. 1992), and bicubic remapping (Press et al. 1992).

For second-order remapping, Jones (1998) found that remapping from a fine grid to a coarse grid (essentially an averaging) has "virtually no advantage to using the second-order method in such cases" and we see this same result in Figure 3.6 (Jones 1998).

Bilinear remapping is commonly employed when processing climate model output, but the remapping tool is applicable for logically rectangular grids only (Jones 1998), though a newly developed method allows bilinear remapping to be used on irregular grids (Kim et al. 2019). Bicubic remapping suffers from this same limitation. As a result, we use first-order conservative remapping as our base method in the main chapters.

In Figure 3.6, as an example, we report the difference in the temperature-time series found using a second-order remapping, bilinear, or bicubic remapping versus the first-order remapping for each historical observational dataset over Australasia,

one of the seven continental-scale regions defined by PAGES2k. As expected, the difference between the two conservative mapping techniques is small. The difference between the bilinear and bicubic techniques and the standard remapping are within $\pm 0.5^\circ$ C, with the largest difference from the HadCRUT dataset.

For other regions the difference between the two conservative mapping techniques is also small. Only over North America for NOAAGlobalTemp is the difference noticeably larger at a max difference of $\pm 0.0125^\circ$ C. In addition, the differences between the bilinear and bicubic techniques and the standard remapping are within $\pm 0.5^\circ$ C, with the exception of the polar regions. In the polar regions we see larger differences of $\pm 2^\circ$ C for datasets with poor polar spatiotemporal coverage, such as HadCRUT in the Arctic region.

In general, the regridding technique has minimal impact on the temperature time-series, especially compared to other factor impacting data reliability, such as spatial coverage or infilling technique.

3.4 Spectral Density Estimates Using Base Method

In Chapter 4, we compare variability in several complex climate to two historical observational datasets, HadCRUT and GISTEMP, in frequency space using estimates of spectral density found using the multitaper method (see Chapter 4 for methods; Riedel and Sidorenko 1995; Thomson 1982; Walden et al. 1998).

In Figure 3.7, we show the spectral densities for all observational datasets listed above using the multitaper spectral estimation technique. We see regionally

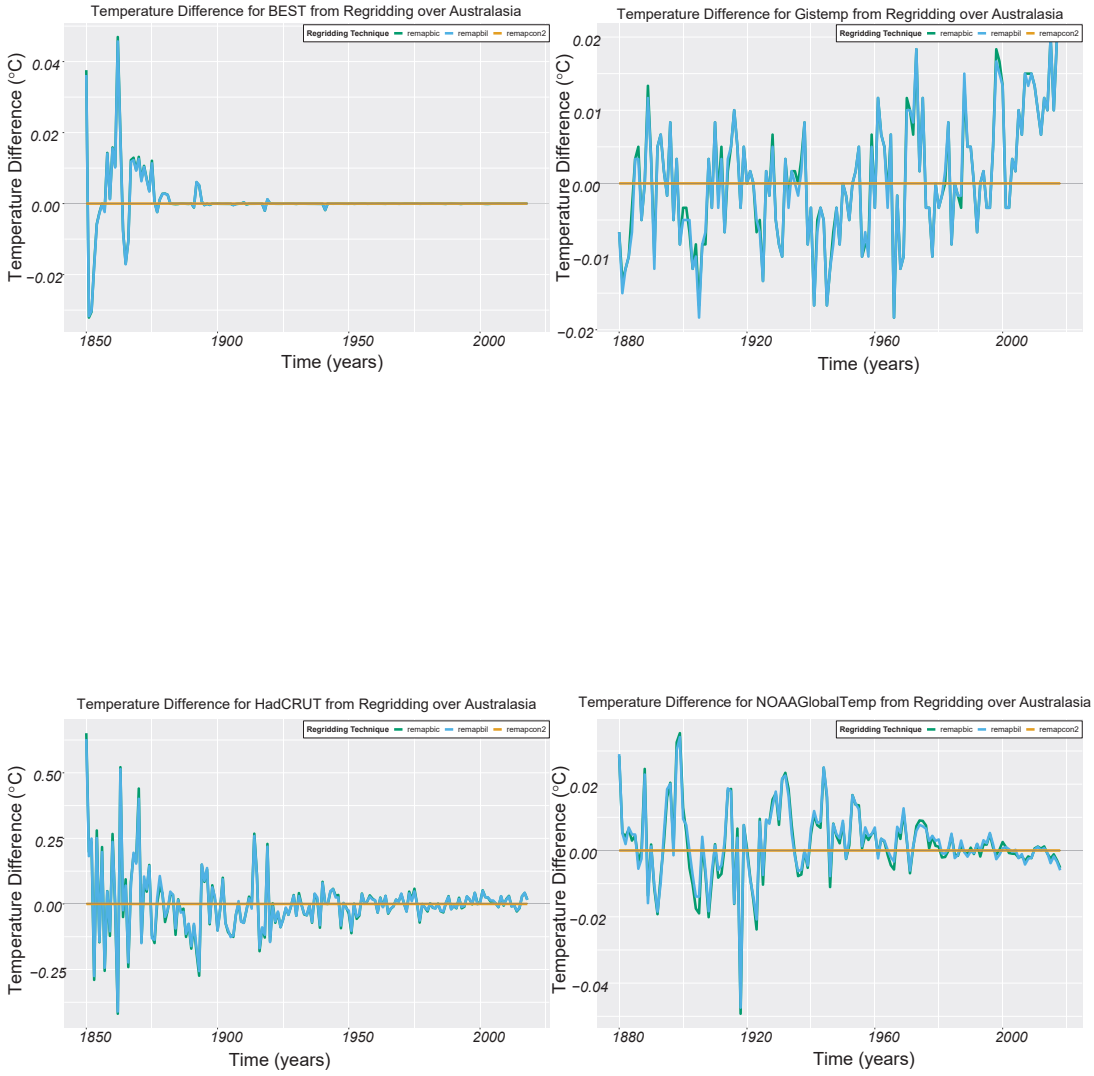


Fig. 3.6: Comparison of temperature-time series produced using various regridding techniques for each in-situ dataset over Australasia.

specific differences between the spectral densities of the PAGES2k reconstruction and the historical datasets (see discussion in Bothe et al. 2013a). Additionally, we see similarities in the spectral densities across the historical datasets, though there are some datasets with lower spectral densities in certain regions including Antarctica, Europe, and South America. Below, we discuss some of the interesting regional features.

Taking Antarctica as an example, we see that BEST has a smaller spectral density by a factor of 100 (note the log scale) at long time scales compared to HadCRUT and GISTEMP. The difference is likely a result of the BEST dataset providing surface temperature anomalies over the land surface for a limited area of the Antarctic region compared to other infilled datasets, such as GISTEMP. We also note that NOAAGlobalTemp does not provide surface temperature results over Antarctica. In addition, PAGES2k has a lower spectral density across all time scales compared to HadCRUT and GISTEMP.

By contrast the datasets have a strong coherence in North America where temperature trends have been well documented, especially in the satellite era. Some studies indicate, however, that land datasets may underestimate warming in northern areas (Wang et al. 2017; Way et al. 2017), though we do not see this here with BEST. We also note that PAGES2k is highly smoothed in this region because, unlike the other datasets, paleoclimate reconstructions over North America are limited. We use the pollen record reconstruction provided from PAGES2k at a 30-year resolution as mentioned above.

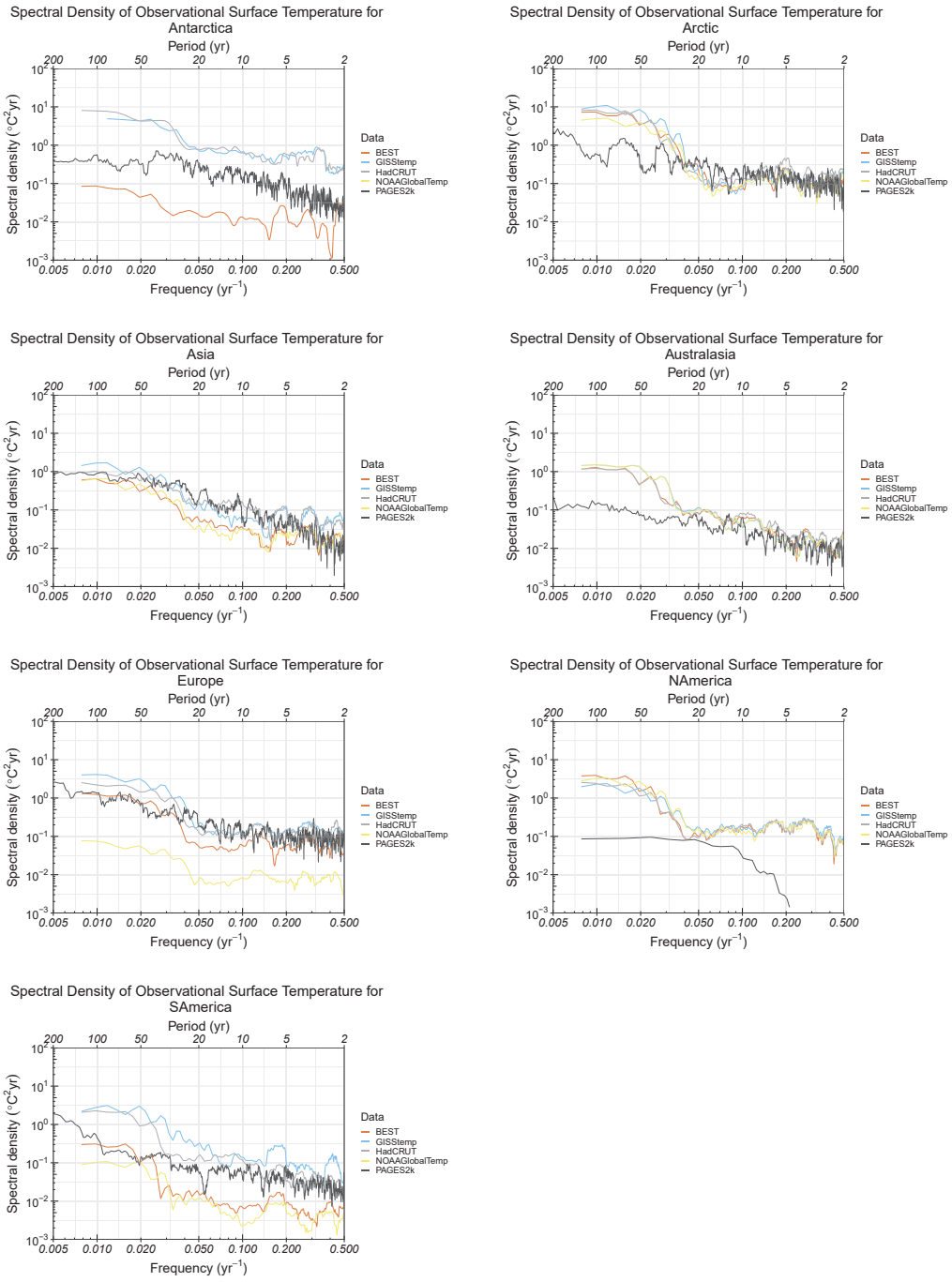


Fig. 3.7: Spectral density estimates of observational data over the PAGES2k regions.

Region	Size	Min lat	Max lat	Min lon	Max lon
Asia	Large-land (Lland)	25	50	80	120
	Smaller-land (Sland)	30	50	90	115
	Large-landocean (Llandocean)	25	50	100	140
	Smaller-landocean (Slandocean)	30	50	110	135
North America	Large-land (Lland)	30	55	240	280
	Smaller-land (Sland)	35	50	245	270
	Large-landocean (Llandocean)	30	55	220	260
	Smaller-landocean (Slandocean)	35	50	225	250

Tab. 3.3: Region definitions for sensitivity analysis (Smith et al. 2015).

3.4.1 Region size sensitivity tests

We next perform a test to examine how the size of the regions impacts the spectral density results for two select regions, Asia and North America. Details on the region bounds used in our main analysis are available in Table 3.2. We expand upon these regions to test if doubling the regional area, or shifting a region to include more ocean area consistently impacts the variability of the observations across time scales. Table 3.3 describes the region definitions for this sensitivity analysis.

We examine the variability for two main historical datasets, HadCRUT and GISTEMP, for each region in Table 3 at each spectral bin of interest also used

in our main analyses (see Figure 3.9). Overall, the impact of the shifting regions is smaller than the differences in the variability differences across the regions (see Table 7.2 in Appendix 7). A separate study from Smith et al. (2015) similarly found that region selection had a limited impact on their results, though the authors were studying temperature trends in the time domain. In addition, we note that we did not explicitly examine the impacts of shifting regions on PAGES2k data, though such domain changes can have differences due to the mix of proxies used in the regions as some types respond differently to temperature changes, in addition to the spatial and temporal sampling differences. Thus, even though we see some shift in the spectral density over the different land-ocean partitions and region sizes, more fundamental regional differences can account for larger portion of the differences between the regions.

3.5 Conclusions

Here, we summarize the choice to use the HadCRUT and GISTEMP data, while excluding others in subsequent analyses. We note that the HadCRUT dataset separates land and ocean data sets, with less analysis and smoothing/interpolation than comparable data sets for land and ocean domains. HadCRUT also has less spatial coverage, resulting in more gaps as compared with in-filled datasets like GISTEMP. GISTEMP has limited coverage of polar regions. Other datasets are also explored, but because some datasets have lower spectral density over longer time periods that we consider to be robust over certain regions (see Chapter 4), we

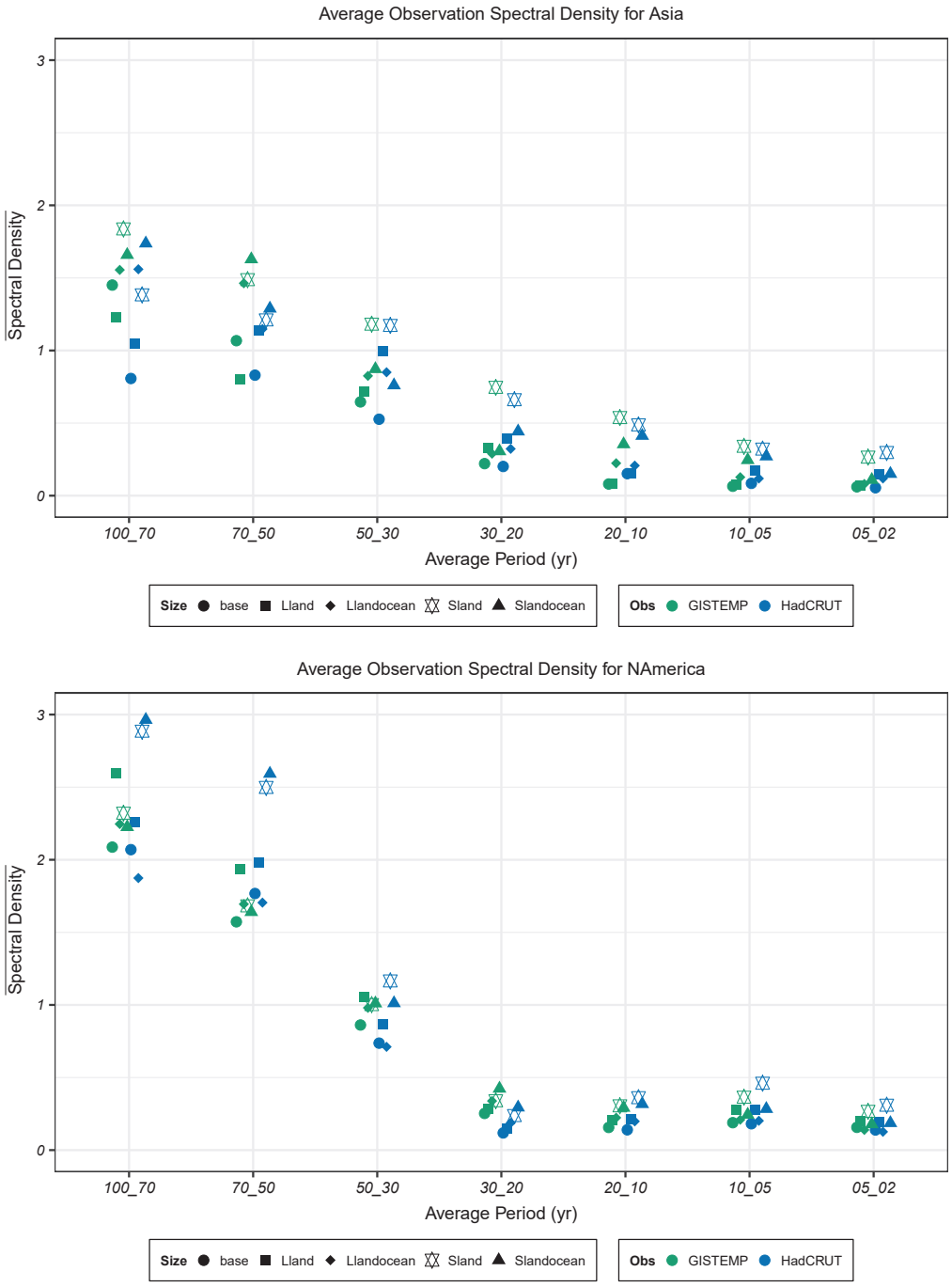


Fig. 3.8: Average spectral density for GISTEMP and HadCRUT over Asia (top) and North America (bottom) with changing region sizes.

do not include all datasets in subsequent analyses.

Chapter 4: Characterization of Unforced Model Variability in CMIP5 (in prep for submission to Journal of Geophysical Research - Atmospheres)

4.1 Introduction

Contributions from both unforced variability (e.g., ocean-atmosphere interactions) and external forcing (e.g., anthropogenic greenhouse gas emissions, volcanic activity, changes in solar intensity) drive changes in Earth's global mean surface temperature (GMST; Hawkins and Sutton 2009). Understanding the relative contributions from these different components has been of great interest to the scientific community and society, more generally.

In the context of climate change, it is important to decipher the forcing contributions from anthropogenic sources, and those from "natural" sources. Detection and attribution studies, for example, often investigate important time periods (e.g., Anthropocene) and regions (e.g., Northern Hemisphere, Arctic) in an effort to identify the magnitude of human influence on the climate (e.g., Santer et al. 2013 and citations there-in). Yet other studies consider the time at which climate change 'signal' emerges from the 'noise' of natural variability, termed the time of

emergence, for a number of key variables (see Giorgi and Bi 2009; Hawkins and Sutton 2012; Sui et al. 2014). We adopt the convention common in such literature and refer to external forcings as the signal and unforced variability as noise (Brown et al. 2015).

Many studies have also sought to disentangle the internal climate mechanisms that also influence GMST trends (e.g., Cai et al. 2015; Canty et al. 2013; Easterling and Wehner 2009; England et al. 2014; Hope et al. 2017; Kosaka et al. 2013; Meehl et al. 2007; Perry et al. 2017; Stocker et al. 2013). Internal mechanisms can vary on seasonal to multi-decadal to centennial timescales, for example oscillations with large extents such as the El Niño–Southern Oscillation, Atlantic Meridional Overturning Circulation, Asian Monsoon, or North Atlantic Oscillation, etc.

Often, climate models are used in these analyses to assess our ability to predict future climate change, for example. However, a single run from a complex climate model, a tool commonly used to understand changes in the climate system (e.g., temperature trends, teleconnection patterns), contains both the forced signal and unforced noise, however. Isolating the components can be accomplished by averaging over several independent realizations (i.e. different initial conditions) of the forced and unforced model runs. Several studies (e.g., Bothe et al. 2013a,b; Coats et al. 2015a,b; Covey et al. 2003; Furrer et al. 2007a,b; Giorgi and Mearns 2002; Goosse 2015; Hargreaves et al. 2013; Knutti et al. 2008; Meehl et al. 1997; Neukom et al. 2014; Santer et al. 2011, 2007; Tebaldi and Knutti 2007) employed this method and used multi-model ensembles to determine the statistical significance of the signal versus the noise.

There are limited studies, however, that have explicitly characterized the model-specific unforced variability. One recent study by Libardoni et al. (2019) found that estimating internal variability from a single model at the global scale produces distributions of important climate indicators (i.e., climate sensitivity) that "do not fully represent the uncertainty in the climate system." However, this analysis does not utilize spectral density to characterize the unforced variability over all time scales.

Spectral analysis in frequency space, which can provide information on climate response or natural and forced perturbations over all time scales, is frequently used in climate science studies to understand modes of unforced variability, such as ENSO (Feldstein 2000), or in analyzing the coherence of paleoclimate data with observational data (Ahmed et al. 2013; Franke et al. 2013; Fyfe et al. 2013).

Fyfe et al. (2013) estimates spectral densities of paleoclimate data over the Arctic in comparison to HadCRUT and climate model data to qualitatively assess the similarities between the reconstructions and simulations (Fyfe et al. 2013). Goosse et al. (2015) goes further and utilized correlations, which can be a problematic technique (Hu et al. 2017), to determine the coherence of historical datasets with the PAGES2k reconstruction (Goosse 2015). Here, we go further and compare the historical data, reconstruction, and model data across all timescales using a robust assessment of model variability to inform the time scales we can consider to be reliable.

In general, observational and reconstructed temperature datasets are often used to test the fidelity of model results to past and current climate conditions (e.g.,

Haustein et al. 2019). Comparing observations and simulations is difficult because these represent two different representations of the behavior of the actual climate system, creating more uncertainties (Bothe et al. 2013a). Observational datasets provide insights into perturbations of our climate over the last century, with higher resolution particularly available after the satellite era (> 1970s), though none of the datasets used in this dissertation rely on satellite data. Many satellite- and in-situ-based temperature datasets, however, are limited by their temporal and spatial coverage, especially during times of war or in areas of interest, but with low human population (e.g., Arctic).

Paleoclimate reconstructions overcome some of these limitations (and add a few others) by providing millennial scale temperature time-series. For example, comparing observations and simulations is difficult because these represent two different representations of the behavior of the actual climate system, creating more uncertainties (Bothe et al. 2013a). Additionally, some techniques to compare these disparate datasets must be used with caution (Hu et al. 2017).

The long time series, however, allows for comparisons of models with past climatic variations. For example, reconstructions and simulations can also be used together to evaluate estimates of climate sensitivity to external radiative forcing (e.g. Braconnot et al. 2012; Hegerl et al. 2006; Rohling et al. 2012).

In this study, we draw from the PAGES2k Consortium as our source of the paleoclimate data in addition to uses the continentia-scale regions it defines. Looking at the PAGES2k data, studies have postulated that the regional differences in the evolution of temperature at multi-decadal to centennial time scales could result from

major modes of natural atmospheric variability (Ahmed et al. 2013; Fernández-Donado et al. 2013; Goosse et al. 2012). In addition, it has been noted that the simulated magnitude of regional changes is often not as large as the observed magnitude (Braconnot et al. 2012). This may be because models underestimate internal variability or do not have the correct magnitude of repones to external forcing, or it could be caused by the absence of or incorrect representation of key feedback mechanisms (Braconnot et al. 2012; Valdes 2011). These past studies provide an important backdrop to our work, providing support for some regional-specific conclusions.

In our work, we utilize the full suite of climate data including models, observations, and reconstructions to characterize variability in the models. Understanding unforced variability is crucial to extracting the forced component in these disparate datasets (Bothe et al. 2013a). Though many studies draw conclusions using estimate of, or assumptions about, unforced variability, a robust characterization of unforced model variability has not been attempted in the literature. We fill this gap by determining how realistic unforced model variability in complex models is compared to observations for regions and time periods important to human systems.

4.2 Data and Methods

We use model output from several complex model experiments from the CMIP5, a commonly cited dataset: PiControl (pre-industrial control, run for

hundreds of years), past1000 (paleoclimate run from 850 to 1850), and historical (historical simulation run for hundreds of years from ∼1850 to 2012) (Taylor et al. 2012). Our analysis focusses on those models with sufficiently-long PiControl series and past1000 simulations available: CCSM4, CSIRO-Mk3L-1-2, HadCM3, IPSL-CM5A-LR, MPI-ESM, MIROC-ESM, MRI-CGCM3 (see Table 4.1).

Raw output from the selected models was obtained from the CMIP5 data archive $http : //cmip - pcmdi.llnl.gov/cmip5/data_portal.html$) and the World Data Center for Climate site $http : //cera - www.dkrz.de/WDCC/ui/Index.jsp$. Monthly temperature data are aggregated to the global annual mean level using code developed using CDOs (see CDO 2018: Climate Data Operators. Available at $http : //www.mpimet.mpg.de/cdo$). The long-term drift is removed from the CMIP5 model data by subtracting the linear trend from the corresponding pre-industrial control run (Gupta et al. 2013) prior to conducting subsequent analyses.

Additionally, we use two well-known observational datasets that provide surface temperature anomalies for the recent historical period: HadCRUTv4.6 and GISTEMP to understand the unforced variability in the CMIP5 models (Sutton et al. 2007; see Chapter 3). Each dataset has advantages and disadvantages. The HadCRUT dataset separates land and ocean data sets, with less analysis and smoothing/interpolation than comparable data sets for land and ocean domains. HadCRUT has less spatial coverage, resulting in more gaps as compared with in-filled datasets like GISTEMP. GISTEMP has limited coverage of polar regions. Other datasets are also explored in Chapter 3, but we find significant problems with some of the datasets in the context of the spectral analysis we conduct here.

Centre(s)	Model name	PiControl	Historical	Past1000
National Center for Atmospheric Research USA	CCSM4	1051	155	1001
Commonwealth Scientific and Industrial Research Organization/Queensland Climate Change Centre of Excellence Australia	CSIRO-Mk3L-1-2	1000	149	1000
Met Office Hadley Centre UK	HadCM3	1200	146	1001
Institut Pierre Simon Laplace France	IPSL-CM5A-LR	1000	155	1001
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology Japan	MIROC-ESM	630	155	1000
Max Planck Institute for Meteorology Germany	MPI-ESM-P	1155	155	1000
Meteorological Research Institute Japan	MRI-CGCM3	500	155	1001

Tab. 4.1: CMIP5 data and experiment description for the models used in Chapter 4. The table includes the centre that produced the model, the model's names, and the length of the data available for the PiControl, Historical, and past1000 experiments. More information on the model specifications is available in Table 7.1 Appendix 7.

Robust statistical analysis of climate trends requires a long time series of data. Despite techniques to fill observational records, the temporally limited records pose a challenge in understanding the forces driving variability. To account for this limitation, several efforts have been made to expand the historical data available via paleoclimate reconstructions. Here, we use PAGES2k (Ahmed et al. 2013), a multi-centennial reconstruction from seven continents, to add statistical robustness to our analysis otherwise limited by datasets with less temporal coverage. Short-term climate dynamics, however, are not easily resolved from the PAGES2k dataset. It had been suggested, for example, that these data might be most reliable for trends longer than 40 years (Smith et al. 2015).

We select analysis regions matching PAGES2k: South America, North America, Europe, Arctic, Antarctica, Australasia, and Asia (see Table 3.2). Model

and observational data were processed to correspond to the temporal and area-weighted spatial averaging of the PAGES2k temperature reconstructions (Ahmed et al. 2013) for each continental-scale region. In addition, we matched the CMIP5 data seasonally to the season targets in PAGES2k. See Chapter 3 for more information on the methods of subsetting data by region and season.

We also conduct our analysis in frequency space using the power spectral density of the complex model simulations from the past1000, PiControl, and Historical compared with the relevant observational datasets and the PAGES2k paleoclimate reconstruction. For example, we compare the PAGES2k dataset with the past1000 model run, and the HadCRUT and GISTEMP datasets with the Historical model results.

The spectral densities were generated by using the multitaper method (Riedel and Sidorenko 1995; Thomson 1982; Walden et al. 1998) in R through the *multitaper* package (Rahim et al. 2014), which achieves an optimal tradeoff between leakage and resolution (Yiou et al. 1996). It uses K modified periodograms with each one obtained using a different Slepian sequence as the window. Let $S_f(f)$, defined by,

$$S_f(f) = \Delta t \left| \sum_{n=0}^{N-1} g_{k,n} x_n e^{-i2\pi f n \Delta t} \right|^2 \tag{4.1}$$

denote the modified periodogram obtained with the k-th Slepian sequence, $g_{k,n}$. In the simplest form, the multitaper method simply averages the K modified periodograms to produce the multitaper power spectral density estimate.

$$S^{MT}(f) = \frac{1}{K} \sum_{k=0}^{K-1} S_k(f) \tag{4.2}$$

The multitaper spectral estimation technique uses the weighted average of

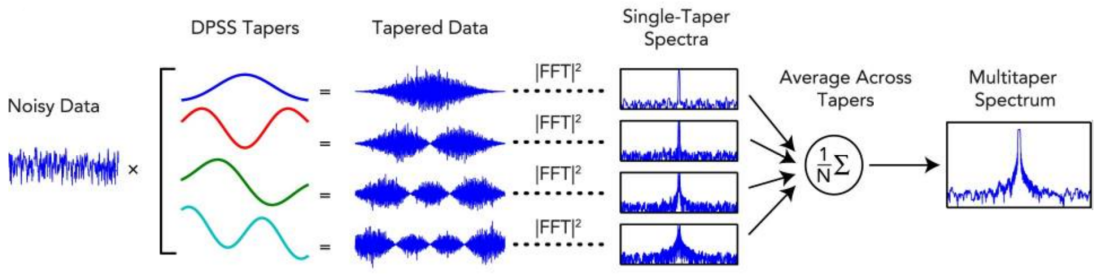


Fig. 4.1: Schematic of multitaper method (Figure 4b in Prerau et al. 2016; used with permission by publisher under licence number 4754810545289).

several direct spectral estimates over the entire signal, each computed using a different member of a family of orthogonal tapers. This is graphically represented in Figure4.1.

We use a half-bandwidth parameter of 4, which favors statistical significance over resolution, and is therefore the most conservative setting. When estimating the spectral density, the mean is removed from model data and observations at either the annual or seasonal level depending on the temporal scale of the data (see Table 3.2), with the exception of PAGES2k, to avoid biasing the estimate.

Figure 4.2 provides an example of the resulting spectral densities of the CMIP5 models and observational datasets used in this study over Australasia.

In analyzing the results, we compare the average spectral densities of the models and observations by differencing the values in select time bins (i.e., 02-05 to 70-100 years; Huybers and Curry 2006; Zhu et al. 2019). We select the time bins based on the periods we are interested in understanding ranging from short time scales (02-30 years) to long time periods (50-100 years). The equation describing

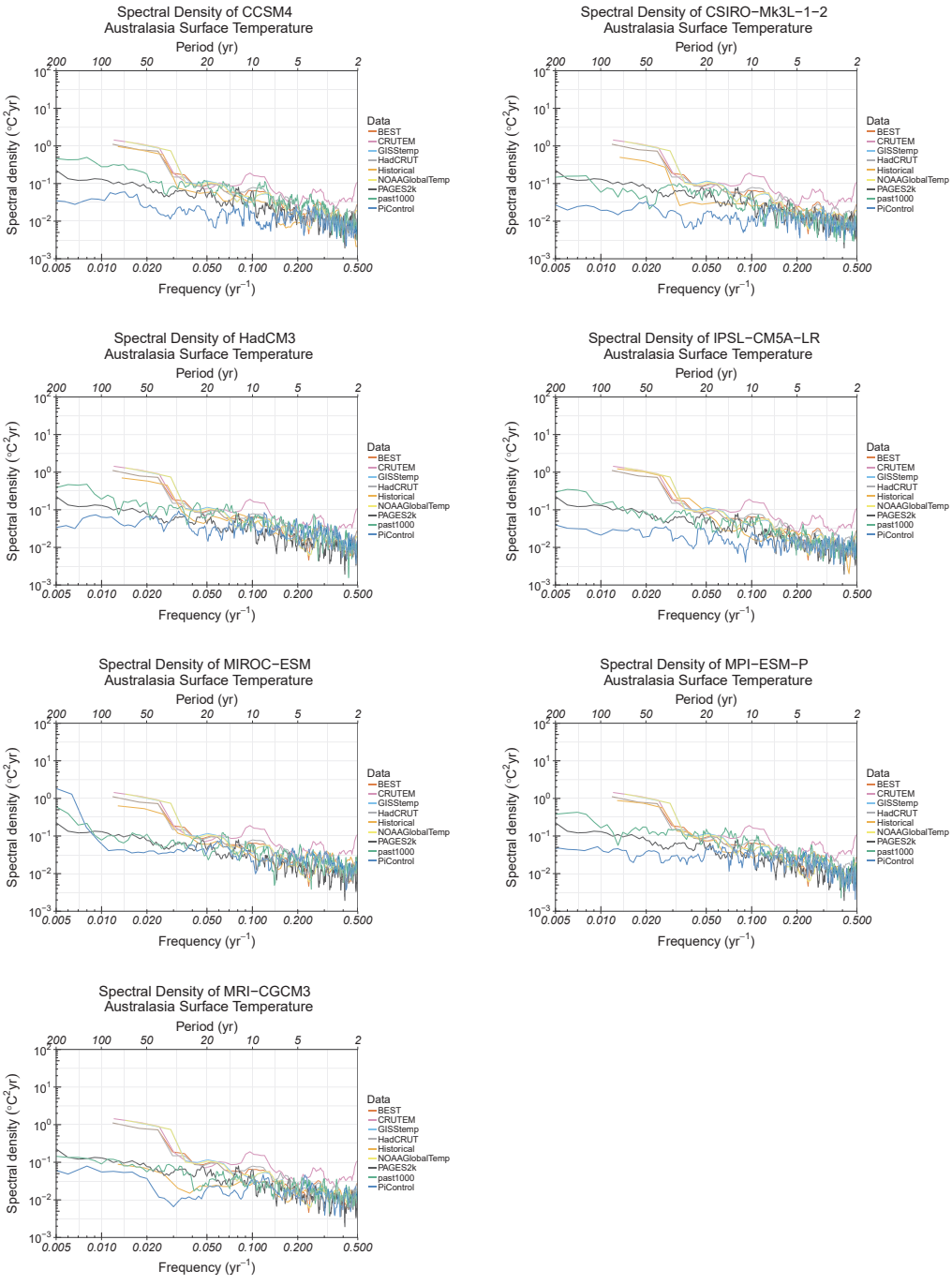


Fig. 4.2: Spectral Density of Models Across all Regions.

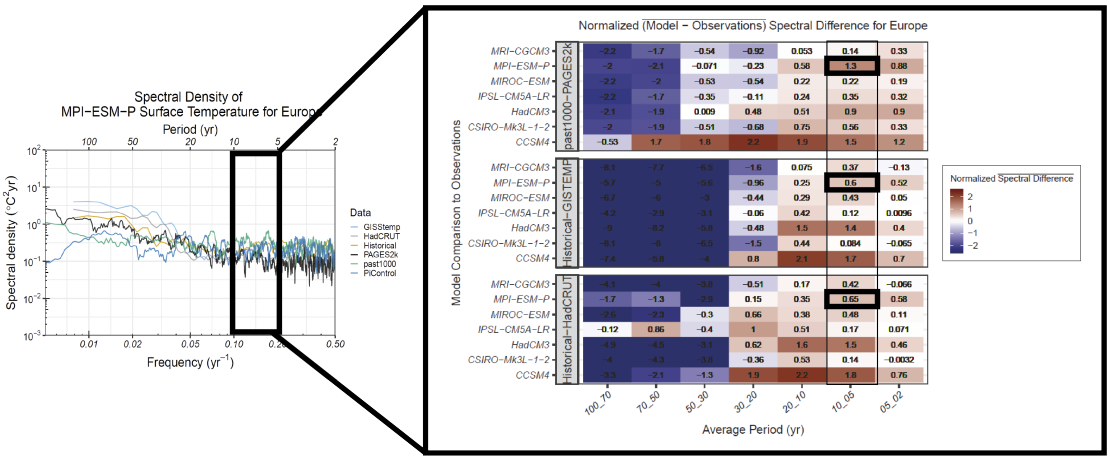


Fig. 4.3: Schematic of Method to Find Normalized Spectral Difference.

this procedure to compare models and observations is as follows:

$$\overline{S_{normalized\ difference}(f')} = \frac{\overline{S_{model}(f')} - \overline{S_{observation}(f')}}{\overline{S_{multi-model\ mean}(f')}} \tag{4.3}$$

where $\overline{S}$ is the average spectral density for a given frequency band, f'. We normalize by the spectral density multi-model mean, produced by averaging the spectral density of the long model PiControl runs, in order to compare across our results.

Biases associated with nonuniform sampling in frequency space—whereby more points are located in higher-frequencies than low-frequencies—have been addressed in the literature (Babu and Stoica 2010; Zhu et al. 2019), and we sample our data at the same frequencies in addition to using the frequency binning procedure to address these issues (Huybers and Curry 2006; Zhu et al. 2019).

Rather than simply providing qualitative descriptions of the spectral densities, we produce box plots for each region that show the normalized average spectral differences. Figure 4.3 illustrates how we arrive at this graphical representation of Equation 4.3. To produce normalized average spectral differences (right), we

first estimate the spectral density over all time periods (left). We then average each dataset separately over the spectral bins of interest (black outline on left) and find the difference between relevant datasets (e.g., past1000-PAGES2k, Historical-GISTEMP, Historical-HadCRUT). The result, the average normalized difference, is the represented graphically for each period of interest using a color scale to indicate the magnitude of the difference (right). Positive differences or the red side of the scale indicate that models have greater variability than the comparison observational dataset. The opposite is the case for negative average normalized differences or the blue side of the color scale. The intensity of the color indicates the magnitude of difference.

4.3 Results and Discussion

Here, we walk through each step of our results, introducing the relevant graphical presentations of the analysis and discussing their relevance. We step through our analysis of model control run variability, our comparison of models to models, and finally the comparison of models to observations. This will be followed by a region-by-region assessment of the results.

4.3.1 Control Run Variability

We begin by characterizing the PiControl run variability, which is important for determining over which time scales we can consider our subsequent results reliable. We carry this out by conducting a bootstrap estimation of the normalized

average spectral density ($^{\circ}C^2$/year) in the PiControl models for a given region and time span. We sample each PiControl temperature-time series at a given length, L, before estimating the spectral density. We then follow the same procedure as above to find the average normalized spectral density for each of these samples.

We conduct this analysis at L=150 years to understand the robustness of our spectral density estimates of the ∼150-year-long historical observational data, HadCRUT and GISTEMP, compared to the models. We report the halfwidth of the normalized spectral density when describing the robustness of the control run. We consider the normalized spectral density results we report to be robust if the difference exceeds the halfwidth of the normalized spectral density found from the PiControl run.

Figure 4.4 shows the normalized spectral density, and therefore variability, for each L=150 year sample as a point. We see that the results range from ±1 over longer time periods (lower frequency) across the suite of models used in this study, implying that differences between models, or models and historical observations, of this magnitude could be caused by sampling. We determine that the last two bins (50-70 and 70-100 period bins) are less reliable and so we do not provide results over these period bins in the main discussion here.

Although, we note that the results over the aforementioned bins are not always less reliable for all models. Some models in certain regions have a robust, though low, spectral power across all frequencies. Taking Australasia as an example we see that CSIRO-Mk3L-1-2 and IPSL-CM5A-LR have consistently lower variability across all time scales, while other models do not. We further discuss the bifurcation,

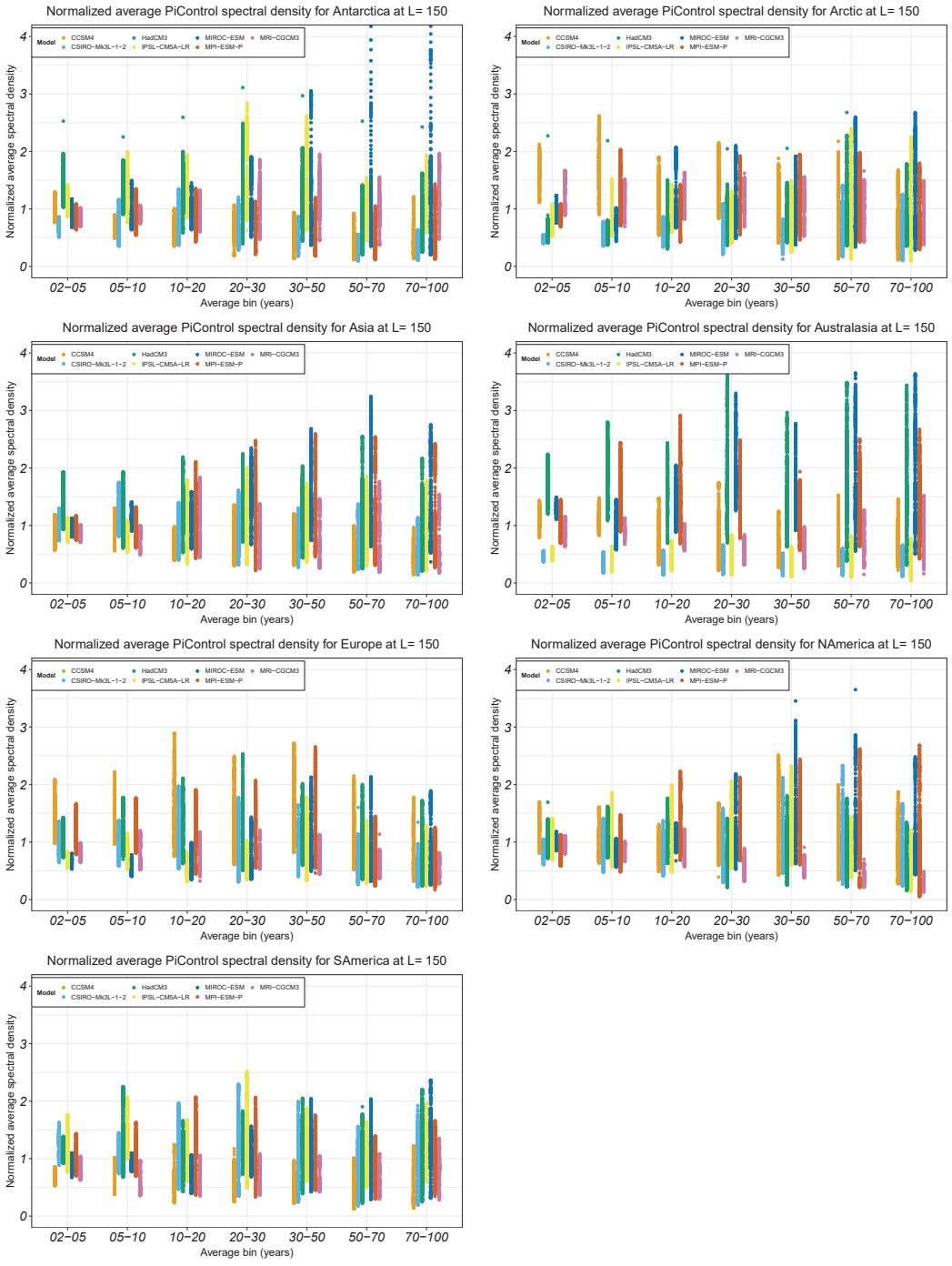


Fig. 4.4: Bootstrap Estimation of PiControl Internal Variability at L=150 years. Each model-specific data point represents a PiControl sample of length, L.

or split, in the variability for this particular region below.

The logic behind this robustness analysis applies to observations and also reconstructions. Reconstructions represent only one realization of the Earth system, and therefore, we can only expect models to agree with observations within the natural variability of the Earth system. Additionally, the actual background variability is unknown, presenting an irreducible level of uncertainty in this analysis and the literature, in general.

We also conduct the bootstrap estimation at L=500 years (Figure 4.5), which is the length of the shortest control run experiment used here, providing a general estimate for the magnitude beneath which differences between models and observations are not reliable. The magnitude and halfwidth values for our analyses are available in Table 7.2 in Appendix 7.

Looking across the models in the robustness analysis, we see that the range in normalized average spectral densities often do not overlap, leading us to conclude that the models examined here are distinct in their variability. This could be exploited in the future for studies of model finger printing (Allen and Stott 2003; Allen and Tett 1999; Stott et al. 2003).

One of the models we use in this study, MRI-CGCM3, has a PiControl run 500 years in length, and therefore, is not included in Figure 4.5 because we cannot sample the model at this length. We can instead rely on the L=150 analysis because MRI-CGCM3's average spectral density is similar to other models used in this study that do have longer control runs, such as MPI-ESM-P or MIROC-ESM. Looking at the results where this model is valid in the L=150 year bootstrap analysis, we find

no evidence that 500 years of PiControl data is insufficient to be considered reliable.

However, the L=500 analysis may underestimate the robustness of the results for most PiControl runs, because we are not sampling the full length of the control run time series, which is generally longer than 500 years. In the later results, we look at the deviation from the full control run to assess this rather than limiting our analysis to just 500 years. We further discuss region-specific characteristics below.

4.3.2 Model Noise and Forced Variability

In comparing models to models, we begin by looking at the normalized difference between the individual PiControl model runs and the PiControl multi-model mean. As mentioned above, the multi-model mean is found using those models with (1) sufficiently-long PiControl series ($\geq$ 500 years) and (2) available past1000 simulations (top of Figure 4.7). We normalize our results for comparability across analyses, placing all our results on the same "scale" (see Equation 4.3). As a result, the normalized average spectral density differences are unitless. In Figure 4.6, we provide the normalized model-to-model spectral differences found in Figure 4.10 with overlaid values.

From Figure 4.7 (top), we see that noise in the individual models differ from the multi-model mean across timescales and regions with little consistency. This analysis allows us to understand the magnitude of noise in each model as we step through subsequent analyses.

We can also explore how the models respond to perturbations (bottom of Figure 4.7) in the past1000 experiment (i.e., volcanic forcing, solar cycle) by looking

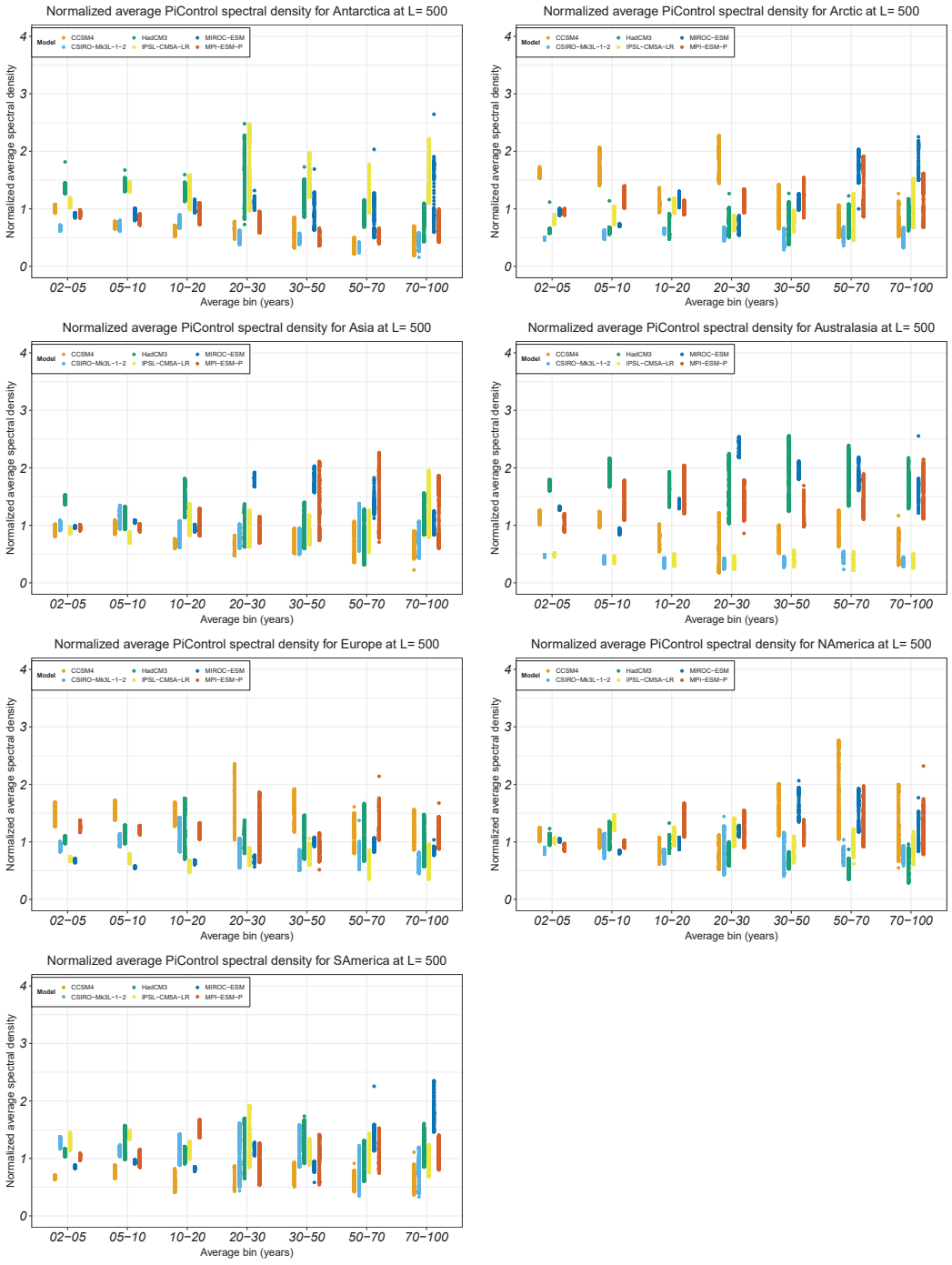


Fig. 4.5: Bootstrap estimation of PiControl internal variability at L=500 years. Each model-specific data point represents a PiControl sample of length, L. The magnitude and halfwidth values for this analysis are available in Table 7.2 in Appendix 7.

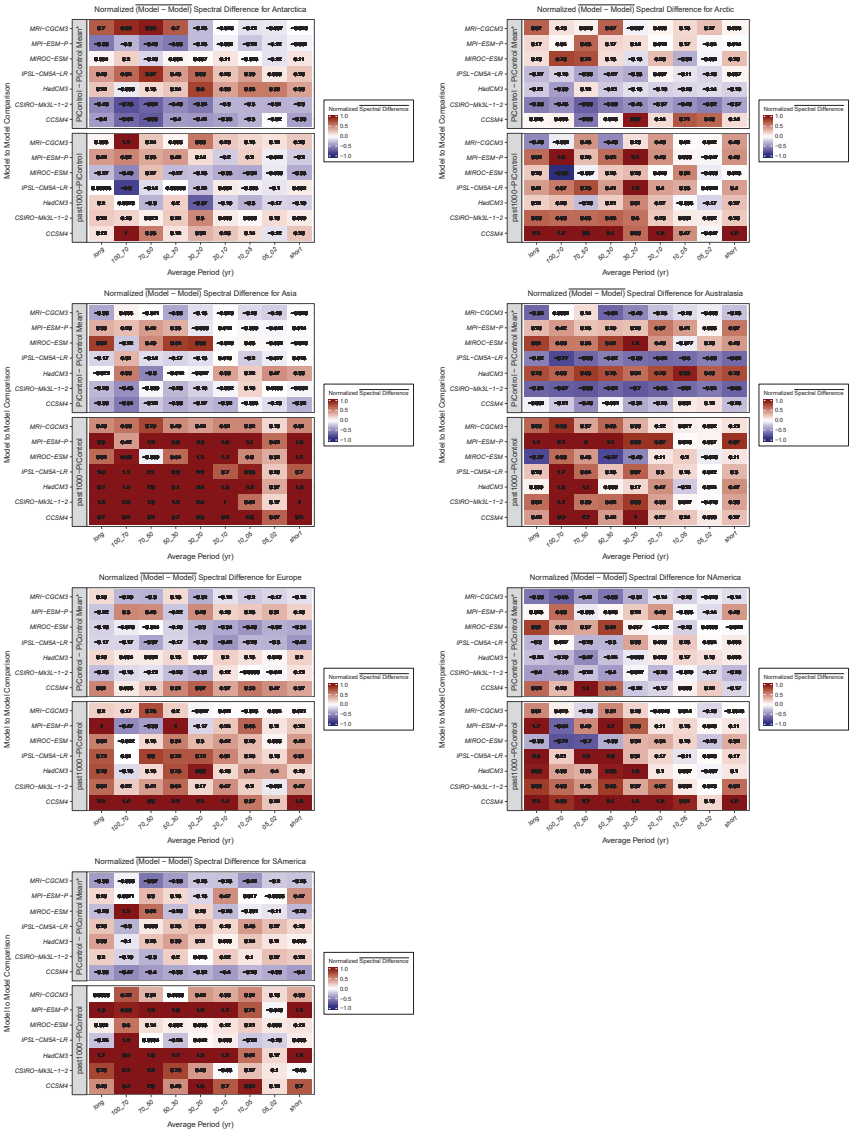


Fig. 4.6: Normalized model-to-model control run spectral difference with values. Positive differences or the red side of the scale indicates models have greater variability than the multi-model mean (top) or respond more strongly to perturbations than the multi-model mean (bottom). The opposite is the case for negative differences or the blue side of the color scale. The intensity of the color indicates the magnitude of difference.

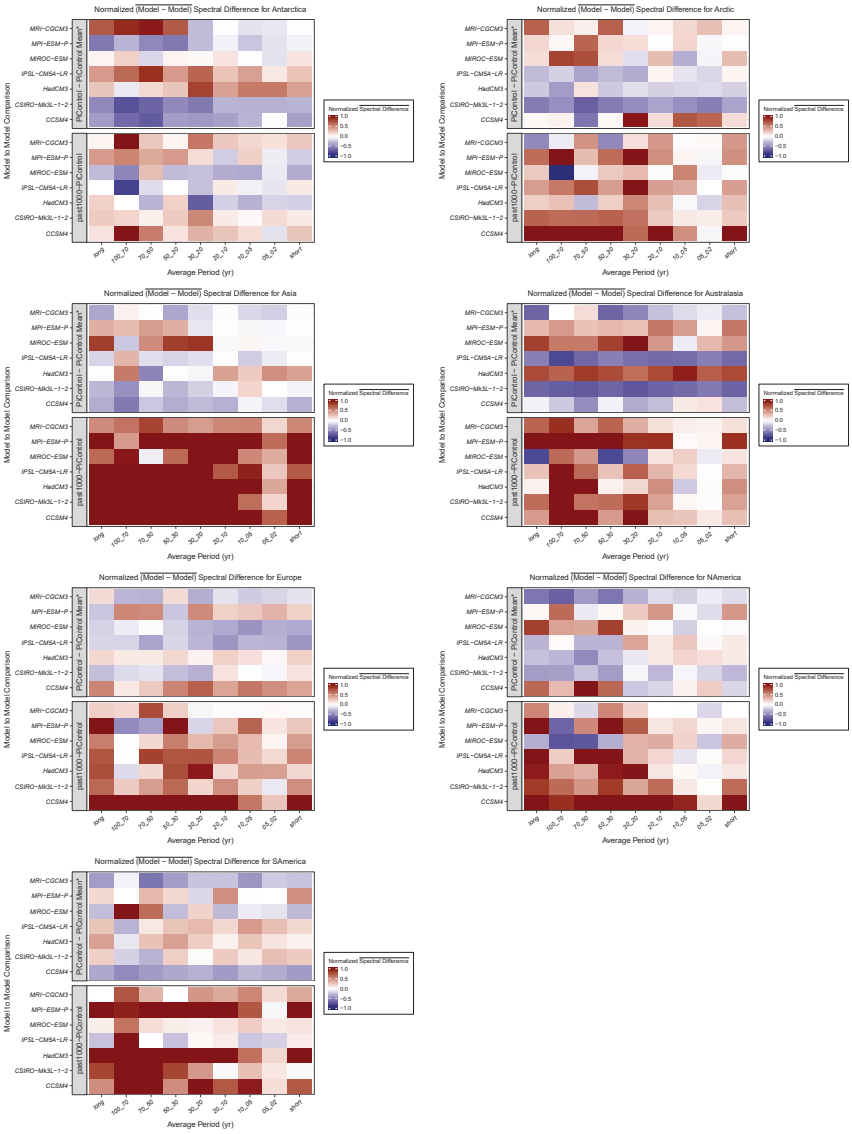


Fig. 4.7: Normalized model-to-model control run spectral difference. Positive differences or the red side of the scale indicates models have greater variability than the multi-model mean (top) or respond more strongly to perturbations than the multi-model mean (bottom). The opposite is the case for negative differences or the blue side of the color scale. The intensity of the color indicates the magnitude of difference. Values are available in Figure 4.6.

at the normalized average spectral differences of the past1000 experimental forcing response and unforced variability represented by the control runs. We see that the models have varied responses across the regions.

In addition to interpreting the results for select bins, we also compare the responses over long (50-70 year periods) and short (10-30 year periods) time scales. We find that in the control run, many models have characteristic long and short variability generally consistently across regions. Combined with the boostrap analysis, these two analyses need to be used together to get a complete picture of the response and its robustness. The bootstrap figure shows how the results vary depending on the sample chosen for the control run, while our other results show the mean response, not the variability.

4.3.3 Model-to-Observations Comparison

When comparing the models to observations of the recent ∼150 years, we also determine that the last two longer period bins (50-70 and 70-100 year periods), which have a variability of ±1, are less reliable and have larger noise. Therefore, we do not present results for these periods in our main analysis. This is also supported numerically and we provide the full set of values in Figure 4.9. In Figure 4.9, we provide the full normalized model-to-observation spectral differences with values. In the remainder of the chapter, we do not report the differences over all period bins.

In looking at our PAGES2k comparison, we rely on the literature to inform the time scales at which we find the data reliable. From the PAGES2k Consortium, itself, we know that this data is highly smoothed, even at annual time scales (Bothe

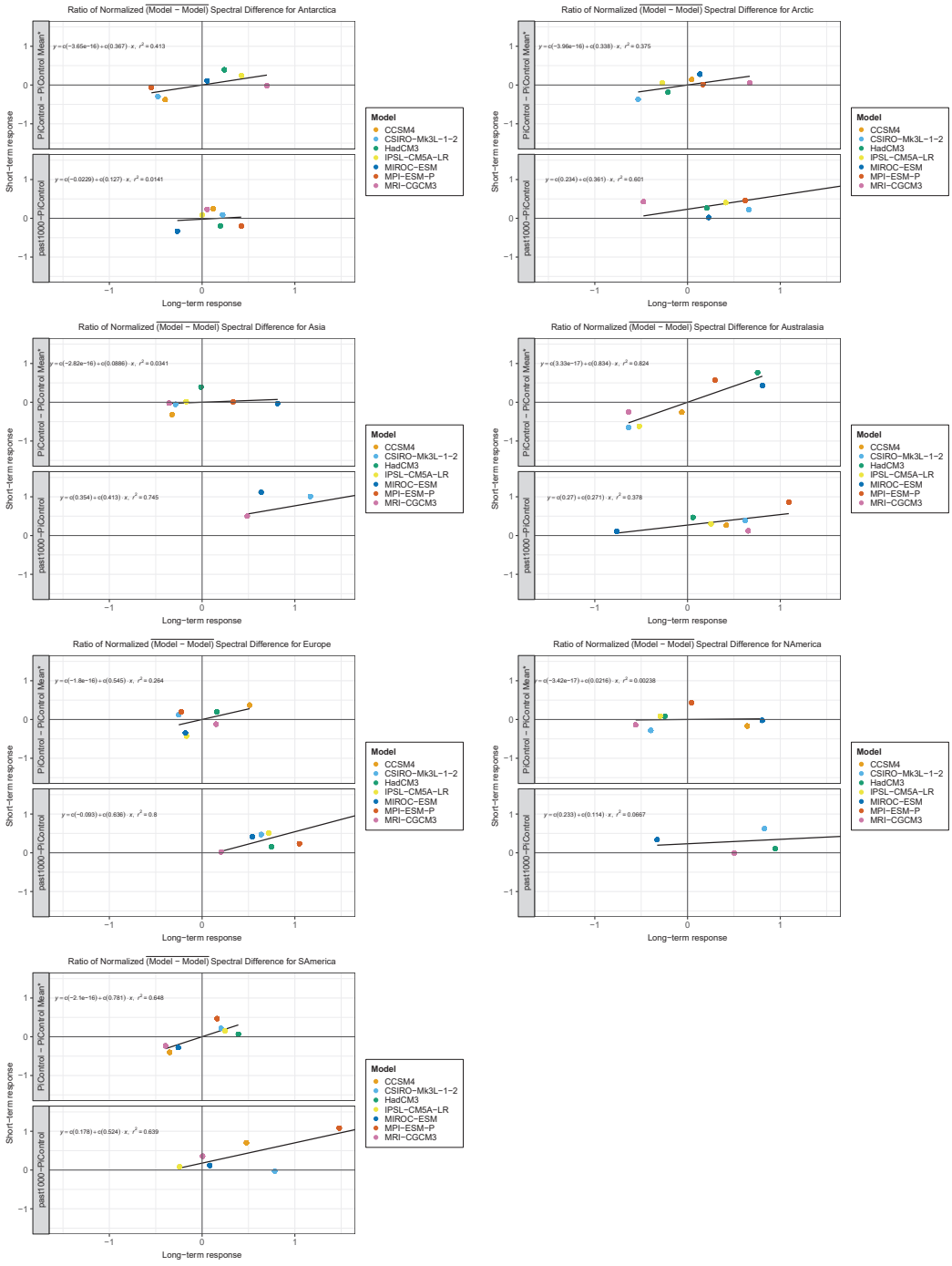


Fig. 4.8: Ratio of Short-term to Long-term Normalized Average Spectral Difference Between Models.

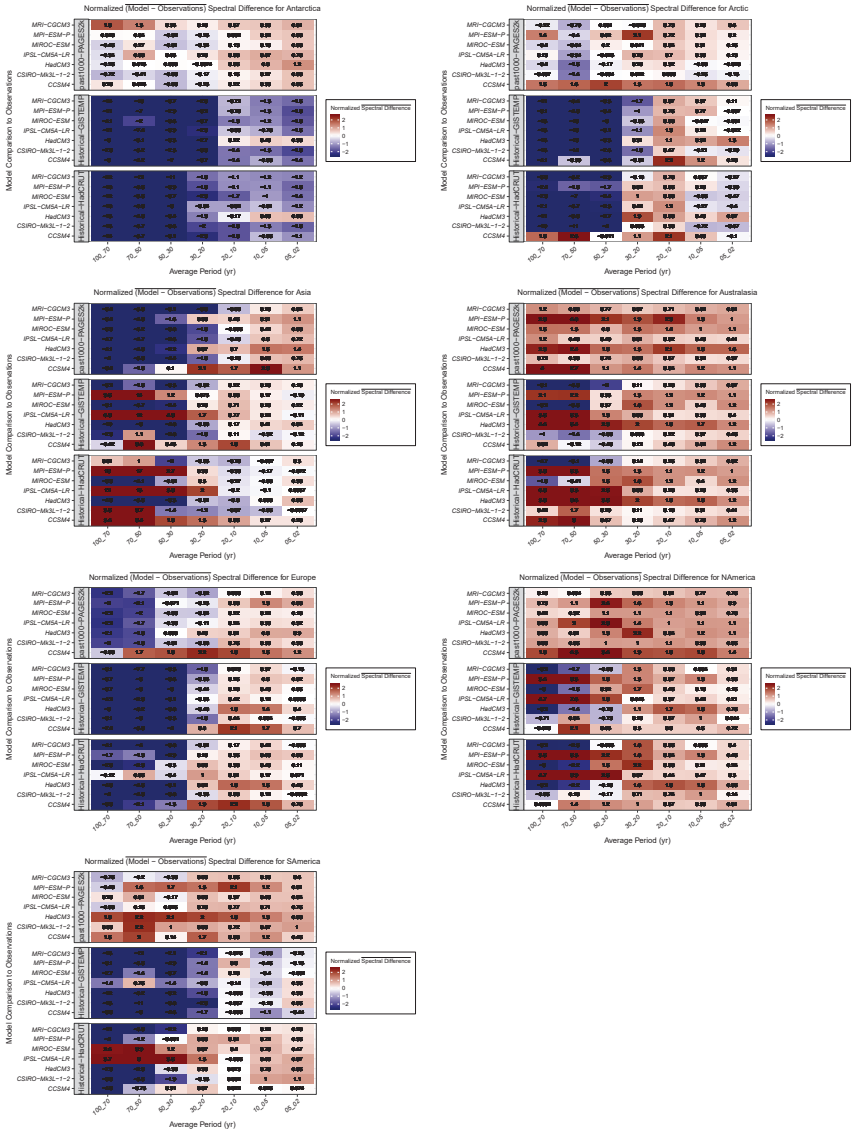


Fig. 4.9: Normalized model-to-observation spectral difference with values.

et al. 2013a. As a result, we also remove short periods (or high frequency) results from the main PAGES2k comparisons.

4.4 Regional Analysis

Below, we explore the results region-by-region, drawing conclusions about the differences in internal model variability across the models, as well as comparing variability in both the models and observations. We provide a thorough exploration of the results for the first region, Australasia, and highlight primary results for the remaining regions.

4.4.1 Australasia

Control Run Variability. For Figure 4.5 at L=500 years, we find that the halfwidth between the 20-30 year period bin up to the 70-100 year period bin is ±1 across the range of models, while the shorter time periods up to 10-20 year period bins have a range of normalized spectral density of ±0.6. Thus in this region, the models have higher robustness at shorter time periods. We also see a bifurcation in the model variability in Figure 4.5, which shows the robustness of internal model variability over different timescales sampled at L=500 years.

Two models, CSIRO-Mk3L-1-2 and IPSL-CM5A-LR, have lower spectral density, and therefore variability, at all periods compared to the other models. We report both the magnitude of the variability (center value) and the robustness of the results (halfwidth). CSIRO-Mk3L-1-2 has an average normalized spectral density

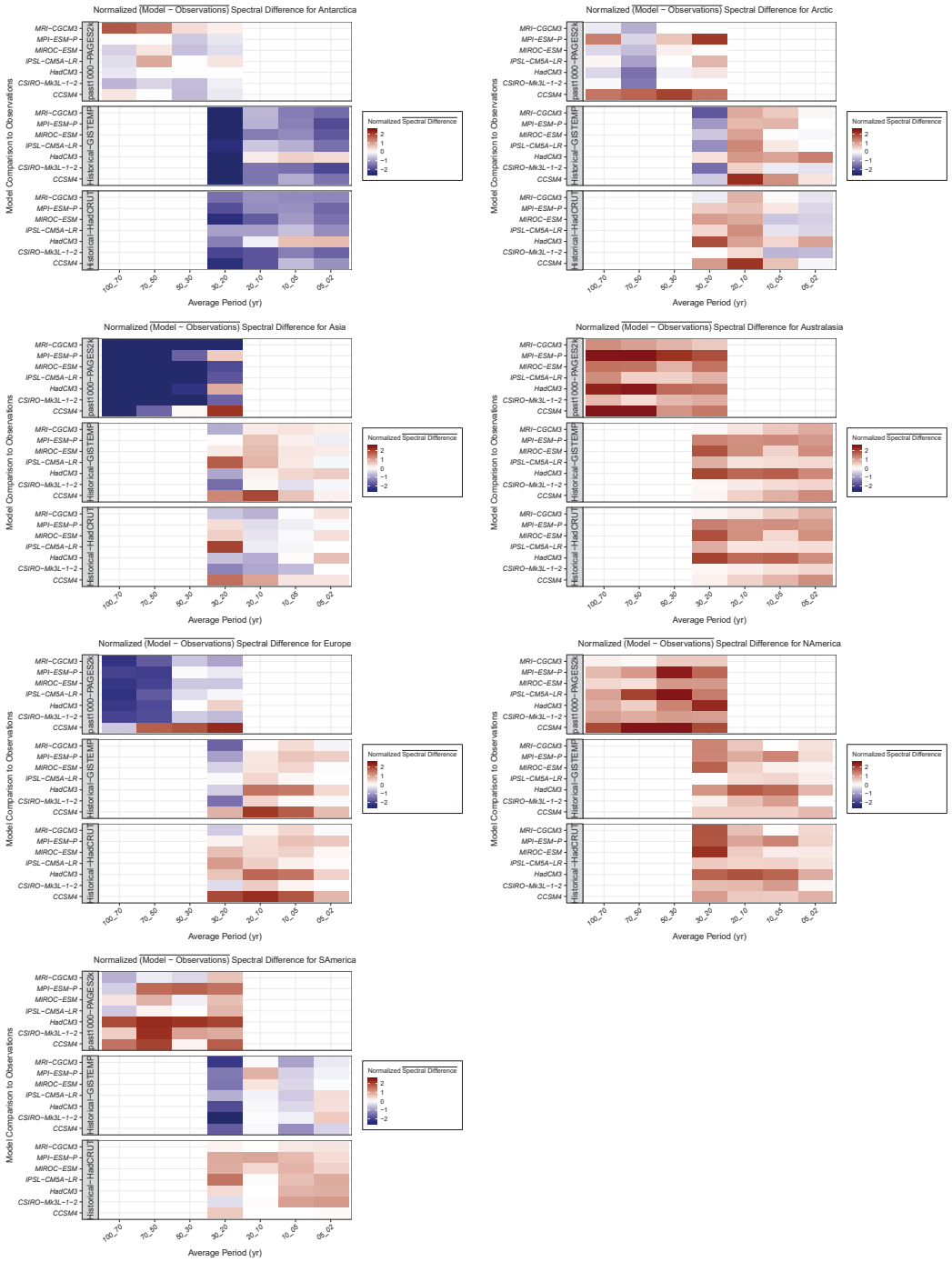


Fig. 4.10: Normalized Model-to-Observation Spectral Difference. Values are available in Figure 4.9, which shows full results including those excluded in the main analysis.

ranging from 0.45±0.018 over the 02-05 period bin to 0.34±0.13 over the 50-70 year period bin. Similarly, IPSL-CM5A-LR has an average normalized spectral density ranging from 0.47±0.03 to 0.33±0.14 over those same period bins. We note that we provide a better estimate of the spectral density, below, when using the full time series of each of the control runs, rather than just the 500 year samples.

Looking at Figure 4.7, we see that the magnitude of the difference from the multi-model mean for these two models is more negative than other models. For instance, CSIRO-Mk3L-1-2 has a normalized spectral difference of −0.54 over the 02-05 period bin to −0.69 over the 50-70 year period bin. (Recall that the negative values indicate that the models have smaller variability compared to the multi-model mean.) Similarly, IPSL-CM5A-LR has a normalized spectral difference ranging from −0.53 to −0.63 over those same period bins.

Comparing the overall magnitude of the control run deviation (Figure 4.7) and the robustness of these results (Figure 4.5) to characterize these individual models, we see that CSIRO-Mk3L-1-2 and IPSL-CM5A-lR are robust with the lowest overall spectral power in all bins, and larger, less robust deviation from the multi-model mean.

We also see that some models, HadCM3 and MPI-ESM-P, have higher spectral density, and therefore variability at all periods compared to the other models. HadCM3 has an average normalized spectral density range from 1.64±0.092 over the 02-05 period bin to 1.6±0.46 over the 50-70 year period bin. Similarly, MPI-ESM-P has an average normalized spectral density ranging from 1.0±0.15 to 1.3±0.34 over those same period bins.

CCSM4 has a high, but consistent control run difference, and the range of spectral density over the period bins decreases over longer periods. In addition, we note that the results appears to change state in the periods with low robustness. So, for example, in the 20-30 period bin, depending on what 500-year period was chosen, CCSM4 might have relatively low spectral power, comparable to the two low variability models, or might have spectral power more comparable to the higher variability models. It seems CCSM4, therefore, might undergo some sort of regime shift in terms of variability in this frequency range.

Model Noise and Forced Variability. The bottom panel of Figure 4.7 shows the strength of the model response to perturbations in the past1000 experiment. In this particular region, Australasia, we see stronger responses relative to the control run across the models at longer time periods (lower frequencies).

In Figure 4.7, MPI-ESM-P has a larger response to the past1000 forcings over longer periods (averaging 1.1 over the 50-100 year period), compared to the other models, but it also has lower levels of internal variability compared to the multi-model mean (0.29 over the same period bin). We also see that MPI-ESM-P has generally small responses at shorter periods (high frequencies). By contrast, MIROC has the weakest response to past1000 forcings (−0.77 over the 50-100 year period), but the highest level of internal variability (0.81 over the same period). In these two examples, we have one model with low noise and a high response (MPI-ESM-P), and vice versa for the other (MIROC). This result suggests that for this region, the response to the past1000 forcing is not correlated with the noise.

Model-to-Observations Comparison. Australasia presents an interesting

case when comparing the variability of the models and observations, where the models have higher levels of variability across all time scales. Only two models, IPSL-CM5A-LR and CSIRO-Mk3L-1-2, are close to the observations as illustrated by Figure 4.9. In comparison to HadCRUT, IPSL-CM5A-LR has an normalized spectral difference ranging from 0.28-0.35 for the 10-20 to 02-05 spectral bins, respectively. (We note that the observational datasets have a high coherence in this region and so will rely on HadCRUT when explicitly stating results). Over the same comparison periods, CSIRO-Mk3L-1-2 has an normalized spectral difference ranging from 0.19-0.41, respectively.

According to our analysis, the spectral density of the past1000 response in these two models is larger than the halfwidth range from unforced model noise (Figure 4.5; see discussion above), and so the results can be considered robust from the standpoint of internal model variability. The similarity in the variability of the observations and CSIRO-Mk3L-1-2 is of particular interest because this model is produced by an institute in Australia, and literature on the model indicates that it simulated oceanic circulation including the strength of the Antarctic Circumpolar Current, reasonably well compared to observations. CSIRO-Mk3L-1-2 also produces reasonable representations of the leading modes of internal climate variability (Phipps et al. 2011).

For other models, the higher level of variability across all time scales suggests some model processes result in higher spectral differences, and therefore variability, across all time scales compared to the observations. The higher variability could be caused by modes of variability in the models that are particularly strong and not

captured by the observations, or caused by a consistent deficiency in the models. Our conclusion is particularly robust across both paleoclimate data and the two instrumental records, and across all time scales, with the exception of IPSL-CM5A-LR and CSIRO-Mk3L-1-2. IPSL-CM5A-LR and CSIRO-Mk3L-1-2 are also close to the PAGES2k reconstruction. IPSL-CM5A-LR has a normalized average spectral difference of 0.49 over the 50-70 year period bin and CSIRO-Mk3L-1-2 as a difference of 0.38 over the same period. We can only conclude that the variability of these two models is larger than the reconstruction, which may itself have biased-low variability due to smoothing.

Regional Conclusions.

1. The internal model variability in this region is bifurcated, such that two models have lower levels of variability, while two others have higher levels of variability compared to the multi-model mean.

2. The bifurcation and variability in the model responses suggests one should use caution when making assumptions about or relying on model noise to be a good predictor of natural variability.

3. In this region, the model noise and response do not appear to be correlated.

4. All models have larger variability compared to observations across all time scales.

4.4.2 Antarctica

Exploring Control Run Variability. Antarctica presents another particularly interesting case where the model variability in Figure 4.5 is bifurcated across all spectral bins. Again we see that CSIRO-Mk3L-1-2 has lower internal variability compared to the other models with a halfwidth between the 20-30 year period bin up to the 70-100 year period bin ranging from 0.49±0.11 - 0.46±0.26, respectively, while the shorter time periods up to 10-20 year period bins have a range of normalized spectral density of ∼±0.7.

CCSM4 also serves as an "intermediate case" in Antarctica. Initially, the internal variability is grouped with other models, but over longer period bins, we see that the variability decreases compared to the other models to 0.56±0.32 at the 70-100 year period bin.

By contrast, IPSL-CM5A-LR in Antarctica has a high level of model noise that varies greatly over time from 1.1±0.08 for the 02-05 year bin to 2.1±0.70 for the 70-100 period bin. In general, it is unsurprising that models would have high levels of internal variability over Antarctica because physical processes across a number of timescales are not well constrained by observations.

Model Noise and Forced Variability. Looking again at the bottom panel of Figure 4.7, we see that in this region, though model responses to the perturbations in the past1000 experiment vary for each model over time, the magnitude of the normalized average spectral density difference is lower than for other regions. For instance, if we look across all the models for this region on short time scales, we

have a response range from -0.33 to 0.25, and a range of -0.27 to 0.42 over long time periods.

However, based on our analysis of the robustness of our internal variability estimates, we see that at some intermediate time scales (i.e., 05-10 and 10-20 year periods) the model responses to perturbations in the past1000 experiment are within the halfwidth range of internal model variability (see discussion above). For instance, we find that MPI-ESM-P has an average spectral density difference in response to past1000 forcing of -0.074 and 0.2 over the 02-05 and 05-10 year period bins, respectively, while the halfwidth for those same period bins are ± 0.067 and ± 0.093. Thus, the results cannot be considered robust on these time scales.

However, at other time scales some models have responses robust against internal variability. The exception is CCSM4, which has stronger responses than the control run that are robust across most timescales, other than over the 30-50 year period bin. One will recall that this model's internal variability decreases over longer time scales. Thus, CCSM4 has a high response, and increasingly lower noise relative to other models.

This suggests that for Antarctica, as with the previous region, that model response and noise are not correlated. However, this conclusion is not robust over all time scales where some models may have high noise and a low response, or other combinations therein.

Model-to-Observations Comparison. In Antarctica, we see consistently robust and negative normalized spectral density differences between the model and the observations across all models and time scales, with a general coherence between

the two historical observation datasets.

Only HadCM3 has a positive spectral difference, especially over shorter time scales, indicating that for this region, the model has a stronger response compared to the observations. However, over the majority of period bins, these results are within the halfwidth, suggesting that they are within variability due to sampling. Only over the 02-05 and 05-10 period bins can we consider the results robust because the difference is not within the halfwidth.

Thus, either the models fail to capture important features in the observational datasets, or the observations are underrepresenting variability present in the actual climate system in this region, or a combination of both. Because both models and observations are known to be problematic in this region, it is not possible with our analysis to draw conclusions about which explanation is most likely.

The normalized average spectral difference between the past1000 experiment and PAGES2k shows mixed results. Several models have forced response spectral differences close to the observations including HadCM3. We also note that the normalized magnitude of the model and observational spectral densities is less than the halfwidth, making it difficult to draw a conclusion for these models since the spectral density signals are not robust, even for HadCM3. Thus, it is important to exercise caution when using models internal variability in this region.

Regional Conclusions.

1. We note that for Antarctica, the bifurcation and variability in the model responses suggests one should use caution when making assumptions about

or relying on model noise to be a good predictor of the natural variability, particularly at the low end of the spectrum (or high frequencies).

2. Compared to the historical observational datasets, the majority of the models have robustly negative normalized spectral difference, and therefore variability, indicating a bias in either the model or observations.

3. Because of the high level of internal model variability, or noise, in this region, it is not possible to draw an overarching conclusion about the model responses across all time scales.

4.4.3 Arctic

Exploring Control Run Variability. Looking at our second polar region, the Arctic, we do not have a clear bifurcation in the ranges of internal variability, as with Australasia, but there are some interesting features that are worth exploring here (Figure 4.5).

For example, CCSM4 has a more positive normalized spectral difference, and therefore, higher internal variability and level of noise, especially in the short time periods (high frequencies) compared to other models. For instance, CCSM4 has robust normalized spectral differences of 0.14 over the 02-05 period bin to -0.56 over the 50-70 year period bin.

And we see that the results for CSIRO-Mk3L-1-12 are similar in strength to CCSM4, though opposite in magnitude. The average spectral density halfwidth range from 0.53±0.025 over the 02-05 year period to maximum of 0.56±0.22 over

the 30-50 year period. Over these same spectral bins, the difference from the multi-model mean is -0.52 and increases to -0.66, and so can be considered robust.

We also see that just as in Antarctica, the HadCM3 model has one 500 year time slice with a normalized average spectral density that is outside the range of other slices in the model. Because of this, depending on the time range selected in our study, one may report a much higher level of internal variability for this model. It is important, therefore, to understand the range in internal variability by conducting such an analysis in order to prevent the selection of a potentially spurious level of internal variability.

Model Noise and Forced Variability. In the Arctic region, we also generally see a stronger response to past1000 forcing relative to the control run across the models at longer time periods (lower frequencies).

CCSM4, for example, has a particularly strong response to the past1000 forcing compared to the control run over long periods. The average normalized spectral difference, for instance, is 1.7 over the 70-100 year bin, which is robust for this model. As in Antarctica, one will recall that this model's internal variability decreases over longer time scales. Thus, this model has a high response, and low noise.

Other models in the region, including CSIRO-Mk3L-1-2, HadCM3, IPSL-CM5A-LR, also have strong responses to the past1000 experiment compared to the control run and low noise. However, the response of HadCM3 is not robust across all time scales because the halfwidths are greater than the normalized average spectral differences.

The three other models, MRI-CGCM3, MPI-ESM-P, and MIROC-ESM have

higher model noise and mixed results for the response depending on the time scale examined. For example, looking across these three models over the 70-100 year bin, we see that noise in MPI-ESM-P is similar to the multi-model mean (0.04), but the model has a strong response (1.6) that is robust compared to the internal model variability ($\pm$0.49).

Model-to-Observations Comparison. When comparing to observations, we see that the robustness and variability of the model responses are dependent on the time period bin being examined. Over the 02-05 period bin, for instance, the normalized average spectral density differences compared to HadCRUT are initially robust and low across the majority of the models, and increase over the 05-10 and 10-20 period bins. Relating this to the physical results, we find that the observations have higher spectral density, and thus higher variability, over the 02-05 time period, while the models have higher spectral density over the 05-10 and 10-20 period bins.

Over longer period bins where we rely on PAGES2k for a comparison, we can see that the differences have a somewhat inconsistent trend in normalized spectral density. For instance, over the 50-70 year period bin, PAGES2k has a lower spectral density than the majority of models, while over other longer time bins the models have a similar level of variability compared to the PAGES2k reconstruction.

CCSM4 does not follow the trends discussed above, and has a consistently high normalized spectral density difference, and thus a higher variability, compared to all the observational datasets.

We also note that over the Arctic, the responses between the two historical observation datasets are not coherent, such that for most periods the responses

across the datasets are different in magnitude and sign.

Regional Conclusions.

1. The limited overlap in the variability in the model ranges of internal variability suggests one should use caution when making assumptions about or relying on model noise to be a good predictor of the natural variability.

2. This region has little consistency in the model response to past1000 experiment across the modes. MRI-CGCM3, MPI-ESM-P, and MIROC-ESM have high model noise and mixed results for the response to the past1000 experiment depending on the time scale examined. CSIRO-Mk3L-1-2, HadCM3, IPSL-CM5A-LR have stronger responses to the past1000 experiment compared to the control run and low noise.

3. CCMS4 has a higher response to forcing relative to both observations and other models

4. Because of the high level of internal variability, or noise, in this region, it is not possible to draw an overarching conclusion about the model responses across all time scales.

4.4.4 Asia

Exploring Control Run Variability. In Figure 4.5, we see that the internal variability in the models is increasing in range over longer periods (lower frequencies). In addition, the internal variability across the models overlap, especially over longer periods, indicating that the internal variability estimated

from any one 500 year-long time series in a model would be generally consistent with one another. Other regions, such as in North and South America, also have a similar overlap in internal model variability. However, this serves as our first example explored here where a region generally considered to not be problematic (in comparison to the polar regions), and does not display some of the bifurcation in internal variability across the models seen previously.

We do see that HadCM3 as a higher internal model variability compared to the other models at the 02-05 year period, and MIROC-ESM has a higher internal model variability at the 20-30 year period. In general, however, the ranges in internal variability across the models overlaps.

Model Noise and Forced Variability. In Asia, we see that four models have consistently low noise, CCSM4, CSIRO-Mk3L-1-2, IPSL-CM5-LR, and MRI-CGCM3. The remaining three have low noise over some periods, such as MPI-ESM-P and MIROC-ESM in the short periods (02-05 to 10-20) and higher difference from the multi-model mean.

In response to the forcing in the past1000 run, we see that in Asia all models have large positive spectral density differences. From Figure 4.7, we see that these results are also robust because the normalized average spectral density difference between the past1000 and control runs are greater than the internal variability halfwidths. Further, across all the regions examined here, the models over Asia have the consistently strongest response to natural forcing compared to the control run.

Model-to-Observations Comparison. Over short-time periods (02-05 to

20-30) the comparison of the model spectral density to the observational spectral density yields mixed results. The normalized average spectral density differences for the two historical datasets bear some similarities in magnitude. However, GISTEMP has lower spectral density compared to the models, while HadCRUT has greater spectral density compared to the models. Drawing conclusions from the observational datasets is difficult because of the disagreement between the normalized average spectral density differences from the observational datasets.

Over longer time periods (greater than 20-30 years), PAGES2k has consistently larger negative spectral density values in comparison to the models. We already discussed that the model response is consistently large and positive in response to the past1000 experiment over all time periods compared to the control run. The higher variability could be caused by modes of variability in the models that are particularly strong and not captured by the paleoclimate reconstruction, or caused by a consistent deficiency in the models over this region and time periods.

Regional Conclusions.

1. Asia has a generally consistent level of internal model variability, thus model noise may be a good predictor of natural variability.

2. The models respond consistently strong, or with larger positive spectral density, to the past1000 experiment over all time periods compared to the control run.

3. It is difficult to draw conclusions about the recent historical period since the sign of the normalized average spectral difference compared to the historical

observational data do not agree.

4.4.5 Europe

Exploring Control Run Variability. We again report both the range in control run variability of the results (halfwidth). In particular, MIROC-ESM and IPSL-CM5A-LR have lower spectral densities, and therefore, lower internal variability than other models. MIROC-ESM has a spectral density halfwidth of ± 0.03 over the 02-05 period bin, ± 0.15 over the 05-10 period bin, and ± 0.33 over the 10-20 period bin. IPSL-CM5A-LR has a spectral density halfwidth of ± 0.35 over the 02-05 period bin, ± 0.74 over the 05-10 period bin, and ± 0.86 over the 10-20 period bin.

By comparison, CCSM4 has higher spectral density halfwidth over the same period: ± 0.19, ± 0.15, and ± 0.19, respectively. Thus, similar to the results of both Australasia and Antarctica, in Europe we also see a slight bifurcation in the model internal variability, however this is largely evident at short time periods (high frequencies).

Model Noise and Forced Variability. In Europe, two models have higher spectral density, and therefore, higher internal variability compared to the multi-model mean: CCSM4 and MPI-ESM-P. CCSM4 has a difference ranging from 0.47 over the 02-05 period to 0.23 over the 50-70 period bin. MPI-ESM-P has a range of difference from 0.31 to 0.48 over the same period bins, respectively.

We also see that CCSM4 has the strongest response to the past1000 experiment, which we also note are robust across all time scales. This model has a

strong response and a high level of noise.

The majority of other models show strong responses to the past1000 experiment and low levels of noise, including MIROC-ESM, IPSL-CM5-LR, and CSIRO-Mk3L-1-2.

As for other regions, we also note that over Europe, we can conclude that the model noise and response are not correlated.

Model-to-Observations Comparison. Over short time scales (02-05 to 10-20 period bins) we see that some models have similar or higher spectral densities compared to the observations. These differences are also coherent across both observational datasets. Over longer time periods the models have lower spectral densities than PAGES2k.

Regional Conclusions.

1. Over Europe, we find that the model noise and response are not correlated.
2. We do not find a trend in the model response to the past1000 experiment and conclude that the response is model specific.
3. Over longer time periods the models have lower spectral densities than PAGES2k.

4.4.6 North America

Exploring Control Run Variability. In Figure 4.5, we see that the internal variability in the models is increasing in range over longer periods (lower frequencies). As with Asia, the internal variability across the models overlaps,

however this is more prominent over shorter periods, indicating that the internal variability estimated from any one 500 year-long time series in a model would not be dramatically different from one another. We do see that CCSM4 as a higher internal model variability compared to the other models over long time periods, while HadCM3 lower internal variability over these periods. In general, however, the ranges in internal variability across the models overlaps.

Model Noise and Forced Variability. For this particular region, we do not see a consistent level of noise in the models across time scales. Instead, MIROC-ESM and CCSM4 have a higher noise compared to the multi-model mean compared to other models in the region, while MRI-CGCM3, IPSL-CM5A-LR, HadCM3, and CSIRO-Mk3L-1-2 have lower noise than the multi-model mean.

In this region as with Australasia, we have we have one model with low noise and a high response (CSIRO-Mk3L-1-2), and vice versa for the other (MIROC-ESM). Again we see that the response to the past1000 experiment is not correlated with the noise. It also worth recalling that for this region, the PAGES2k data is at a coarse temporal resolution and may not be capturing the spectral power seen in the historical datasets (see Chapter 3).

Model-to-Observations Comparison. Similar to Australasia, when comparing the variability of the models and observations, we see that the models have higher levels of variability across all time scales. We also note that the results are robust over all timescales. Further, this suggests that there is a model process that results in higher spectral differences, and therefore variability, across all time scales compared to the observations. We restate that this could be caused by modes

of variability in the models that is particularly strong and not captured by the observations, or caused by a consistent deficiency in the models. This conclusion is particularly robust across both paleoclimate data and the two instrumental records, and across all time scales.

Regional Conclusions.

1. As with Asia, the internal variability across the models overlaps, however this is more prominent over shorter periods.

2. This indicates that the internal variability estimated from any one 500 year-long time series in a model would not be consistent with another.

3. In this region, the model noise and response do not appear to be correlated.

4. All models have larger variability compared to observations across all time scales. However, the larger variability in the models could be caused by deficiencies sin the paleoclimate data due to the course data in North America (30 year for pollen data).

4.4.7 South America

Exploring Control Run Variability. In Figure 4.5, over South America we see that CCSM4 has consistently low internal variability across all time scales compared to the other models in this region. Over the 02-05 period, CCSM4 has a halfwidth of ± 0.034 which increases to ± 0.37 at the 70-100 period bin. By comparison, MIROC-ESM has a large range of variability of ± 0.45 over the 70-100 period.

Model Noise and Forced Variability. In South America, two models have lower spectral density compared to the multi-model mean: CCSM4 and MRI-CGCM3. CCSM4 has a range of difference from −0.33 over the 02-05 year period to −0.4 over the 50-70 year period bin. MRI-CGCM3 has a range of difference from −0.2 to −0.57 over the same period bins, respectively.

We also see that CCSM4 responses more strongly to the past1000 experiment, which we also note is robust across all time scales. This model has a strong response and a low level of noise. The majority of other models show mixed levels of noise. HadCM3, MPI-ESM-P, and CSIRO-Mk3L-1-2, also have strong responses to the past1000 experiment. Thus, as for other regions, we can conclude that the model noise and response are not correlated.

Model-to-Observations Comparison. The model comparison to the observations for South America yields mixed results. Over the 10-20 period bins, the models have a similar spectral density to the observations with coherence between the two historical observational datasets. The other time periods are not consistent between the models in this region and the two historical observational datasets.

In addition, the model response to the past1000 experiment is also not consistent across time periods with the historical observational data. Bothe et al. (2013a) described this difference in the PAGES2k data, finding that the reconstruction displays more centennial-scale variability than the simulations used in their analyses.

Regional Conclusions.

1. We also see that CCSM4 has a strong response to the past1000 experiment, which we also note are robust across all time scales. This model has a strong response and a high level of noise.

2. The majority of other models show mixed levels of noise. HadCM3, MPI-ESM-P, and CSIRO-Mk3L-1-2, also have strong responses to the past1000 experiment.

3. We can conclude that the model noise and response are not correlated.

4.5 Conclusions

Our study has focused on determining how realistic unforced model variability in complex models is compared to observations for regions and time periods important to human systems. It began by processing the climate and model data at continental scales consistent with PAGES2k over all time scales using estimates of the spectral density. This is the first study to systematically characterize the variability of models and observations over all time scales. From this we find several overarching conclusions about the observation and climate model variability.

First, for several continental-scale regions including Australasia, Antarctica, and the Arctic, the internal model variability is bifurcated. The bifurcation and variability across the model responses suggests one should use caution when making assumptions about or relying on model noise to be a good predictor of natural variability. From our analysis, we also note that even 500 year control runs can be insufficient to fully characterize variability in some models.

Second, we find that the model noise and response do not appear to be correlated. Throughout we have made note of models with low noise and a high response, and vice versa. This suggests that the response to the past1000 forcing is not correlated with the noise.

Third, several models have large positive or negative levels of variability across regions and time scales compared to the multi-model mean. For instance, CCSM4 has large positive spectral differences in the Arctic and Europe compared to the multi-model mean, while in Antarctica and Asia it has high negative spectral differences compared to the multi-model mean. Thus, model analyses should be conducted on a scale at which the noise can be characterized. In our study, we demonstrated that model noise is often inconsistent across regions.

Fourth, models have consistently larger variability compared to observations across all time scales for Australasia, and suggests some model processes result in higher spectral differences, and therefore variability, across all time scales compared to the observations. This could be caused by modes of variability in the models that are particularly strong and not captured by the observations, or caused by a consistent deficiency in the models.

In some instances, we also find a lack of coherence in the observational data, which more strongly suggests that for that particular region, the observational data are not capturing an important process.

Chapter 5: Climate Response Over Time (in prep for submission to Journal of Climate)

5.1 Motivation

Responses of the climate system to anthropogenic emissions vary on different time scales depending on the species emitted (Hansen et al. 1997; Manabe and Strickler 1964). Carbon dioxide (CO_2) emissions primarily drive anthropogenic climate change, however emissions of various other radiative forcing agents, such as short-lived climate forcers (SLCFs), contribute significantly to Earth's altered radiative budget (Shine 2000; Shine and de F Forster 1999). Much of the literature focuses on long-term climate responses emphasizing analysis with equilibrium climate sensitivity (ECS) or transient climate response (TCR; Forster and Taylor 2006; Senior and Mitchell 2000; Wigley et al. 2005). There is a more limited set of literature exploring short-term climate change responses occurring in a 20-30 year time horizon (Hasselmann et al. 2003; Pierrehumbert 2014; Shindell 2014; Smith and Mizrahi 2013). To address this gap, we investigate climate dynamics on decadal time scales with the ultimate goal of understanding the implications of near-term emission reductions on climate.

5.2 Background

Understanding spatial distributions of climate responses are important for addressing near-term climate change. In addition, the impacts of climate change will be (and are) felt most acutely at the regional level (Hawkins and Sutton 2012). Further, the rate of the temperature increase over multi-decadal time scales has important implications for century-long climate strategies because of the considerable inertia in the climate system. For example, despite the continued increase in anthropogenic GHG emissions, the expected temperature increase is not fully realized because the oceans have a large heat capacity-absorbing more than 97% of the accumulated energy in the climate system from 1971 to 2010 (Abraham et al. 2013).

By contrast, the land surface covers 29% of the Earth's surface, largely in the Northern Hemisphere (NH), and has a much lower heat capacity. The resulting differential heating of the Earth's surface has implications for seasonal climatology and hemispheric warming differences (Cubasch et al. 2001; Joshi et al. 2008; Lambert and Chiang 2007; Manabe et al. 1991; Meehl et al. 2007; Sutton et al. 2007). Anthropogenic emissions are also mainly generated over the land surface, adding to the importance of understanding hemispheric-scale contributions to climate change.

Though global emphasis has been placed on long-term emission mitigation strategies via CO_2 emission reductions, SLCFs generated over the land surface, including aerosols, play a significant role in modifying climate (Fuzzi et al. 2015; Ødemark et al. 2012; Schmale 2016; Shindell et al. 2012). Atmospheric circulations

allow longer-lived constituents to have relatively uniform concentrations throughout the atmosphere; SLCFs are more quickly removed via chemical and physical reactions. The SLCFs discussed in this dissertation, BC and CH_4, have atmospheric lifetimes of 4-12 days and at steady-state, 12 years, respectively (Bond et al. 2013; Lelieveld et al. 1993). In an effort to reduce the unprecedented rate of temperature increase in recent history, there is growing interest in near-term climate mitigation strategies specifically targeting SLCFs (Rogelj et al. 2014; Smith and Mizrahi 2013).

5.3 Methods

Understanding the temporal and spatial variations of SLCFs influences by perturbing CMIP5 models requires large computing resources and generates noisy output so few CMIP5 model results specifically investigate aerosol impacts. Therefore, in this study, we employ a quantitative approach to determine the range of responses in a stylized CMIP5 experiment.

We investigate the climate responses from two CMIP5 experiments: piControl (pre-industrial control, run for hundreds of years) and 1pctCO_2 (1% annual increase in [CO_2]; Taylor et al. 2012). Climate model output from 18 models was obtained from the CMIP5 data archive. The long-term drift is removed from the perturbed CMIP5 model data by subtracting the linear trend from the pre-industrial control run (Gupta et al. 2013).

Initially, we are investigating the range of responses from stylized CMIP5 experiments (e.g. 1% yr^{-1} CO_2 increase), which can elucidate changes in climate

responses on certain time scales. The stylized CMIP5 experiments allow the opportunity to better quantify model noise in our sub-global analysis.

The range of responses within CMIP5 GHG-only or full-forcing runs can be defined using two climate variables, the rate of temperature response over time (RT_t) and land-ocean warming ratios (ϕ). RT_t is the linear fit of drift adjusted temperature series, T_{adj}. The drift adjusted temperature series is given by the following equation:

$$T_{adj} = T_s(t) - (\delta T_n(t) \times t) \tag{5.1}$$

where $T_s(t)$ is the temperature at time t from the 1pctCO_2 experiment, δT_n is the linear slope of the full temperature series from the corresponding piControl experiment, and t is time. RT_t is useful for understanding the climate response on shorter time scales than other climate variables such as TCR, as CO_2 concentrations double in 70 years (Millar et al. 2015; see Table 9.5 from Flato et al. 2014). Using RT_t we can investigate the climate response on a decadal time scale (20-30 years), capturing both the near-term climate response and the changing response over time.

Consistently processing data to sub-global spatial scales allows comparisons of the temperature responses in four major boxes to understand the spatial differences within the models. The four hemispheric boxes are: NH, SH, land, and ocean. Because aerosols, a type of SLCF, are predominately influential over land, with the majority located in the Northern Hemisphere, comparing land and ocean responses of hemispheric responses can imform the range of responses within the CMIP5 models. Thus, RT_t can be used in combination with a well-known climate change

variable, land-sea surface warming ratio (ϕ), a property which describes the rate the land surface warms, compared to the ocean, under climate change conditions. ϕ is given by the following equation:

$$\phi = \frac{RT_{tland}}{RT_{tocean}} \tag{5.2}$$

Because of the larger land surface in the NH, this hemisphere exhibits stronger ϕ supported by both models and observations of the 20th century (Joshi et al. 2013; Lambert and Chiang 2007; Sutton et al. 2007).

Therefore, while RT_t provides understanding of the temporal response changes over time in our analysis, ϕ is a significant geographical variable for determining the pattern of climate response (Geoffroy et al. 2015; Joshi et al. 2013). By explicitly examining the spatial and temporal evolution of our two climate variables, we can explore the range of responses in coupled models to inform our understanding of short term climate responses and ultimately, SLCFs.

5.4 Preliminary Results

Here we report our preliminary results from analyzing the spread of climate responses from the CMIP5 1pctCO_2 experiment. The literature suggests that when spatially varying climate forcers, such as SLCFs, constitute a large fraction of the climate response, ϕ might change over time (Joshi et al. 2013; hereafter Joshi et al.). Figure 5.1 shows our results for ϕ over time (in years) for 18 CMIP5 models. The spread in ϕ varies over the time series, with significant model spread early in

the time series. Numerically, when ϕ has a large spread early in the time series, the variability comes from changes in ocean temperature, the denominator of the ratio (see Equation 5.2).

The spread narrows over the time series and converges to an average value of $\phi = 1.55$. Joshi et al. conducted a similar analysis of the 1pctCO_2 runs with 16 models and found the models converge to $\phi = 1.6$. Our ϕ value range, 1.32-1.65, is within the range found from observations (Drost et al. 2012), though it differs from the range found by Joshi et al., 1.4-1.9. The differences can be accounted for in model choices and methodology that may increase the spread in the Joshi et al. range. While Joshi et al. (2008) defined ϕ as the ratio of the difference in area-weighted GMST, our method followed Equation 5.2.

Similar RT_t ratio comparisons can be made using the NH and SH. Figure 5.2 shows the ratio of the NH RT_t to the SH RT_t in same 18 CMIP5 models presented in Figure 5.1. Compared to the ϕ results, there is more model spread in the NH/SH comparison and a less delineated convergence point. The results show a large range of 1.15-1.81, though the majority of models still center above or below a value of 1.5. Because the range of responses are greater than 1, the NH response has more variability, especially early in the time series.

Each of the hemispheric ratios presented in Figure 5.1 and Figure 5.2 are summarized in Figure 5.3 and Figure 5.4. Figure 5.3 shows the hemispheric warming ratio (NH/SH, land/ocean) plotted with increasing transient climate response (TCR; left to right) values from each CMIP5 model investigated in our 1pctCO_2 analysis. The model responses varied, with four models showing a greater NH/SH hemispheric

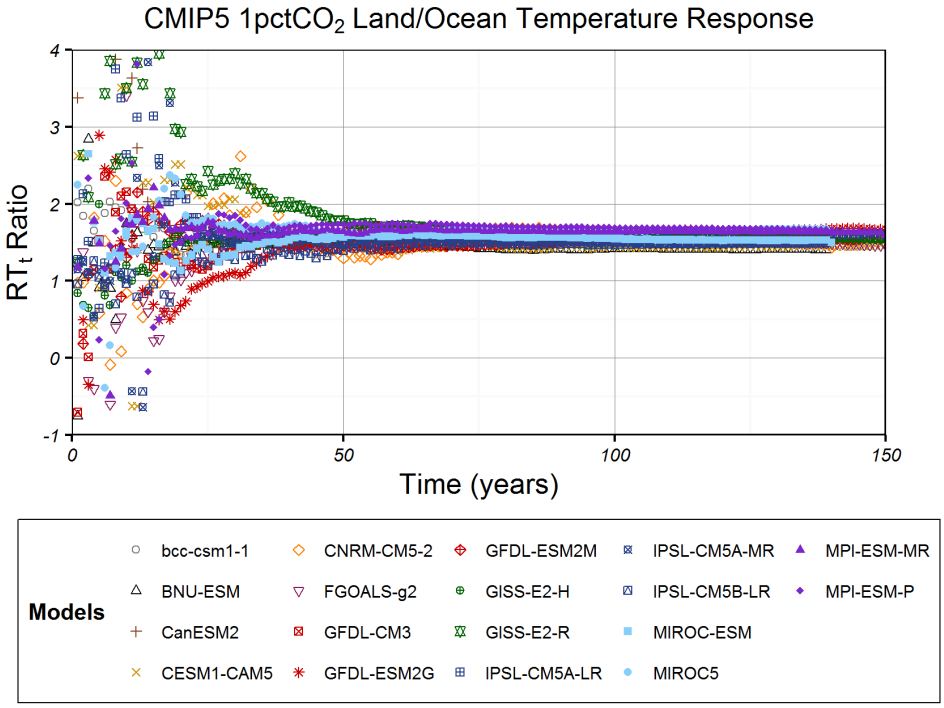


Fig. 5.1: Ratio of land to ocean warming produced from CMIP5 1pctCO_2 RT_t series.

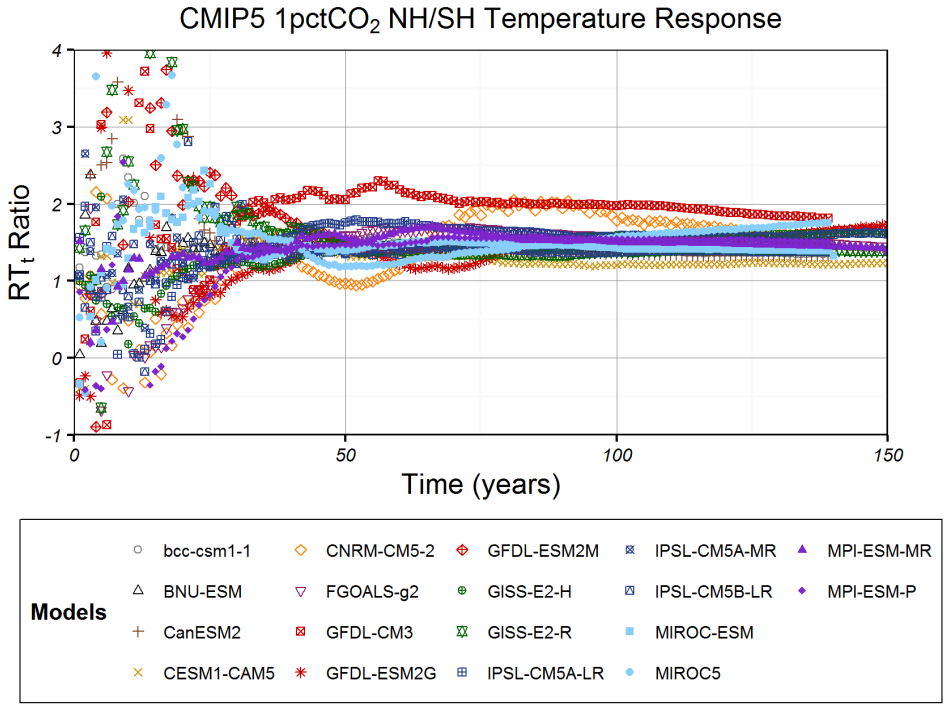


Fig. 5.2: Ratio of NH to SH warming produced from CMIP5 1pctCO_2 RT_t series.

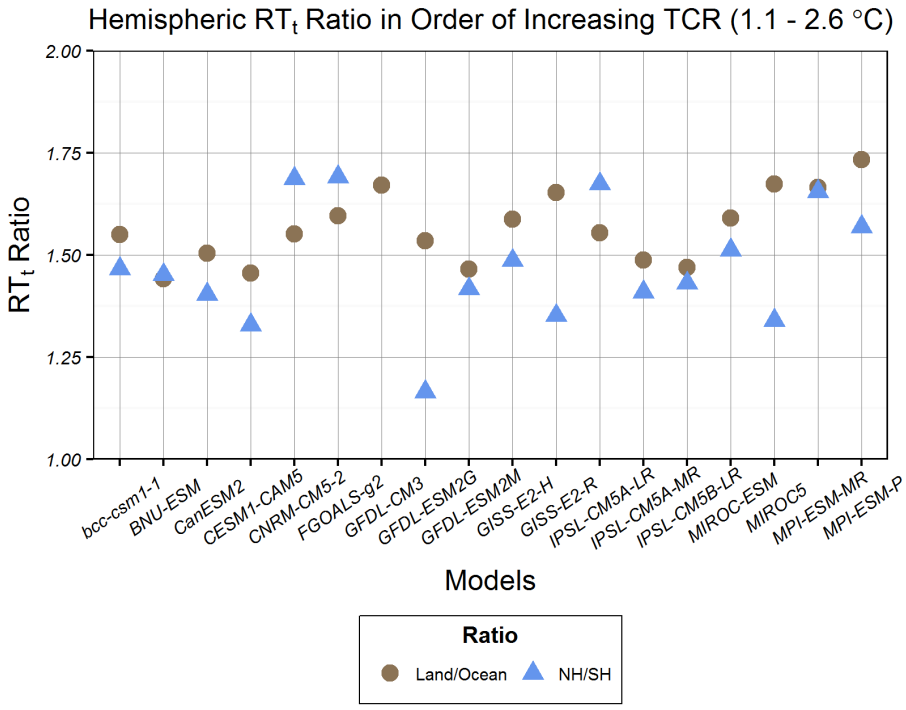


Fig. 5.3: CMIP5 model comparison of the regional (NH/SH, land/ocean) RT_t ratios in order of increasing transient climate response values (TCR in °C).

ratio than a land/ocean ratio.

Figure 5.4 shows a box and whisker plot of the hemispheric response ratios. The CMIP5 model responses show a much larger range in NH/SH ratio as compared to the land/ocean ratio. We can compare our NH/SH ratio results to a similar study conducted by Shindell et al. (2014). In our analysis, we calculated the RT_t from two hemispheric levels, NH/SH and the land/ocean. Unlike Shindell et al. (2014), we did not specifically investigate the extratropics, however, the ranges of the results are similar.

Other spatial areas of interest include the Arctic and Antarctic; where signal to

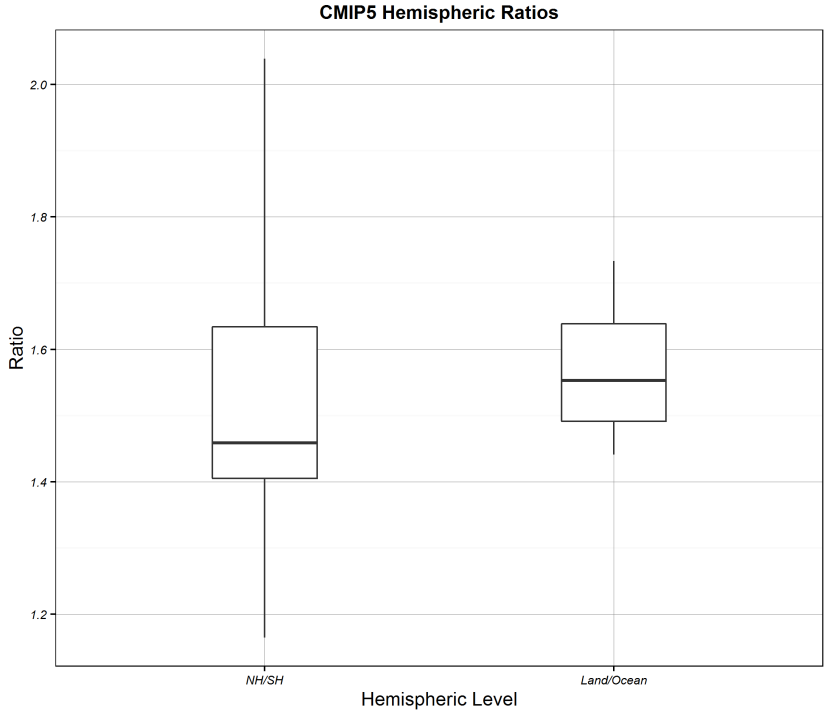


Fig. 5.4: Mean and range of the hemispheric (NH/SH, land/ocean) RT_t ratios. The outer edge of the box represents the 25th and 75th quartile, the thick black line is the median value, the range is represented by the whiskers of the plot.

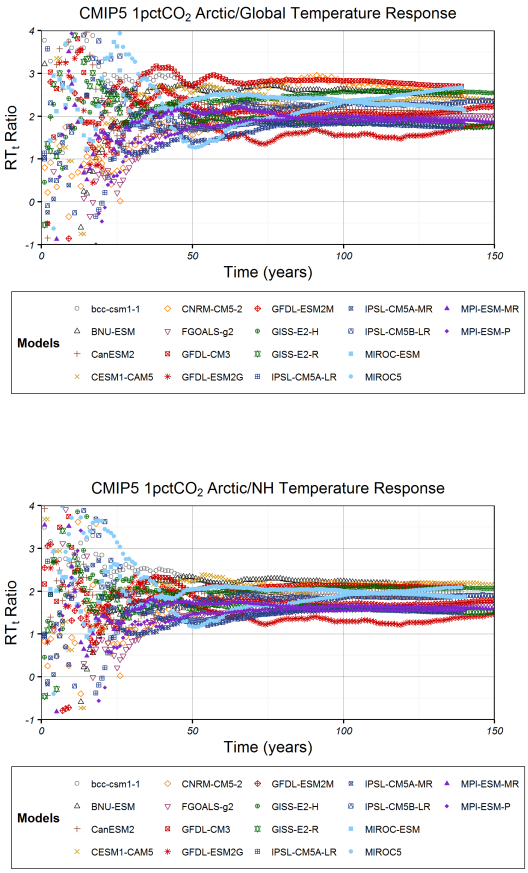


Fig. 5.5: Ratio of Arctic hemispheric-level warming produced from CMIP5 1pctCO_2 RT_t series.

noise ratios will present challenges when interpreting values of RT_t and ϕ. However, given the well-documented climate impacts from SLCFs investigating responses in the polar region can provide interesting results and insights into climate responses from near-term climate change (Bond et al. 2013; Flanner 2013; Schmale 2016). Figure 5.5 shows the Arctic/Global RT_t ratio (top) and the Arctic/NH RT_t ratio (bottom) with a very large spread throughout the time series.

We see that the signal to noise ratio is larger in Figure 5.6, which shows

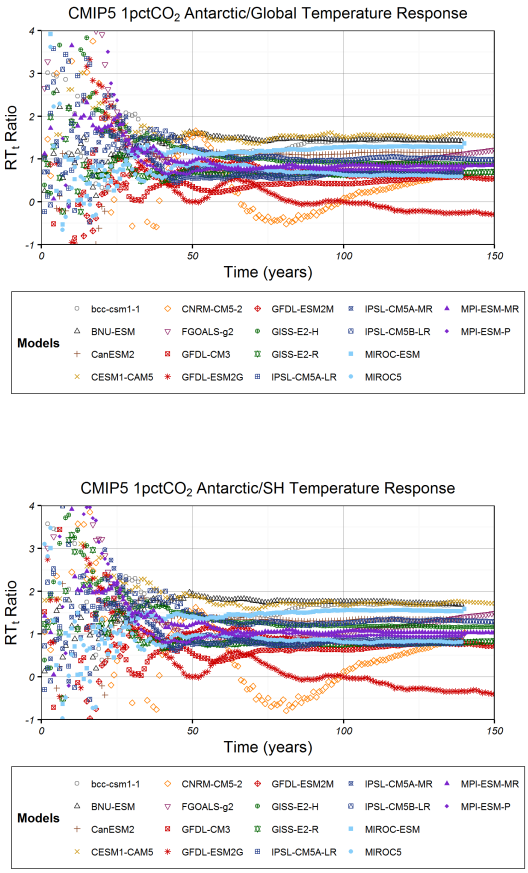


Fig. 5.6: Ratio of Antarctic hemispheric-level warming produced from CMIP5 1pctCO_2 RT_t series.

the Antarctic/Global RT_t ratio (top) and the Antarctic/SH RT_t ratio (bottom) compared to Figure 5.5. The models have very little agreement in response, with one model having a negative RT_t. However, the magnitude of the response for the Antarctic/Global is similar to the magnitude of the Antarctic/SH response.

The large amount of noise in Figure 5.5 and Figure 5.6 illustrate the need for a robust analysis of the 1pctCO_2 RT_t series. To understand when the 1pctCO_2 time series becomes a true signal, we conducted a signal to noise analysis using

the corresponding piControl experiments. We identified the climate change signal to be the hemispheric-level 1pctCO_2 time series and the variability to derive from 100-year random time sequences of the piControl experiment.

Figure 5.8 shows the time of emergence (ToE) values plotted for each model. The shaded portion representing the 5-95% range of the noise (piControl experiment), while the black line shows the global 1pctCO_2 series (signal). We defined the ToE as the first year in which the signal permanently exceeds the 5-95% range of the noise threshold (Hawkins and Sutton 2012). The red dashed line represents the ToE for the global 1pctCO_2 series. The majority of the model signals emerge from the noise before 20 years, the decadal period of interest in this work, suggesting that we can be confident in our signal when comparing near-term climate responses at 20 years to longer-term responses at 70 years.

However, our analysis does not allow us to understand if the temperature response exceeds the internal model variability. What the analysis in Chapter 4 has shown, the noise envelope analysis here might not be robust for models with shorter control runs (and the model control runs are likely not long enough for many of these models). Identifying where this is the case is the last step that will be needed for this analysis.

To understand the factors driving the emergence of the forced signal, we compared: (1) the half-width of the noise spectrum, (2) the hemispheric ratios, and (3) the short-term and longer-term response, to the $RT_t(70)$ x 70. This response value was found by multiplying the value the $RT_t(70)$ by time $(t) = 70$, the time at $[CO_2]$ doubling. We calculated this response value because it is comparable to

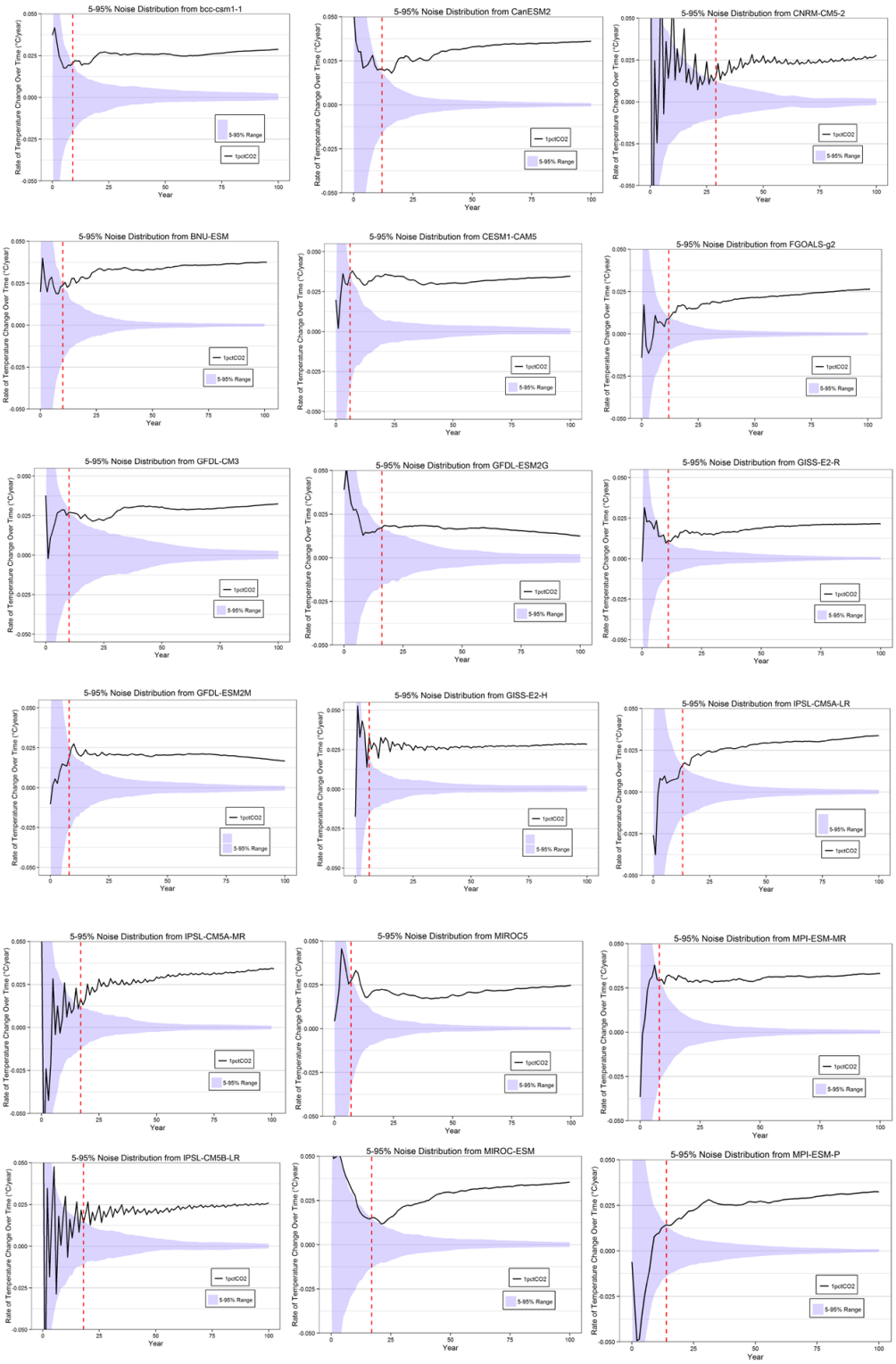


Fig. 5.7: Signal to noise analysis, with the red dashed line showing the ToE for each model.

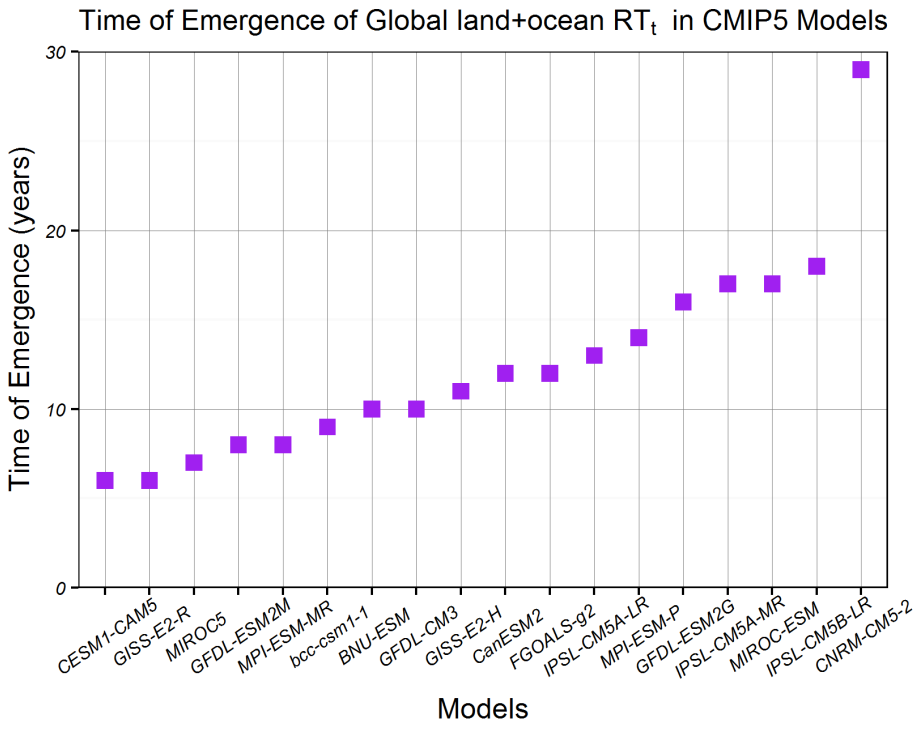


Fig. 5.8: Time of Emergence (in years) values plotted for the global-level RT_t series from each of the CMIP5 model used in this study.

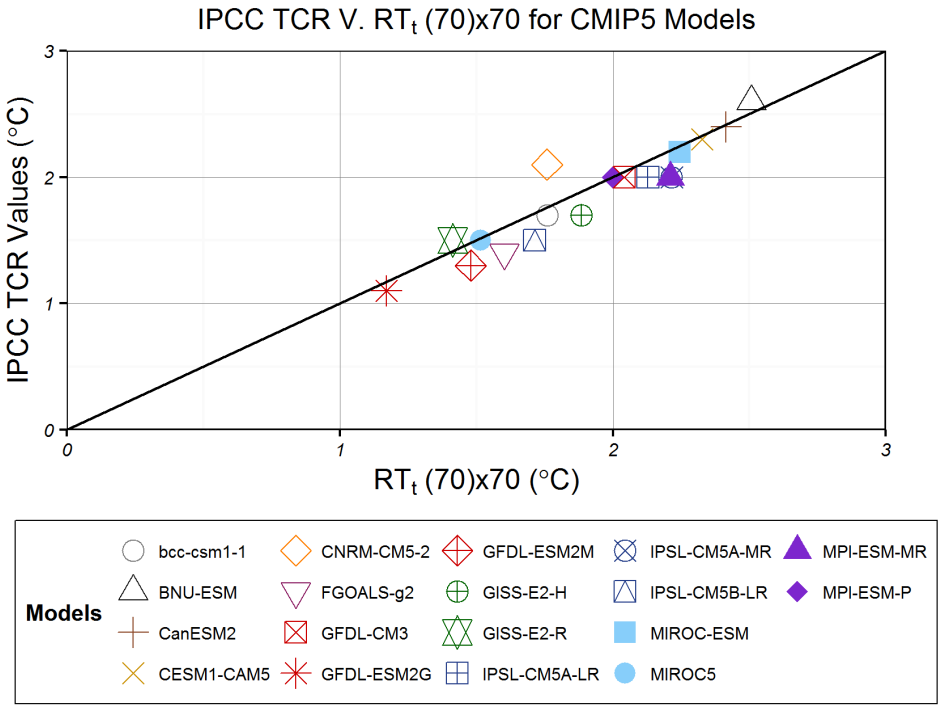


Fig. 5.9: IPCC Transient Climate Response values compared to the $RT_t(70)$ multiplied by $t = 70$ for each CMIP5 model. The one-to-one line is in black.

the reported TCR values in Table 9.5 of the Intergovernmental Panel on Climate Change's Fifth Assessment Report (IPCC AR5; Stocker et al. 2013). Figure 5.9 illustrates that our calculated values are close to the reported TCR values. The majority of the models fall close to the one-to-one line (black). The differences in the values can be accounted for because we are using a linear fit to obtain the RT_t values, rather than the 20 year mean centered on the year of $[CO_2]$ doubling, as defined by the IPCC.

The analysis (Appendix 7 B: Correlation Analysis) showed no correlation of $RT_t(70)$ to the ToE, suggesting it is still unclear the factors driving the signal

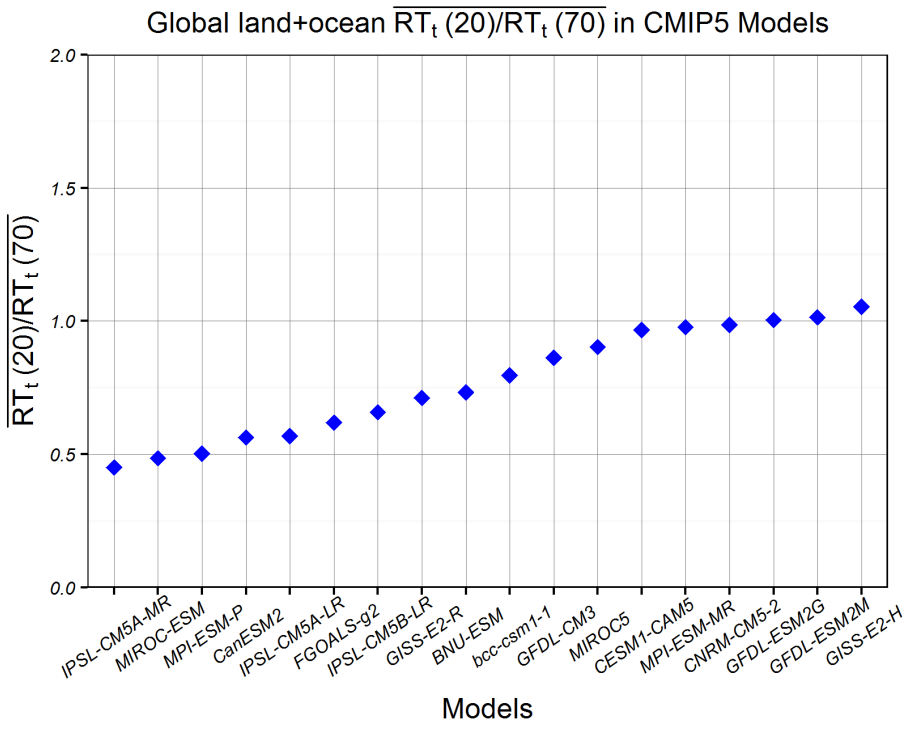


Fig. 5.10: Global ratio of near-term, $\overline{RT_t(20)}$, to longer-term, $\overline{RT_t(70)}$, response for each CMIP5 model.

emergence (see Appendix 7 B: Correlation Analysis).

From this work, we can begin to classify the 18 CMIP5 models used in this study according to their response. We compared the ratio of the 5-year average of $RT_t(20)/RT_t(70)$, given by $\overline{RT_t(20)}/\overline{RT_t(70)}$ for each model to the hemispheric levels. Figure 5.10 shows the ratio of the near-term response, $\overline{RT_t}$ at $t = 20$, to the longer-term response, $\overline{RT_t}$ at $t = 70$. The results show a group of 9 models within 20% of $\overline{RT_t(20)}/\overline{RT_t(70)} = 1$. The remaining 9 models have a distinctly lower response at 20 years versus 70 years.

5.5 *Conclusions and Work Plans*

Here we have used CMIP5 model results from a CO_2 forced experiment to investigate the range of spatial and temporal climate responses within the models. Figure 5.10 shows a factor of two range in the temporal response of the near-term, $\overline{RT_t(20)}$, to longer-term, $\overline{RT_t(70)}$, which is previously not seen in the literature. Our next step is to determine the factors driving the large range in models responses in the short-term, compared to the longer-term. The literature suggestes investigating ocean heat flux in the CMIP5 models might provide one path forward. In particular, Gregory et al. (2004) (hereafter Gregory et al.) used the linear regression of the top of the atmosphere radiative forcing and the effective temperature to estimate the effective climate sensitivity under a highly non-linear regime (i.e. $4\text{x}CO_2$ experiment). Known as the Gregory method, the results produced a concave scatter plot pattern, illustrating that climate sensitivity increases as the climate warms. Since the Gregory et al. publication, it has been noted in the literature that the concavity of climate sensitivity over time in the CMIP5 models may suggest an active role in the ocean on the decadal scale.

The Gregory et al. results hinted at the possible short-term implications for the concave pattern of response. Using our methodology, we are able to investigate short-term decadal responses in a linear CO_2 forcing regime without relying on both temperature and limited forcing data. Furthermore, we produced results showing the change in response over time at a sub-global scale (Figure 5.7), illustrating the robustness of our methods.

Chapter 6: Concluding Remarks and Future Work

6.1 Concluding Remarks

The following overarching findings are provided to the scientific community by the completion of this thesis:

6.1.1 Fundamental Impulse response tests in SCMs

SCMs are widely used in the literature and in decision-making context, e.g., within Intergovernmental Panel on Climate Change(IPCC) Reports, coupled with Integrated Assessment Models. A paper describing a commonly used SCM, MAGICC 6.0, has been cited 371 times in the literature and policy contexts. Another model, the impulse response model used in the IPCC Fifth Assessment Report (AR5-IR), is heavily used by the scientific community to support decision making.

Despite their importance, the fundamental responses of SCMs are not fully characterized. The U.S. National Academies of Science (2016) specifically suggested that SCMs be, "assessed on the basis of the response to a pulse of emissions," which we do here. Additionally, we provide a set of tests that we recommend as a standard validation suite for any SCM.

SCMs range in complexity from comprehensive SCMs (e.g., MAGICC 6.0)

to idealized SCMs (e.g., AR5-IR) and we test SCMs in both categories. We find that idealized SCMs often fail to capture the responses of more complex Earth system models. Biases in the model responses to perturbations will affect reported climate metrics, such as Global Warming Potential, and subsequent climate policy analysis. While SCMs are able to capture some responses of more complex models, comprehensive SCMs still do not fully capture the response time scales from BC perturbations found in more complex models, for example, indicating such SCM responses should be reevaluated. We suggests improvements should be made to SCMs, which affect numerous scientific endeavors and illustrates the necessity of integrating unit tests into SCM development.

6.1.2 Characterizing unforced variability in CMIP5 models

Contributions from both unforced variability (e.g., ocean-atmosphere interactions) and external forcing (e.g., anthropogenic greenhouse gas emissions, volcanic activity, variations in solar intensity) drive changes in Earth's global mean surface temperature. Understanding the relative contributions from these different components is of great interest to the scientific community (e.g., detection and attribution studies, time of emergence estimates). Though many studies draw conclusions using estimates of, or assumptions about, unforced variability using multi-model ensembles, a robust characterization of model-specific unforced model variability has not been conducted.

We fill this gap by determining how realistic unforced model variability in complex models is compared to observations for regions and time periods important

to human systems. From this, we can determine how realistic complex model variability is compared to observations. We compare CMIP5 model results from the past1000, PiControl, and Historical experiments with two historical observational datasets(GISTEMP, HadCRUT) and paleoclimate reconstructions from PAGES2k. We first examine variability in those CMIP5 models with sufficient data (e.g., published past1000 output and long PiControl runs).

We use power spectra of temperature-time series to compare variability across all time scales. We also investigate model variability at the regional level, using the sub-global regions defined by PAGES2k. Our approach allows for a more robust assessment of complex model variability at time periods and regional levels important to human systems.

6.1.3 Changes in Climate Response Over Time

The time scales of climate system responses to anthropogenic emissions vary depending on the chemical species emitted. Though carbon dioxide(CO_2) emissions primarily drive anthropogenic climate change, emissions of various other radiative forcing agents, including short-lived climate forcers (SLCFs), also contribute significantly to Earth's altered radiative budget. Much of the literature focuses on long-term climate responses emphasizing analysis with equilibrium climate sensitivity (ECS) or transient climate sensitivity (TCR). We seek to clarify climate dynamics on decadal time scales with the ultimate goal of understanding the implications of near-term emission reductions on climate. Using observations and coupled model results from CMIP5, we analyze the range of temperature response

over time, with specific attention to the 20-30 year time sequence. Similarly, we also explore sub-global temperature responses at a hemispheric-scale. We find that the range of responses of land/ocean varied less than the range of hemispheric responses. Our results are a first step of better quantifying the short-term climate responses to change in SLCFs.

6.2 Future Work

Though global emphasis has been placed on long-term emission mitigation strategies via CO_2 emissions reductions, short-term climate mitigation strategies are a potentially valuable option for addressing anthropogenic climate change, with importance placed on reducing the rate of climate change in the near-term.

Short-lived climate forcers (SLCFs) generated over the land surface, including aerosols, play a significant role in modifying climate (Fuzzi et al. 2015; Ødemark et al. 2012; Schmale 2016; Shindell et al. 2012). Atmospheric circulations allow longer-lived constituents and CH_4 to have relatively uniform concentrations throughout the atmosphere; SLCFs are more quickly removed via chemical and physical reactions. The SLCFs discussed in this dissertation, BC and CH_4, have atmospheric lifetimes of 4-12 days and at steady-state, 12 years, respectively (Bond et al. 2013; Lelieveld et al. 1993). In an effort to reduce the unprecedented rate of temperature increase in recent history, there is a growing interest in near-term climate mitigation strategies specifically targeting SLCFs (Pierrehumbert 2014; Rogelj et al. 2014; Shindell 2014; Smith and Mizrahi 2013).

Addressing near-term emission reductions, however, requires understanding climate impacts on shorter time scales. Utilizing the most recent and commonly used climate model data set, CMIP5, to investigate SLCFs influences presents its own challenges. Perturbing the CMIP5-class models requires large computing resources and generates noisy output. Instead, climate metrics, such as the rate of temperature response over time and land-ocean warming ratios, can be used define the range of responses within CMIP5 GHG-only or full-forcing runs.

Emulating the range of responses from CMIP5 climate metrics in SCMs can be a useful tool for clarifying the role SLCFs play in future climate. Reduced form models, or SCMs, are computationally efficient—simulating thousands of scenarios in minutes—with representations of the most fundamental climate components, such as SLCF or ocean heat uptake. Though recent literature investigated the hemispheric differences in the CMIP5 experiments and inferred that SCMs would do a poor job at representing realistic spatial temperature responses due to the omission of spatial forcing details in these SCMs, some models, such as the Model for Greenhouse Gas Induced Climate Change (MAGICC), take this differential hemispheric forcing into account.

Future studies could develop temperature uncertainty ranges illustrating possible responses to changes in short-lived climate forcer (SLCF) emissions, among other climate factors. Specifically, we could explore the implications of CH_4 and BC emissions reductions on climate in the next 20-30 years. Additionally, we can change parameters within simplified models (e.g. Land-ocean warming ratio, Land-ocean exchange coefficient, NH-SH exchange coefficient, global ocean heat diffusivity value,

the upwelling rate), informed by observations and stylized CMIP5 experiments, to capture ranges and uncertainty in the climate responses. This dissertation can inform climate policies by contributing uncertainty ranges produced from SCMs, which may lead to future improvements in model representations of SLCFs within the models.

Further, following the cause-effect chain, we note that changes in emissions ultimately cause climate impacts, which can be translated to economic losses expressed in monetary units (e.g., dollars; Bosello et al. 2006; Pearce et al. 1996; Tol 2003). Additional future work could estimate economic damages from climate impacts depend on a litany of assumptions (e.g., spatial and temporal scale, exposure dosage assumptions, valuation techniques), hence the increasing uncertainty down the chain. Nonetheless, understanding climate impacts and placing them in the context of economic consequences has increasing relevance to human systems and provides important context to decision makers. Though outside the scope of this dissertation, the fundamental questions addressed here do have implications for the scientific community's attempts to meaningfully address climate change.

6.2.1 CMIP5 Emulation in MAGICC 5.3

In future work, one can investigate the range of CMIP5 responses in MAGICC 5.3. Previous studies emulated CMIP-class model results in MAGICC. Raper et al. (1996; 2001) used MAGICC 5.3, an upwelling diffusion (UD) model from Wigley and Raper (1987), to emulate results from globally coupled ocean atmosphere general circulation models (O/AGCMs; Raper and Cubasch 1996; Raper et al. 2001; Wigley

and Raper 1987).

Raper et al. (2001) varied parameters such as climate sensitivity, Land-ocean warming ratio, Land-ocean exchange coefficient, NH-SH exchange coefficient to produce good simulations of the land/ocean and NH/SH temperature change contrasts present in the HadCM2 results. Further, the authors altered the global ocean heat diffusivity value (κ), the upwelling rate, and the parameter determining the temperature change of incoming bottom water. These three parameters can be varied to produce changes in the ocean heat uptake and were used by Raper et al. (2001) to emulate the thermal expansion from the HadCM2 results.

MAGICC 5.3 is of particular interest in this study because the model has user-defined parameters, such as climate sensitivity and ϕ, with regional temperature output from the four major boxes investigated in our CMIP5 response runs: global, ocean, land, and hemispheres (Wigley and Raper 1987). Similar to the Raper et al. studies, we can adjust parameters such as the land-ocean warming ratio, land-ocean exchange coefficient, NH-SH exchange coefficient to simulate the hemispheric-level temperature response found in Chapter 5 (Raper and Cubasch 1996; Raper et al. 2001).

Figure 6.1 is a preliminary investigation of the model response in MAGICC 5.3. Using the same analysis as Figure 5.10, Figure 6.1 shows the ratio of near-term, $RT_t(20)$, to longer-term, $RT_t(70)$, response for each CMIP5 model (diamonds) and MAGICC 5.3 (squares) under different climate sensitivities (high = 4.5 °C, mid = 3 °C, low = 2 °C). Even with simple changes to MAGICC 5.3 climate sensitivity values, we are able to capture a narrow range of CMIP5 responses.

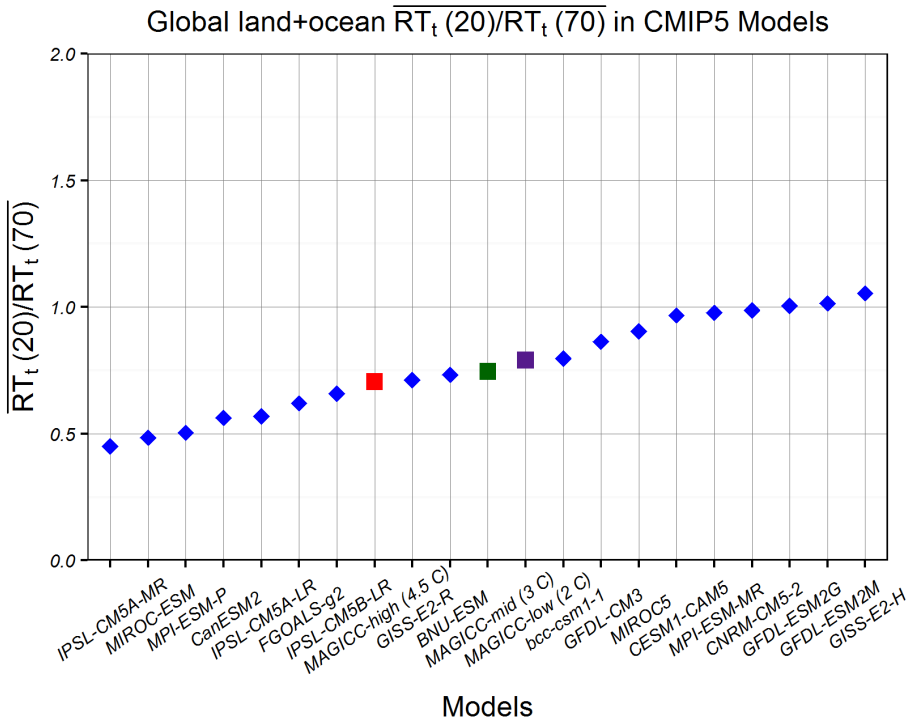


Fig. 6.1: $\overline{RT_t(20)}/\overline{RT_t(70)}$ response for each CMIP5 model and MAGICC5.3 for different climate sensitivities (high = 4.5 °C in red, mid = 3 °C in green, low = 2 °C) in purple.

In future work, one could change the global ocean heat diffusivity value (k), the upwelling rate ($W0$), and the parameter determining the temperature change of incoming bottom water (PI or π). Altering these values may allow us to more closely reproduce not only the full range of the CMIP5 results, but the fundamental climate dynamics driving the responses found in Chapter 5.

6.2.2 Understanding Climate Response to SLCFs

To understand the implications of SLCF emissions reductions we need to relate the climate responses found under the CMIP5 CO_2 forced experiments to SLCFs. CH_4 is often categorized as a SLCF, though it is a WMGHG with similar spatial forcing to CO_2. Therefore, results produced under the CO_2 forced experiments can be directly mapped to CH_4. Reducing CH_4 emissions has recently garnered attention from the scientific community as a means to limit near-term climate change. However, the range of results in Figure 5.10 suggest that near-term climate mitigation of CH_4 may not be as beneficial as previously expressed in the literature. We will use SCMs to explore changes in CH_4 emissions, investigating the implications of such emission reductions on climate.

Similarly, one could use SCMs to investigate changes in aerosol emissions, particularly BC. Aerosols, are inhomogenous forcers with differential impacts over land and the Northern Hemisphere, where the largest concentration of aerosol emissions occurs. It is commonly known in the literature that aerosol RF estimates have large uncertainties. Our efforts have added additional levels of uncertainty by uncovering the large range in $\overline{RT_t(20)}/\overline{RT_t(70)}$, NH/SH, and land/ocean results. By

coupling changes in BC emissions with user-defined parameters, we can investigate the climate response to changes in emissions in the near-term.

SCMs are computationally efficient, so we will be able to investigate numerous pathways while changing the parameter space, and produce results quickly. These features provide us the opportunity to capture ranges and uncertainty in the SCMs results, which has implications for decision support.Reducing SLCFs emissions has attracted attention from both the scientific and policy community as a potential method for reducing near-term climate change. Furthermore, SCMs are often used to inform possible ranges of climate responses under a multitude of possible climate futures. By exploring our three sources of uncertainty in the SCMs, we can develop temperature uncertainty ranges illustrating possible responses to changes in SLCF emissions, among other climate factors.

APPENDIX

Chapter 7: Appendix

7.1 Additional Results for Chapter 4

Here we present additional results and figures from Chapter 4.

7.1.1 Additional CMIP5 Data Description

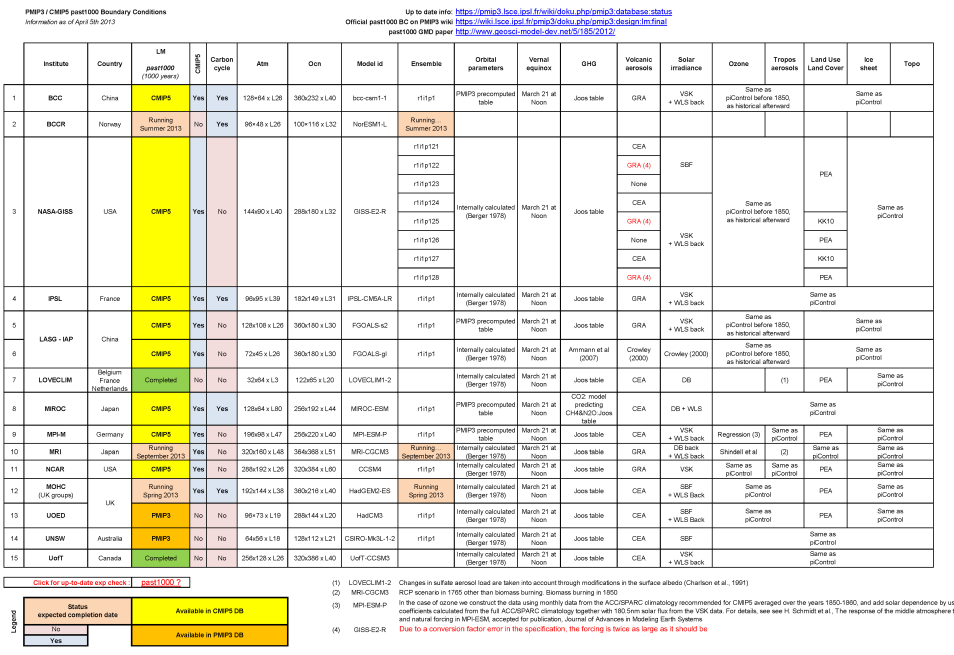
Table 7.1 provides additional details of the past1000 CMIP5 data used in Chapter 4 of this dissertation.

7.1.2 Control Run Variability

In Table 7.2, we present the numerical values from Figure 4.5. We cite the halfwidths through the chapter to characterize internal climate model variability and to explore the robustness of our analyses.

7.2 Additional Results for Chapter 5

Here we present additional results and figures from Chapter 5.



PMIP3 / CMIP5 past1000 Boundary Conditions
Information as of April 5th 2013

Up to date info: https://pmip3.lsce.ipsl.fr/wiki/doku.php/pmip3:database:status
Official past1000 BC on PMIP3 wiki https://wiki.lsce.ipsl.fr/pmip3/doku.php/pmip3:design:lm:final
past1000 GMD paper http://www.geosci-model-dev.net/5/185/2012/

	Institute	Country	LM *past1000* (1000 years)	CMIP5	Carbon cycle	Atm	Ocn	Model id	Ensemble	Orbital parameters	Vernal equinox	GHG	Volcanic aerosols	Solar irradiance	Ozone	Tropos aerosols	Land Use Land Cover	Ice sheet	Topo
1	**BCC**	China	**CMIP5**	**Yes**	**Yes**	128×64 x L26	360x232 x L40	bcc-csm1-1	r1i1p1	PMIP3 precomputed table	March 21 at Noon	Joos table	GRA	VSK + WLS back	Same as piControl before 1850, as historical afterward			Same as piControl	
2	**BCCR**	Norway	Running Summer 2013	No	**Yes**	96×48 x L26	100×116 x L32	NorESM1-L	Running... Summer 2013										
3	**NASA-GISS**	USA	**CMIP5**	**Yes**	No	144x90 x L40	288x180 x L32	GISS-E2-R	r1i1p121	Internally calculated (Berger 1978)	March 21 at Noon	Joos table	CEA	SBF	Same as piControl before 1850, as historical afterward		PEA	Same as piControl	
									r1i1p122				GRA (4)						
									r1i1p123				None						
									r1i1p124				CEA	VSK + WLS back			KK10		
									r1i1p125				GRA (4)						
									r1i1p126				None				PEA		
									r1i1p127				CEA				KK10		
									r1i1p128				GRA (4)				PEA		
4	**IPSL**	France	**CMIP5**	**Yes**	**Yes**	96x95 x L39	182x149 x L31	IPSL-CM5A-LR	r1i1p1	Internally calculated (Berger 1978)	March 21 at Noon	Joos table	GRA	VSK + WLS back	Same as piControl				
5	**LASG - IAP**	China	**CMIP5**	**Yes**	No	128x108 x L26	360x180 x L30	FGOALS-s2	r1i1p1	PMIP3 precomputed table	March 21 at Noon	Joos table	GRA	VSK + WLS back	Same as piControl before 1850, as historical afterward		Same as piControl		
6			**CMIP5**	**Yes**	No	72x45 x L26	360x180 x L30	FGOALS-gl	r1i1p1	Internally calculated (Berger 1978)	March 21 at Noon	Ammann et al (2007)	Crowley (2000)	Crowley (2000)	Same as piControl before 1850, as historical afterward		Same as piControl		
7	**LOVECLIM**	Belgium France Netherlands	Completed	No	No	32x64 x L3	122x95 x L20	LOVECLIM1-2		Internally calculated (Berger 1978)	March 21 at Noon	Joos table	CEA	DB		(1)	PEA	Same as piControl	
8	**MIROC**	Japan	**CMIP5**	**Yes**	**Yes**	128x64 x L80	256x192 x L44	MIROC-ESM	r1i1p1	PMIP3 precomputed table	March 21 at Noon	CO_2: model predicting. CH_4&N_2O: Joos table	CEA	DB + WLS	Same as piControl				
9	**MPI-M**	Germany	**CMIP5**	**Yes**	No	196x98 x L47	256x220 x L40	MPI-ESM-P	r1i1p1	PMIP3 precomputed table	March 21 at Noon	Joos table	CEA	VSK + WLS back	Regression (3)	Same as piControl	PEA	Same as piControl	
10	**MRI**	Japan	Running September 2013	**Yes**	No	320x160 x L48	364x368 x L51	MRI-CGCM3	Running September 2013	Internally calculated (Berger 1978)	March 21 at Noon	Joos table	GRA	DB back + WLS back	Shindell et al	(2)	Same as piControl	Same as piControl	
11	**NCAR**	USA	**CMIP5**	**Yes**	No	288x192 x L26	320x384 x L60	CCSM4	r1i1p1	Internally calculated (Berger 1978)	March 21 at Noon	Joos table	GRA	VSK	Same as piControl	Same as piControl	PEA	Same as piControl	
12	**MOHC** (UK groups)	UK	Running Spring 2013	**Yes**	**Yes**	192x144 x L38	360x216 x L40	HadGEM2-ES	Running Spring 2013	Internally calculated (Berger 1978)	March 21 at Noon	Joos table	CEA	SBF + WLS Back	Same as piControl		PEA	Same as piControl	
13	**UOED**		**PMIP3**	No	No	96×73 x L19	288x144 x L20	HadCM3	r1i1p1	Internally calculated (Berger 1978)	March 21 at Noon	Joos table	CEA	SBF + WLS Back	Same as piControl		PEA	Same as piControl	
14	**UNSW**	Australia	**PMIP3**	No	No	64x56 x L18	128x112 x L21	CSIRO-Mk3L-1-2	r1i1p1	Internally calculated (Berger 1978)	March 21 at Noon	Joos table	CEA	SBF	Same as piControl				
15	**UofT**	Canada	Completed	No	No	256x128 x L26	320x386 x L40	UofT-CCSM3		Internally calculated (Berger 1978)	March 21 at Noon	Joos table	CEA	VSK + WLS back	Same as piControl				

Click for up-to-date exp check: past1000.7

Legend

Status expected completion date	Available in CMIP5 DB
No	
Yes	Available in PMIP3 DB

(1) LOVECLIM1-2 Changes in sulfate aerosol load are taken into account through modifications in the surface albedo (Charlson et al., 1991)
(2) MRI-CGCM3 RCP scenario in 1765 other than biomass burning. Biomass burning in 1850
(3) MPI-ESM-P In the case of ozone we construct the data using monthly data from the ACCI/SPARC climatology recommended for CMIP5 averaged over the years 1850-1860, and add solar dependence by using regression coefficients calculated from the full ACCI/SPARC climatology together with 180.5nm solar flux from the VSK data. For details, see see H. Schmidt et al., The response of the middle atmosphere to anthropogenic and natural forcing in MPI-ESM, accepted for publication, Journal of Advances in Modeling Earth Systems
(4) GISS-E2-R Due to a conversion factor error in the specification, the forcing is twice as large as it should be

Tab. 7.1: Additional past1000 CMIP5 data description from PMIP wiki (Schmidt et al. 2012).

		Average spectral density and halfwidth for each average period bin ($°C^2$ year)						
Region	**Model**	**02-05 (years)**	**05-10 (years)**	**10-20 (years)**	**20-30 (years)**	**30-50 (years)**	**50-70 (years)**	**70-100 (years)**
Australasia	CCSM4	1.1±0.12	0.98±0.12	0.74±0.22	0.71±0.53	0.74±0.24	0.82±0.28	0.70±0.41
	CSIRO-Mk3L-1-2	0.45±0.018	0.35±0.058	0.32±0.079	0.34±0.090	0.35±0.089	0.34±0.13	0.35±0.078
	HadCM3	1.6±0.090	1.7±0.22	1.5±0.29	1.7±0.62	1.8±0.63	1.6±0.46	1.6±0.42
	IPSL-CM5A-LR	0.47±0.030	0.36±0.052	0.37±0.099	0.36±0.12	0.41±0.14	0.33±0.14	0.36±0.11
	MIROC-ESM	1.3±0.032	0.79±0.049	1.3±0.079	2.4±0.19	1.9±0.15	1.6±0.25	1.8±0.64
	MPI-ESM-P	1.0±0.15	1.3±0.30	1.5±0.39	1.3±0.47	1.3±0.34	1.3±0.34	1.6±0.49
Antarctica	CCSM4	0.98±0.072	0.71±0.062	0.61±0.095	0.57±0.14	0.67±0.30	0.53±0.21	0.56±0.32
	CSIRO-Mk3L-1-2	0.65±0.047	0.69±0.088	0.78±0.11	0.46±0.11	0.55±0.11	0.49±0.13	0.46±0.27
	HadCM3	1.5±0.27	1.5±0.18	1.4±0.23	1.5±0.80	1.5±0.50	1.4±0.35	0.95±0.42
	IPSL-CM5A-LR	1.1±0.081	1.4±0.083	1.3±0.30	1.6±0.69	1.8±0.45	2.0±0.63	2.1±0.70
	MIROC-ESM	0.87±0.045	0.89±0.11	1.1±0.12	1.0±0.16	1.3±0.61	1.9±1.1	2.0±1.3
	MPI-ESM-P	0.89±0.067	0.81±0.093	0.92±0.19	0.69±0.17	0.59±0.18	0.78±0.19	0.89±0.36
Arctic	CCSM4	1.8±0.11	1.7±0.33	1.3±0.24	1.8±0.41	1.2±0.46	0.84±0.29	0.94±0.39
	CSIRO-Mk3L-1-2	0.53±0.025	0.55±0.078	0.67±0.050	0.56±0.11	0.56±0.22	0.56±0.17	0.53±0.18
	HadCM3	0.95±0.30	0.85±0.29	0.90±0.38	0.88±0.37	0.99±0.53	0.92±0.39	0.95±0.29
	IPSL-CM5A-LR	0.91±0.10	0.89±0.15	1.2±0.14	0.74±0.13	0.95±0.22	0.92±0.43	1.2±0.45
	MIROC-ESM	1.1±0.06	0.72±0.02	1.3±0.15	0.71±0.17	1.4±0.17	1.6±0.55	1.9±0.40
	MPI-ESM-P	1.1±0.06	1.2±0.19	1.1±0.13	1.1±0.20	1.4±0.42	1.5±0.56	1.2±0.49
Asia	CCSM4	0.86±0.1	0.96±0.1	0.65±0.1	0.72±0.2	0.75±0.2	0.70±0.3	0.54±0.3
	CSIRO-Mk3L-1-2	0.94±0.08	1.1±0.20	0.81±0.22	0.89±0.23	0.75±0.23	0.95±0.40	0.72±0.31
	HadCM3	1.4±0.07	1.1±0.19	1.4±0.32	1.1±0.41	1.0±0.41	0.8±0.48	1.2±0.35
	IPSL-CM5A-LR	0.85±0.05	0.78±0.09	1.0±0.26	1.0±0.35	0.95±0.27	0.87±0.36	1.3±0.56
	MIROC-ESM	0.92±0.02	1.1±0.03	0.97±0.06	2.1±0.14	2.1±0.24	1.8±0.34	1.1±0.20
	MPI-ESM-P	0.91±0.05	0.95±0.07	1.0±0.22	1.0±0-.25	1.5±0.71	1.4±0.76	1.2±0.61
Europe	CCSM4	1.4±0.19	1.4±0.15	1.3±0.19	1.5±0.58	1.5±0.36	1.2±0.39	1.3±0.38
	CSIRO-Mk3L-1-2	0.85±0.08	0.92±0.10	1.02±0.27	0.71±0.22	0.66±0.17	0.92±0.41	0.68±0.20
	HadCM3	0.97±0.06	1.0±0.14	1.1±0.47	0.95±0.25	1.0±0.37	1.1±0.49	1.1±0.49
	IPSL-CM5A-LR	0.65±0.04	0.63±0.07	0.52±0.09	0.65±0.13	0.80±0.22	0.59±0.24	0.70±0.33
	MIROC-ESM	0.63±0.03	0.50±0.02	0.58±0.03	0.58±0.09	0.96±0.08	0.92±0.12	0.97±0.15
	MPI-ESM-P	1.2±0.10	1.1±0.07	1.1±0.13	1.1±0.53	0.80±0.31	1.5±0.54	1.4±0.43
North America	CCSM4	1.1±0.12	0.99±0.15	0.83±0.22	0.81±0.29	1.5±0.43	1.8±0.81	1.4±0.77
	CSIRO-Mk3L-1-2	0.84±0.06	0.88±0.20	0.73±0.12	0.92±0.50	0.74±0.36	0.77±0.21	0.80±0.18
	HadCM3	1.1±0.14	1.1±0.23	1.1±0.26	0.77±0.20	0.65±0.14	0.58±0.25	0.66±0.36
	IPSL-CM5A-LR	1.0±0.05	1.3±0.13	1.1±0.16	1.2±0.24	0.82±0.22	0.87±0.28	0.95±0.30
	MIROC-ESM	1.0±0.02	0.78±0.03	0.96±0.10	1.2±0.10	1.6±0.34	1.5±0.36	1.4±0.50
	MPI-ESM-P	0.90±0.06	0.91±0.06	1.4±0.29	1.2±0.32	1.1±0.21	1.4±0.50	1.6±0.82
South America	CCSM4	1.1±0.12	0.99±0.15	0.83±0.22	0.81±0.29	1.5±0.43	1.8±0.81	1.4±0.77
	CSIRO-Mk3L-1-2	0.84±0.06	0.88±0.2	0.73±0.12	0.92±0.5	0.74±0.36	0.77±0.21	0.8±0.18
	HadCM3	1.1±0.14	1.0±0.23	1.0±0.26	0.77±0.2	0.65±0.14	0.58±0.25	0.66±0.36
	IPSL-CM5A-LR	1.0±0.05	1.3±0.13	1.1±0.16	1.15±0.24	0.82±0.22	0.87±0.28	0.95±0.3
	MIROC-ESM	1.0±0.02	0.78±0.03	0.96±0.1	1.2±0.1	1.6±0.34	1.5±0.36	1.4±0.5
	MPI-ESM-P	0.90±0.06	0.91±0.06	1.4±0.29	1.2±0.32	1.1±0.21	1.4±0.5	1.6±0.82

Tab. 7.2: Robustness of Average Spectral Density and Halfwidths for Each Model, Region, and Period Bin at L=500

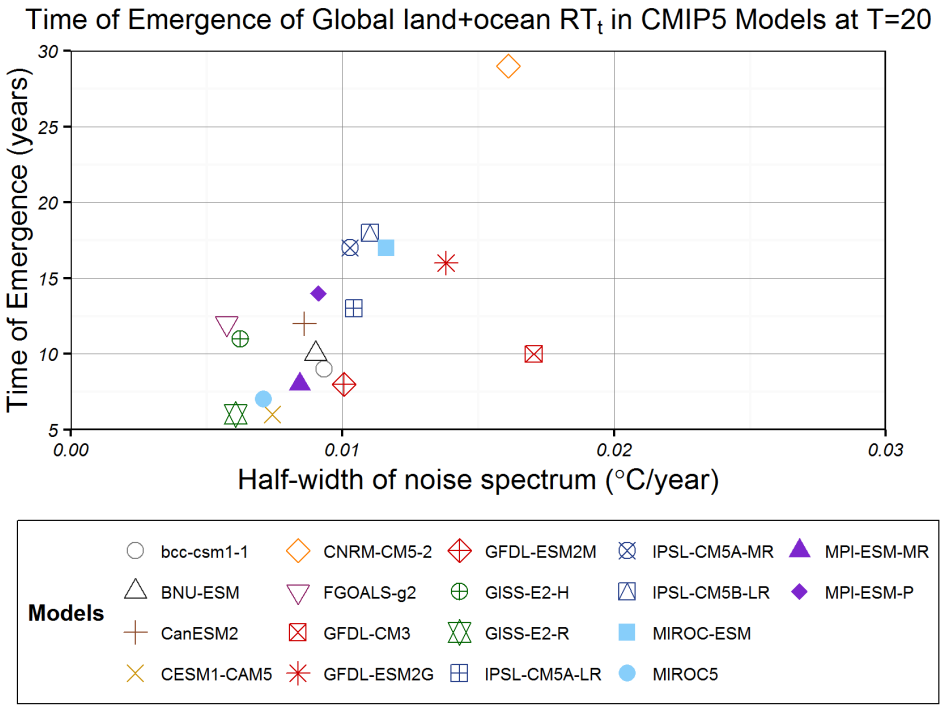


Fig. 7.1: ToE values for each model plotted against the half-width at year 20.

7.2.1 Appendix B: Correlation Analysis

The ToE values for each model are plotted against the half-width of the 5-95% noise range. Figure 7.1 shows the results at $t = 20$. We also looked at $t = 10$ and 15 and found that the model signals clustered with increasing time, inferring that the models are in general agreement in their response. However, there is no obvious trend in the results, suggesting the model noise, or internal variability, is not the primary driver of the ToE.

In Figure 7.2, we further investigated the $\overline{RT_t(20)}/\overline{RT_t(70)}$ compared to the hemispheric values ϕ (top) and NH/SH (bottom). We found the model responses

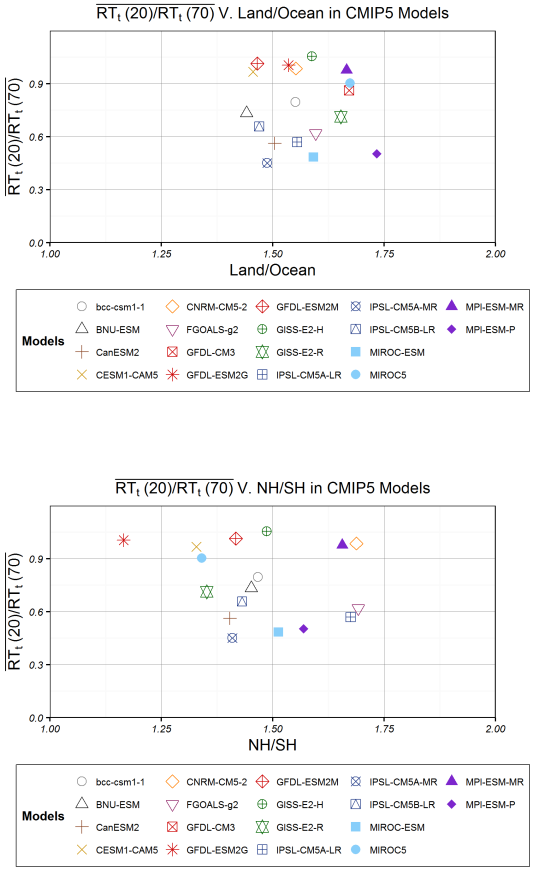


Fig. 7.2: $\overline{RT_t(20)}/\overline{RT_t(70)}$ values for each model plotted against the hemispheric values of ϕ (left) and NH/SH (right).

are not linear, but instead cluster in groups according to the shorter term response compared to the longer term response.

Similarly, we compared the 5-year average of the $RT_t(20)$ and $RT_t(70)$ for each model to the ToE, given by $\overline{RT_t(20)}$ and $\overline{RT_t(70)}$. We see no clear linear relationship, which infers that the signal strength is also not apparently driving the ToE. It is still unclear what is driving the emergence of the signal.

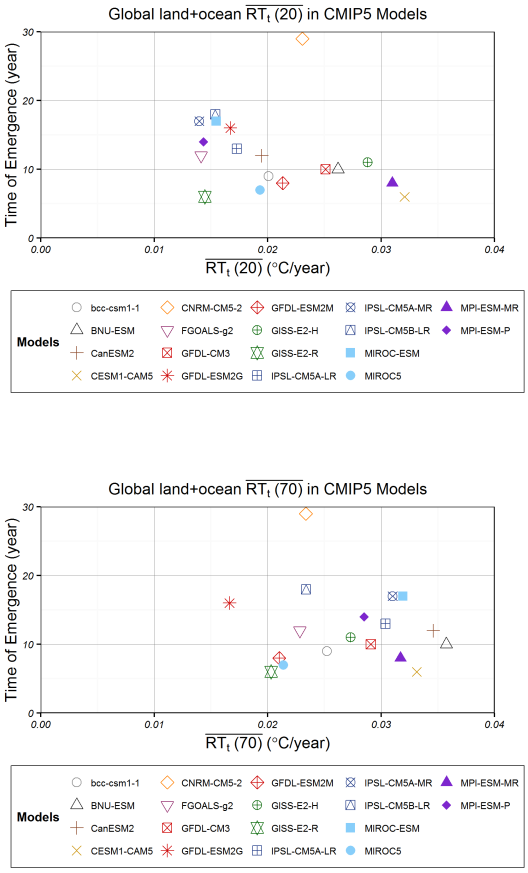


Fig. 7.3: $\overline{RT_t(20)}$ (top) and $\overline{RT_t(70)}$ (bottom) values for each model plotted against the Time of Emergence.

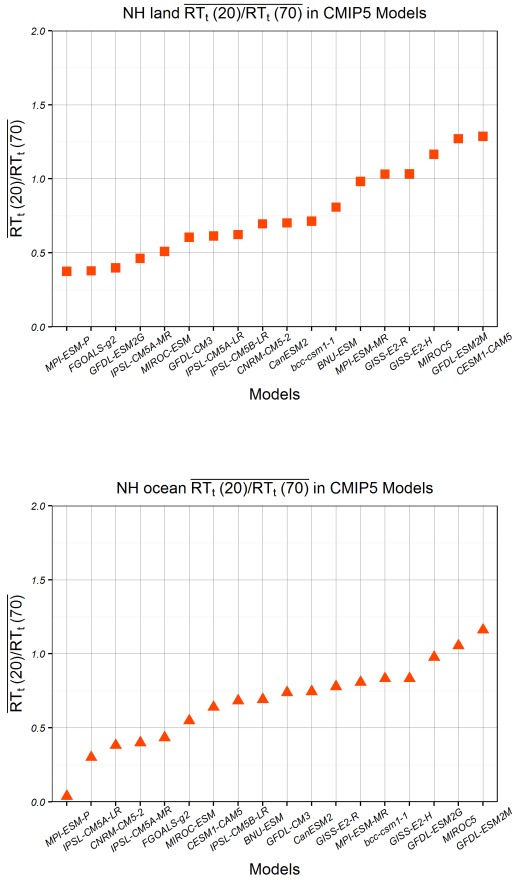


Fig. 7.4: Northern Hemispheric ratio of near-term, $\overline{RT_t(20)}$, to longer-term, $\overline{RT_t(70)}$, response for each CMIP5 model.

7.2.2 Appendix C: Decomposing the Spatial and Temporal Responses

We can decompose Figure 5.10 into the hemispheric levels to understand the response at sub-global levels. Figure 7.4 shows the ratio of $\overline{RT_t(20)}/\overline{RT_t(70)}$ over the land (top) and ocean (bottom). The results show a similar distribution of model response to that observed in Figure 5.10; 7 models are within 20% of $\overline{RT_t(20)}/\overline{RT_t(70)} = 1$. The remaining 11 models have a distinctly lower response at 20 years versus 70 years.

More interesting are the results from the SH. Figure 7.5 shows the ratio of $\overline{RT_t(20)}/\overline{RT_t(70)}$ over the land (top) and ocean (bottom). While, Figure 7.5 shows a similar distribution in the SH land as previously observed for other hemispheric levels, there is a greater distribution in the SH ocean (bottom). Here we see a group of 7 models within 50% of $\overline{RT_t(20)}/\overline{RT_t(70)} = 1.5$. The remaining 10 models have a lower response at 20 years versus 70 years (i.e., ratio less than 1). One model, GFDL-ESM2M, has a higher near-term response compared to longer-term response. This range of response might suggest that the SH ocean plays a important role in the response of CMIP5 models.

Figure 7.6 provides a summary of the above information for a clear comparison of the CMIP5 model response.

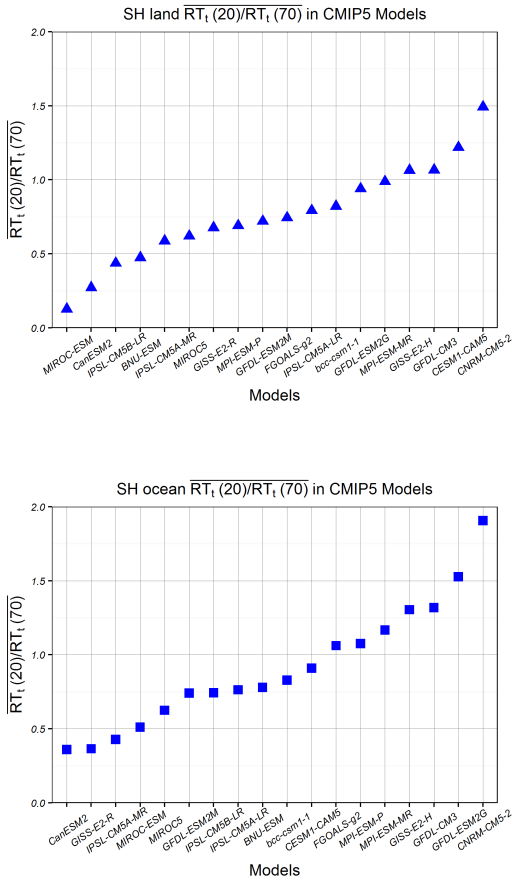


Fig. 7.5: Southern Hemispheric ratio of near-term, $\overline{RT_t(20)}$, to longer-term, $\overline{RT_t(70)}$, response for each CMIP5 model.

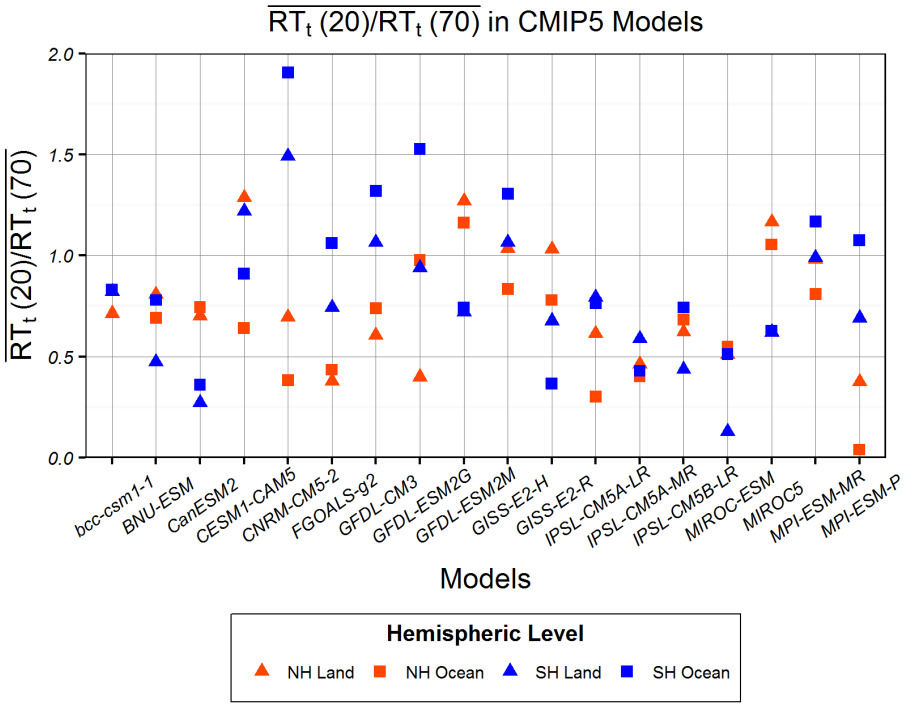


Fig. 7.6: Summary of the ratios of near-term, $\overline{RT_t(20)}$, to longer-term, $\overline{RT_t(70)}$, response for each CMIP5 model for each hemispheric level.

Bibliography

Aamaas, B., Peters, G., and Fuglestvedt, J. (2013). Simple emission metrics for climate impacts. *Earth System Dynamics*, 4(1):145–170.

Abraham, J. P., Baringer, M., Bindoff, N., Boyer, T., Cheng, L., Church, J., Conroy, J., Domingues, C., Fasullo, J., Gilson, J., et al. (2013). A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Reviews of Geophysics*, 51(3):450–483.

Abram, N. J., McGregor, H. V., Tierney, J. E., Evans, M. N., McKay, N. P., Kaufman, D. S., Thirumalai, K., Martrat, B., Goosse, H., Phipps, S. J., et al. (2016). Early onset of industrial-era warming across the oceans and continents. *Nature*, 536(7617):411.

Ahmed, M., Anchukaitis, K. J., Asrat, A., Borgaonkar, H. P., Braida, M., Buckley, B. M., Büntgen, U., Chase, B. M., Christie, D. A., Cook, E. R., et al. (2013). Continental-scale temperature variability during the past two millennia. *Nature geoscience*, 6(5):339–346.

Allen, M., Mustafa, B., et al. (2018a). Global warming of 1.5 c, summary for policy makers. *Intergovernmental Panel on Climate Change. October*, 6.

Allen, M. R., Shine, K. P., Fuglestvedt, J. S., Millar, R. J., Cain, M., Frame, D. J., and Macey, A. H. (2018b). A solution to the misrepresentations of co 2-equivalent emissions of short-lived climate pollutants under ambitious mitigation. *Npj Climate and Atmospheric Science*, 1(1):16.

Allen, M. R. and Stott, P. A. (2003). Estimating signal amplitudes in optimal fingerprinting, part i: Theory. *Climate Dynamics*, 21(5-6):477–491.

Allen, M. R. and Tett, S. F. (1999). Checking for model consistency in optimal fingerprinting. *Climate Dynamics*, 15(6):419–434.

Anenberg, S. C., Schwartz, J., Shindell, D., Amann, M., Faluvegi, G., Klimont, Z., Janssens-Maenhout, G., Pozzoli, L., Van Dingenen, R., Vignati, E., et al. (2012). Global air quality and health co-benefits of mitigating near-term climate change through methane and black carbon emission controls. *Environmental Health Perspectives*, 120(6):831–839.

Annan, J. D. and Hargreaves, J. C. (2017). On the meaning of independence in climate science. *Earth System Dynamics*, 8(1):211–224.

Babu, P. and Stoica, P. (2010). Spectral analysis of nonuniformly sampled data–a review. *Digital Signal Processing*, 20(2):359–378.

Bakker, A. M., Wong, T. E., Ruckert, K. L., and Keller, K. (2017). Sea-level projections representing the deeply uncertain contribution of the west antarctic ice sheet. *Scientific reports*, 7(1):3880.

Berntsen, T. and Fuglestvedt, J. (2008). Global temperature responses to current emissions from the transport sectors. *Proceedings of the National Academy of Sciences*, 105(49):19154–19159.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., et al. (2013). Bounding the role of black carbon in the climate system: A scientific assessment. *Journal of Geophysical Research: Atmospheres*, 118(11):5380–5552.

Bosello, F., Roson, R., and Tol, R. S. (2006). Economy-wide estimates of the implications of climate change: Human health. *Ecological Economics*, 58(3):579–591.

Bothe, O., Jungclaus, J., Zanchettin, D., and Zorita, E. (2013a). Climate of the last millenium: Ensemble consistency of simulations and reconstructions. *Climate of the Past*, 9:1089–1110.

Bothe, O., Jungclaus, J. H., and Zanchettin, D. (2013b). Consistency of the multi-model cmip5/pmip3-past1000 ensemble. *Climate of the Past*, 9:2471–2487.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., et al. (2013). Clouds and aerosols. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 571–657. Cambridge University Press.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., and Zhao, Y. (2012). Evaluation of climate models using palaeoclimatic data. *Nature Climate Change*, 2(6):417.

Broccoli, A. J., Delworth, T. L., and Lau, N.-C. (2001). The effect of changes in observational coverage on the association between surface temperature and the arctic oscillation. *Journal of Climate*, 14(11):2481–2485.

Brown, P. T., Li, W., Cordero, E. C., and Mauget, S. A. (2015). Comparing the model-simulated global warming signal to observations using empirical estimates of unforced noise. *Scientific reports*, 5:9957.

Bryan, K., Komro, F., Manabe, S., and Spelman, M. (1982). Transient climate response to increasing atmospheric carbon dioxide. *Science*, 215(4528):56–58.

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

Byrne, B., Jones, D., Strong, K., Zeng, Z.-C., Deng, F., and Liu, J. (2017). Sensitivity of co2 surface flux constraints to observational coverage. *Journal of Geophysical Research: Atmospheres*, 122(12):6672–6694.

Byrne, M. P. and O'Gorman, P. A. (2013). Link between land-ocean warming contrast and surface relative humidities in simulations with coupled climate models. *Geophysical Research Letters*, 40(19):5223–5227.

Cai, W., Santoso, A., Wang, G., Yeh, S.-W., An, S.-I., Cobb, K. M., Collins, M., Guilyardi, E., Jin, F.-F., Kug, J.-S., et al. (2015). Enso and greenhouse warming. *Nature Climate Change*, 5(9):849–859.

Calel, R. and Stainforth, D. A. (2017). On the physics of three integrated assessment models. *Bulletin of the American Meteorological Society*, 98(6):1199–1216.

Canty, T., Mascioli, N., Smarte, M., and Salawitch, R. (2013). An empirical model of global climate-part 1: A critical evaluation of volcanic cooling. *Atmospheric Chemistry and Physics*, 13(8):3997.

Ceppi, P., Brient, F., Zelinka, M. D., and Hartmann, D. L. (2017). Cloud feedback mechanisms and their representation in global climate models. *Wiley Interdisciplinary Reviews: Climate Change*, 8(4):e465.

Charlson, R. J., Schwartz, S., Hales, J., Cess, R. D., Coakley, J. J., Hansen, J., and Hofmann, D. (1992). Climate forcing by anthropogenic aerosols. *Science*, 255(5043):423–430.

Christensen, T. R., Johansson, T., Åkerman, H. J., Mastepanov, M., Malmer, N., Friborg, T., Crill, P., and Svensson, B. H. (2004). Thawing sub-arctic permafrost: Effects on vegetation and methane emissions. *Geophysical research letters*, 31(4).

Cicerone, R. J. and Oremland, R. S. (1988). Biogeochemical aspects of atmospheric methane. *Global biogeochemical cycles*, 2(4):299–327.

Claussen, M., Mysak, L., Weaver, A., Crucifix, M., Fichefet, T., Loutre, M.-F., Weber, S., Alcamo, J., Alexeev, V., Berger, A., et al. (2002). Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. *Climate dynamics*, 18(7):579–586.

Clune, T. and Rood, R. (2011). Software testing and verification in climate model development. *IEEE software*, 28(6):49–55.

Coats, S., Cook, B. I., Smerdon, J. E., and Seager, R. (2015a). North american pancontinental droughts in model simulations of the last millennium. *Journal of Climate*, 28(5):2025–2043.

Coats, S., Smerdon, J. E., Cook, B. I., and Seager, R. (2015b). Are simulated megadroughts in the north american southwest forced? *Journal of Climate*, 28(1):124–142.

Covey, C., AchutaRao, K. M., Cubasch, U., Jones, P., Lambert, S. J., Mann, M. E., Phillips, T. J., and Taylor, K. E. (2003). An overview of results from the coupled model intercomparison project. *Global and Planetary Change*, 37(1-2):103–133.

Cubasch, U., Meehl, G., Boer, G., Stouffer, R., Dix, M., Noda, A., Senior, C., Raper, S., Yap, K., Abe-Ouchi, A., et al. (2001). Projections of future climate change. In *Climate Change 2001: The scientific basis. Contribution of WG1 to the Third Assessment Report of the IPCC (TAR)*, pages 525–582. Cambridge University Press.

Deser, C., Phillips, A., Bourdette, V., and Teng, H. (2012). Uncertainty in climate change projections: the role of internal variability. *Climate dynamics*, 38(3-4):527–546.

Dommenget, D. and Flöter, J. (2011). Conceptual understanding of climate change with a globally resolved energy balance model. *Climate dynamics*, 37(11-12):2143–2165.

Drost, F., Karoly, D., and Braganza, K. (2012). Communicating global climate change using simple indices: an update. *Climate dynamics*, 39(3-4):989–999.

Easterling, D. R. and Wehner, M. F. (2009). Is the climate warming or cooling? *Geophysical Research Letters*, 36(8).

Eby, M., Weaver, A. J., Alexander, K., Zickfeld, K., Abe-Ouchi, A., Cimatoribus, A., Crespin, E., Drijfhout, S., Edwards, N., Eliseev, A., et al. (2013). Historical and idealized climate model experiments: an intercomparison of earth system models of intermediate complexity. *Climate of the Past*, 9:1111–1140.

England, M. H., McGregor, S., Spence, P., Meehl, G. A., Timmermann, A., Cai, W., Gupta, A. S., McPhaden, M. J., Purich, A., and Santoso, A. (2014). Recent intensification of wind-driven circulation in the pacific and the ongoing warming hiatus. *Nature climate change*, 4(3):222.

Feldstein, S. B. (2000). The timescale, power spectra, and climate noise properties of teleconnection patterns. *Journal of Climate*, 13(24):4430–4440.

Fernández-Donado, J., Raible, C., Ammann, C., Barriopedro, D., Garcia-Bustamante, E., Jungclaus, J., Lorenz, S., Luterbacher, J., Phipps, S., Servonnat, J., et al. (2013). Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium. *Climate of the Past*, 9(1):393–421.

Flanner, M. G. (2013). Arctic climate sensitivity to local black carbon. *Journal of Geophysical Research: Atmospheres*, 118(4):1840–1851.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al. (2014). Evaluation of climate models. In *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 741–866. Cambridge University Press.

Flato, G. M. (2011). Earth system models: an overview. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6):783–800.

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., et al. (2007). Changes in atmospheric constituents and in radiative forcing. chapter 2. In *Climate Change 2007. The Physical Science Basis*.

Forster, P. M. F. and Taylor, K. E. (2006). Climate forcings and climate sensitivities diagnosed from coupled climate model integrations. *Journal of climate*, 19(23):6181–6194.

Franke, J., Frank, D., Raible, C. C., Esper, J., and Brönnimann, S. (2013). Spectral biases in tree-ring climate proxies. *Nature Climate Change*, 3(4):360.

Fricko, O., Havlik, P., Rogelj, J., Klimont, Z., Gusti, M., Johnson, N., Kolp, P., Strubegger, M., Valin, H., Amann, M., et al. (2017). The marker quantification of the shared socioeconomic pathway 2: A middle-of-the-road scenario for the 21st century. *Global Environmental Change*, 42:251–267.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R. (2014). Uncertainties in cmip5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, 27(2):511–526.

Friedman, A. R., Hwang, Y.-T., Chiang, J. C., and Frierson, D. M. (2013). Interhemispheric temperature asymmetry over the twentieth century and in future projections. *Journal of Climate*, 26(15):5419–5433.

Fuglestvedt, J. S., Berntsen, T. K., Godal, O., Sausen, R., Shine, K. P., and Skodvin, T. (2003). Metrics of climate change: Assessing radiative forcing and emission indices. *Climatic Change*, 58(3):267–331.

Fuglestvedt, J. S., Shine, K. P., Berntsen, T., Cook, J., Lee, D., Stenke, A., Skeie, R. B., Velders, G., and Waitz, I. (2010). Transport impacts on atmosphere and climate: Metrics. *Atmospheric Environment*, 44(37):4648–4677.

Furrer, R., Knutti, R., Sain, S., Nychka, D., and Meehl, G. (2007a). Spatial patterns of probabilistic temperature change projections from a multivariate bayesian analysis. *Geophysical Research Letters*, 34(6).

Furrer, R., Sain, S. R., Nychka, D., and Meehl, G. A. (2007b). Multivariate bayesian analysis of atmosphere–ocean general circulation models. *Environmental and ecological statistics*, 14(3):249–266.

Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier van der Gon, H., Facchini, M. C., Fowler, D., Koren, I., Langford, B., Lohmann, U., et al. (2015). Particulate matter, air quality and climate: lessons learned and future needs. *Atmospheric chemistry and physics*, 15(14):8217–8299.

Fyfe, J. C., Gillett, N. P., and Zwiers, F. W. (2013). Overestimated global warming over the past 20 years. *Nature Climate Change*, 3(9):767.

Gasser, T., Peters, G. P., Fuglestvedt, J. S., Collins, W. J., Shindell, D. T., and Ciais, P. (2017). Accounting for the climate-carbon feedback in emission metrics. *Earth System Dynamics*, 8(2):235–253.

Geoffroy, O., Saint-Martin, D., Olivié, D. J., Voldoire, A., Bellon, G., and Tytéca, S. (2013). Transient climate response in a two-layer energy-balance model. part i: Analytical solution and parameter calibration using cmip5 aogcm experiments. *Journal of Climate*, 26(6):1841–1857.

Geoffroy, O., Saint-Martin, D., and Voldoire, A. (2015). Land-sea warming contrast: the role of the horizontal energy transport. *Climate dynamics*, 45(11-12):3493–3511.

Gillett, N. P., Arora, V. K., Matthews, D., and Allen, M. R. (2013). Constraining the ratio of global warming to cumulative co2 emissions using cmip5 simulations. *Journal of Climate*, 26(18):6844–6858.

Giorgi, F. and Bi, X. (2009). Time of emergence (toe) of ghg-forced precipitation change hot-spots. *Geophysical Research Letters*, 36(6).

Giorgi, F. and Mearns, L. O. (2002). Calculation of average, uncertainty range, and reliability of regional climate changes from aogcm simulations via the "reliability ensemble averaging"(rea) method. *Journal of Climate*, 15(10):1141–1158.

Good, P., Gregory, J. M., and Lowe, J. A. (2011). A step-response simple climate model to reconstruct and interpret aogcm projections. *Geophysical Research Letters*, 38(1).

Good, P., Gregory, J. M., Lowe, J. A., and Andrews, T. (2013). Abrupt co2 experiments as tools for predicting and understanding cmip5 representative concentration pathway projections. *Climate dynamics*, 40(3-4):1041–1053.

Goosse, H. (2015). Continental-scale temperature variability in pmip3 simulations and pages 2k regional temperature reconstructions over the past millennium. *Climate of the Past*, 11(12).

Goosse, H., Crespin, E., Dubinkina, S., Loutre, M.-F., Mann, M. E., Renssen, H., Sallaz-Damaz, Y., and Shindell, D. (2012). The role of forcing and internal dynamics in explaining the "medieval climate anomaly". *Climate dynamics*, 39(12):2847–2866.

Graham, N. (1994). Decadal-scale climate variability in the tropical and north pacific during the 1970s and 1980s: Observations and model results. *Climate Dynamics*, 10(3):135–162.

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

Guo, H.-D., Zhang, L., and Zhu, L.-W. (2015). Earth observation big data for climate change research. *Advances in Climate Change Research*, 6(2):108–117.

Gupta, A. S., Jourdain, N. C., Brown, J. N., and Monselesan, D. (2013). Climate drift in the cmip5 models. *Journal of Climate*, 26(21):8597–8615.

Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., Ruedy, R., and Lerner, J. (1984). Climate sensitivity: Analysis of feedback mechanisms. *feedback*, 1:1–3.

Hansen, J., Ruedy, R., Sato, M., and Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, 48(4).

Hansen, J., Sato, M., and Ruedy, R. (1997). Radiative forcing and climate response. *Journal of Geophysical Research: Atmospheres*, 102(D6):6831–6864.

Hargreaves, J. C., Annan, J. D., Ohgaito, R., Paul, A., and Abe-Ouchi, A. (2013). Skill and reliability of climate model ensembles at the last glacial maximum and mid-holocene. *Climate of the Past*, 9(2):811–823.

Harmsen, M. J., van Vuuren, D. P., van den Berg, M., Hof, A. F., Hope, C., Krey, V., Lamarque, J.-F., Marcucci, A., Shindell, D. T., and Schaeffer, M. (2015). How well do integrated assessment models represent non-co 2 radiative forcing? *Climatic change*, 133(4):565–582.

Hartin, C. A., Patel, P., Schwarber, A., Link, R. P., and Bond-Lamberty, B. (2015). A simple object-oriented and open-source model for scientific and policy analyses of the global climate system–hector v1. 0. *Geoscientific Model Development*, 8(4):939–955.

Hartmann, D. L., Tank, A. M. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y. A. R., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., et al. (2013). Observations: atmosphere and surface. In *Climate change 2013 the physical science basis: Working group I contribution to the fifth assessment report of the intergovernmental panel on climate change*, pages 159–254. Cambridge University Press.

Hasselmann, K., Latif, M., Hooss, G., Azar, C., Edenhofer, O., Jaeger, C. C., Johannessen, O. M., Kemfert, C., Welp, M., and Wokaun, A. (2003). The challenge of long-term climate change. *Science*, 302(5652):1923–1925.

Haustein, K., Otto, F. E., Venema, V., Jacobs, P., Cowtan, K., Hausfather, Z., Way, R. G., White, B., Subramanian, A., and Schurer, A. P. (2019). A limited role for unforced internal variability in twentieth-century warming. *Journal of Climate*, 32(16):4893–4917.

Hawkins, E. and Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, 90(8):1095–1108.

Hawkins, E. and Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, 39(1).

Hegerl, G. C., Crowley, T. J., Hyde, W. T., and Frame, D. J. (2006). Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature*, 440(7087):1029.

Hill, C., DeLuca, C., Suarez, M., Da Silva, A., et al. (2004). The architecture of the earth system modeling framework. *Computing in Science & Engineering*, 6(1):18.

Hoffert, M. I., Caldeira, K., Benford, G., Criswell, D. R., Green, C., Herzog, H., Jain, A. K., Kheshgi, H. S., Lackner, K. S., Lewis, J. S., et al. (2002). Advanced technology paths to global climate stability: energy for a greenhouse planet. *science*, 298(5595):981–987.

Hoffert, M. I., Callegari, A. J., and Hsieh, C.-T. (1980). The role of deep sea heat storage in the secular response to climatic forcing. *Journal of Geophysical Research: Oceans*, 85(C11):6667–6679.

Hooß, G., Voss, R., Hasselmann, K., Maier-Reimer, E., and Joos, F. (2001). A nonlinear impulse response model of the coupled carbon cycle-climate system (niccs). *Climate Dynamics*, 18(3-4):189–202.

Hope, A. P., Canty, T. P., Salawitch, R. J., Tribett, W. R., and Bennett, B. F. (2017). Forecasting global warming. In *Paris Climate Agreement: Beacon of Hope*, pages 51–113. Springer, Cham.

Hope, C. (2006). The marginal impact of co2 from page2002: an integrated assessment model incorporating the ipcc's five reasons for concern. *Integrated assessment*, 6(1).

Howarth, R. W. (2015). Methane emissions and climatic warming risk from hydraulic fracturing and shale gas development: implications for policy. *Energy and Emission Control Technologies*, 3:45–54.

Howarth, R. W., Santoro, R., and Ingraffea, A. (2011). Methane and the greenhouse-gas footprint of natural gas from shale formations. *Climatic Change*, 106(4):679.

Hu, J., Emile-Geay, J., and Partin, J. (2017). Correlation-based interpretations of paleoclimate data–where statistics meet past climates. *Earth and Planetary Science Letters*, 459:362–371.

Huang, Y. and Bani Shahabadi, M. (2014). Why logarithmic? a note on the dependence of radiative forcing on gas concentration. *Journal of Geophysical Research: Atmospheres*, 119(24):13–683.

Huybers, P. and Curry, W. (2006). Links between annual, milankovitch and continuum temperature variability. *Nature*, 441(7091):329.

Jones, P. (1998). A users guide for scrip: A spherical coordinate remapping and interpolation package. v 1.4, los alamos national laboratory.

Jones, P., Osborn, T., and Briffa, K. (2001). The evolution of climate over the last millennium. *Science*, 292(5517):662–667.

Joos, F. and Bruno, M. (1996). Pulse response functions are cost-efficient tools to model the link between carbon emissions, atmospheric co2 and global warming. *Physics and Chemistry of the Earth*, 21(5-6):471–476.

Joos, F., Mueller-Fuerstenberger, G., and Stephan, G. (1999). Correcting the carbon cycle representation: How important is it for the economics of climate change? *Environmental Modeling & Assessment*, 4(2-3):133–140.

Joos, F., Roth, R., Fuglestvedt, J. S., Peters, G. P., Enting, I. G., von Bloh, W., Brovkin, V., Burke, E. J., Eby, M., Edwards, N. R., Friedrich, T., Frölicher, T. L., Halloran, P. R., Holden, P. B., Jones, C., Kleinen, T., Mackenzie, F. T., Matsumoto, K., Meinshausen, M., Plattner, G.-K., Reisinger, A., Segschneider, J., Shaffer, G., Steinacher, M., Strassmann, K., Tanaka, K., Timmermann, A., and Weaver, A. J. (2013). Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: a multi-model analysis. *Atmospheric Chemistry and Physics*, 13(5):2793–2825.

Joshi, M. M., Gregory, J. M., Webb, M. J., Sexton, D. M., and Johns, T. C. (2008). Mechanisms for the land/sea warming contrast exhibited by simulations of climate change. *Climate Dynamics*, 30(5):455–465.

Joshi, M. M., Turner, A. G., and Hope, C. (2013). The use of the land-sea warming contrast under climate change to improve impact metrics. *Climatic change*, 117(4):951–960.

Keeling, C. D., Piper, S. C., Bacastow, R. B., Wahlen, M., Whorf, T. P., Heimann, M., and Meijer, H. A. (2005). Atmospheric co 2 and 13 co 2 exchange with the terrestrial biosphere and oceans from 1978 to 2000: Observations and carbon cycle implications. In *A history of atmospheric CO2 and its effects on plants, animals, and ecosystems*, pages 83–113. Springer.

Khodayari, A., Wuebbles, D. J., Olsen, S. C., Fuglestvedt, J. S., Berntsen, T., Lund, M. T., Waitz, I., Wolfe, P., Forster, P. M., Meinshausen, M., et al. (2013). Intercomparison of the capabilities of simplified climate models to project the effects of aviation co2 on climate. *Atmospheric environment*, 75:321–328.

Kim, K.-H., Shim, P.-S., and Shin, S. (2019). An alternative bilinear interpolation method between spherical grids. *Atmosphere*, 10(3):123.

Kleidon, A. and Renner, M. (2017). An explanation for the different climate sensitivities of land and ocean surfaces based on the diurnal cycle. *Earth System Dynamics*, 8(3):849–864.

Knutti, R., Allen, M. R., Friedlingstein, P., Gregory, J. M., Hegerl, G. C., Meehl, G. A., Meinshausen, M., Murphy, J., Plattner, G.-K., Raper, S., et al. (2008). A review of uncertainties in global temperature projections over the twenty-first century. *Journal of Climate*, 21(11):2651–2663.

Knutti, R. and Sedláček, J. (2013). Robustness and uncertainties in the new cmip5 climate model projections. *Nature Climate Change*, 3(4):369.

Kosaka, Y., Xie, S.-P., Lau, N.-C., and Vecchi, G. A. (2013). Origin of seasonal predictability for summer climate over the northwestern pacific. *Proceedings of the National Academy of Sciences*, 110(19):7574–7579.

Kriegler, E. (2005). *Imprecise probability analysis for integrated assessment of climate change*. PhD thesis, Verlag nicht ermittelbar.

Lambert, F. H. and Chiang, J. C. (2007). Control of land-ocean temperature contrast by ocean heat uptake. *Geophysical research letters*, 34(13).

Lelieveld, J., Crutzen, P., and Brühl, C. (1993). Climate effects of atmospheric methane. *Chemosphere*, 26(1-4):739–768.

Libardoni, A. G., Forest, C. E., Sokolov, A. P., and Monier, E. (2019). Underestimating internal variability leads to narrow estimates of climate system properties. *Geophysical Research Letters*.

Mahowald, N., Ward, D. S., Kloster, S., Flanner, M. G., Heald, C. L., Heavens, N. G., Hess, P. G., Lamarque, J.-F., and Chuang, P. Y. (2011). Aerosol impacts on climate and biogeochemistry. *Annual review of environment and resources*, 36:45–74.

Manabe, S., Stouffer, R., Spelman, M., and Bryan, K. (1991). Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric co2. part i. annual mean response. *Journal of Climate*, 4(8):785–818.

Manabe, S. and Strickler, R. F. (1964). Thermal equilibrium of the atmosphere with a convective adjustment. *Journal of the Atmospheric Sciences*, 21(4):361–385.

Mann, M. E., Rutherford, S., Wahl, E., and Ammann, C. (2007). Robustness of proxy-based climate field reconstruction methods. *Journal of Geophysical Research: Atmospheres*, 112(D12).

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J. (1997). Intercomparison makes for a better climate model. *Eos, Transactions American Geophysical Union*, 78(41):445–451.

Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. T., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., et al. (2007). Global climate projections. climate change 2007: the physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change.

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C., Frieler, K., Knutti, R., Frame, D. J., and Allen, M. R. (2009). Greenhouse-gas emission targets for limiting global warming to 2 c. *Nature*, 458(7242):1158.

Meinshausen, M., Raper, S. C., and Wigley, T. M. (2008). Emulating ipcc ar4 atmosphere-ocean and carbon cycle models for projecting global-mean, hemispheric and land/ocean temperatures: Magicc 6.0. *Atmospheric Chemistry and Physics Discussions*, 8:6153–6272.

Meinshausen, M., Raper, S. C., and Wigley, T. M. (2011). Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, magicc6–part 1: Model description and calibration. *Atmospheric Chemistry and Physics*, 11(4):1417–1456.

Millar, R. J., Nicholls, Z. R., Friedlingstein, P., and Allen, M. R. (2017). A modified impulse-response representation of the global near-surface air temperature and atmospheric concentration response to carbon dioxide emissions. *Atmospheric Chemistry and Physics*, 17(11):7213–7228.

Millar, R. J., Otto, A., Forster, P. M., Lowe, J. A., Ingram, W. J., and Allen, M. R. (2015). Model structure in observational constraints on transient climate response. *Climatic Change*, 131(2):199–211.

Monckton, C., Soon, W. W.-H., Legates, D. R., and Briggs, W. M. (2015). Why models run hot: results from an irreducibly simple climate model. *Science Bulletin*, 60(1):122–135.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set. *Journal of Geophysical Research: Atmospheres*, 117(D8).

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., et al.

(2010). The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282):747.

Myhre, G., Shindell, D., and Pongratz, J. (2014). Anthropogenic and natural radiative forcing.

National Center for Atmospheric Research (2014). The climate data guide - global temperature data sets: overview and comparison table.

National Academies of Sciences, E. and Medicine (2016). *Assessment of Approaches to Updating the Social Cost of Carbon: Phase 1 Report on a Near-Term Update*. The National Academies Press.

Neukom, R., Gergis, J., Karoly, D. J., Wanner, H., Curran, M., Elbert, J., González-Rouco, F., Linsley, B. K., Moy, A. D., Mundo, I., et al. (2014). Inter-hemispheric temperature variability over the past millennium. *Nature Climate Change*, 4(5):362.

Ødemark, K., Dalsøren, S., Samset, B., Berntsen, T., Fuglestvedt, J., and Myhre, G. (2012). Short-lived climate forcers from current shipping and petroleum activities in the arctic. *Atmospheric Chemistry and Physics*, 12(4):1979–1993.

Olivié, D., Peters, G., and Saint-Martin, D. (2012). Atmosphere response time scales estimated from aogcm experiments. *Journal of Climate*, 25(22):7956–7972.

Ortiz, R. and Markandya, A. (2010). Literature Review of Integrated Impact Assessment Models of Climate Change with Emphasis on Damage Functions. Working Papers 2009-06, Basque Centre for Climate Change.

Pearce, D. W., Cline, W. R., Achanta, A. N., Fankhauser, S., Pachauri, R. K., Tol, R. S., and Vellinga, P. (1996). The social costs of climate change: greenhouse damage and the benefits of control. *Climate change 1995: Economic and social dimensions of climate change*, pages 179–224.

Perry, S. J., McGregor, S., Gupta, A. S., and England, M. H. (2017). Future changes to el niño–southern oscillation temperature and precipitation teleconnections. *Geophysical Research Letters*, 44(20):10–608.

Peters, G. P., Aamaas, B., Berntsen, T., and Fuglestvedt, J. S. (2011). The integrated global temperature change potential (igtp) and relationships between emission metrics. *Environmental Research Letters*, 6(4):044021.

Phipps, S., Rotstayn, L., Gordon, H., Roberts, J., Hirst, A., and Budd, W. (2011). The csiro mk3l climate system model version 1.0–part 1: Description and evaluation. *Geoscientific Model Development*, 4(2):483–509.

Pierrehumbert, R. (2014). Short-lived climate pollution. *Annual Review of Earth and Planetary Sciences*, 42:341–379.

Pierrehumbert, R. T. (2011). Infrared radiation and planetary temperature. In *AIP Conference Proceedings*, volume 1401, pages 232–244. AIP.

Prerau, M. J., Brown, R. E., Bianchi, M. T., Ellenbogen, J. M., and Purdon, P. L. (2016). Sleep neurophysiological dynamics through the lens of multitaper spectral analysis. *Physiology*, 32(1):60–92.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). Numerical recipes in c++. *The art of scientific computing*, 2:1002.

Provenzale, A. (2014). Climate models. *Rendiconti Lincei*, 25(1):49–58.

Rahim, K. J., Burr, W. S., and Thomson, D. J. (2014). *Applications of Multitaper Spectral Analysis to Nonstationary Data*. PhD thesis, Queen's University. R package version 1.0-14.

Rahmstorf, S., Cazenave, A., Church, J. A., Hansen, J. E., Keeling, R. F., Parker, D. E., and Somerville, R. C. (2007). Recent climate observations compared to projections. *Science*, 316(5825):709–709.

Raper, S. and Cubasch, U. (1996). Emulation of the results from a coupled general circulation model using a simple climate model. *Geophysical Research Letters*, 23(10):1107–1110.

Raper, S., Gregory, J. M., and Osborn, T. (2001). Use of an upwelling-diffusion energy balance climate model to simulate and diagnose a/ogcm results. *Climate Dynamics*, 17(8):601–613.

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., et al. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Global Environmental Change*, 42:153–168.

Riedel, K. S. and Sidorenko, A. (1995). Minimum bias multiple taper spectral estimation. *IEEE Transactions on Signal Processing*, 43(1):188–195.

Rogelj, J., Meinshausen, M., and Knutti, R. (2012). Global warming under old and new scenarios using ipcc climate sensitivity range estimates. *Nature climate change*, 2(4):248.

Rogelj, J., Schaeffer, M., Meinshausen, M., Shindell, D. T., Hare, W., Klimont, Z., Velders, G. J., Amann, M., and Schellnhuber, H. J. (2014). Disentangling the effects of co2 and short-lived climate forcer mitigation. *Proceedings of the National Academy of Sciences*, 111(46):16325–16330.

Rohde, R., Muller, R., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., and Wickham, C. (2013a). A new estimate of the average earth surface land temperature spanning 1753 to 2011. geoinform geostat overv.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., Curry, J., Wickham, C., and Mosher, S. (2013b). Berkeley earth temperature averaging process. *Geoinformatics & Geostatistics: An Overview*, 1(2):1–13.

Rohling, E. J., Sluijs, A., Dijkstra, H. A., Köhler, P., van de Wal, R. S., von der Heydt, A. S., Beerling, D. J., Berger, A., Bijl, P. K., Crucifix, M., et al. (2012). Making sense of palaeoclimate sensitivity. *Nature*, 491(7426):683–691.

Sand, M., Berntsen, T., Von Salzen, K., Flanner, M., Langner, J., and Victor, D. (2016). Response of arctic temperature to changes in emissions of short-lived climate forcers. *Nature Climate Change*, 6(3):286.

Santer, B. D., Mears, C., Doutriaux, C., Caldwell, P., Gleckler, P. J., Wigley, T., Solomon, S., Gillett, N., Ivanova, D., Karl, T. R., et al. (2011). Separating signal and noise in atmospheric temperature changes: The importance of timescale. *Journal of Geophysical Research: Atmospheres*, 116(D22).

Santer, B. D., Mears, C., Wentz, F., Taylor, K., Gleckler, P., Wigley, T., Barnett, T., Boyle, J., Brüggemann, W., Gillett, N., et al. (2007). Identification of human-induced changes in atmospheric moisture content. *Proceedings of the National Academy of Sciences*, 104(39):15248–15253.

Santer, B. D., Painter, J. F., Bonfils, C., Mears, C. A., Solomon, S., Wigley, T. M., Gleckler, P. J., Schmidt, G. A., Doutriaux, C., Gillett, N. P., et al. (2013). Human and natural influences on the changing thermal structure of the atmosphere. *Proceedings of the National Academy of Sciences*, 110(43):17235–17240.

Sarofim, M. C. and Giordano, M. R. (2018). A quantitative approach to evaluating the gwp timescale through implicit discount rates. *Earth System Dynamics*, 9(3):1013–1024.

Sausen, R. and Schumann, U. (2000). Estimates of the climate response to aircraft co 2 and no x emissions scenarios. *Climatic Change*, 44(1-2):27–58.

Schimel, D. S. (1998). Climate change: The carbon equation. *Nature*, 393(6682):208.

Schmale, J. (2016). Arctic warming: Short-term solutions. *Nature Climate Change*, 6(3):234.

Schmidt, G., Jungclaus, J. H., Ammann, C., Bard, E., Braconnot, P., Crowley, T., Delaygue, G., Joos, F., Krivova, N., Muscheler, R., et al. (2012). Climate forcing reconstructions for use in pmip simulations of the last millennium (v1. 1). *Geoscientific Model Development*, (1):185–191.

Schneider, S. H., Thompson, S. L., Taking, A., et al. (2000). A simple climate model used in economic studies of global change. In *In New Directions in the Economics and Integrated Assessment of Global Climate Change (editors,*

DeCanio, SJ, Howarth, RB, Sanstad, AH, Schneider, SH, and SL Thompson. Citeseer.

Schwarber, A., Smith, S., and Hartin, C. (2020a). Characterization of model variability in cmip5. *Geophysical Research Letters - Atmpsheres*.

Schwarber, A., Smith, S., and Hartin, C. (2020b). Characterization of the varying climate response in complex model over time. *Journal of Climate*.

Schwarber, A. K., Smith, S. J., Hartin, C. A., Vega-Westhoff, B. A., and Sriver, R. (2019). Evaluating climate emulation: fundamental impulse testing of simple climate models. *Earth System Dynamics*, 10(4):729–739.

Schwinger, J., Tjiputra, J. F., Heinze, C., Bopp, L., Christian, J. R., Gehlen, M., Ilyina, T., Jones, C. D., Salas-Mélia, D., Segschneider, J., et al. (2014). Nonlinearity of ocean carbon cycle feedbacks in cmip5 earth system models. *Journal of Climate*, 27(11):3869–3888.

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. (2000). The time-dependence of climate sensitivity. *Geophysical Research Letters*, 27(17):2685–2688.

Shindell, D., Kuylenstierna, J. C., Vignati, E., van Dingenen, R., Amann, M., Klimont, Z., Anenberg, S. C., Muller, N., Janssens-Maenhout, G., Raes, F., et al. (2012). Simultaneously mitigating near-term climate change and improving human health and food security. *Science*, 335(6065):183–189.

Shindell, D. T. (2014). Inhomogeneous forcing and transient climate sensitivity. *Nature Climate Change*, 4(4):274.

Shine, K. P. (2000). Radiative forcing of climate change. *Space Science Reviews*, 94(1-2):363–373.

Shine, K. P. (2010). Radiative forcing and climate change. *Encyclopedia of Aerospace Engineering*.

Shine, K. P. and de F Forster, P. M. (1999). The effect of human activity on radiative forcing of climate change: a review of recent developments. *Global and Planetary Change*, 20(4):205–225.

Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., and Regayre, L. A. (2018). Fair v1. 3: A simple emissions-based impulse response and carbon cycle model. *Geoscientific Model Development*, 11(6):2273–2297.

Smith, S. J. and Bond, T. C. (2014). Two hundred fifty years of aerosols and climate: the end of the age of aerosols. *Atmospheric Chemistry and Physics*, 14(2):537–549.

Smith, S. J., Edmonds, J., Hartin, C. A., Mundra, A., and Calvin, K. (2015). Near-term acceleration in the rate of temperature change. *Nature Climate Change*, 5(4):333.

Smith, S. J. and Mizrahi, A. (2013). Near-term climate mitigation by short-lived forcers. *Proceedings of the National Academy of Sciences*, 110(35):14202–14206.

Solomon, S., Qin, D., Manning, M., Averyt, K., Marquis, M., and Tignor, M. M. (2007). *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*, volume 4. Cambridge university press.

Stjern, C. W., Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Ø., Andrews, T., Boucher, O., Faluvegi, G., Iversen, T., Kasoar, M., et al. (2017). Rapid adjustments cause weak surface temperature response to increased black carbon concentrations. *Journal of Geophysical Research: Atmospheres*, 122(21):11–462.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P. M., et al. (2013). Climate change 2013: The physical science basis.

Stott, P. A., Allen, M. R., and Jones, G. S. (2003). Estimating signal amplitudes in optimal fingerprinting. part ii: application to general circulation models. *Climate dynamics*, 21(5-6):493–500.

Stott, P. A., Gillett, N. P., Hegerl, G. C., Karoly, D. J., Stone, D. A., Zhang, X., and Zwiers, F. (2010). Detection and attribution of climate change: a regional perspective. *Wiley Interdisciplinary Reviews: Climate Change*, 1(2):192–211.

Strassmann, K. M. and Joos, F. (2018). The bern simple climate model (bernscm) v1. 0: an extensible and fully documented open-source re-implementation of the bern reduced-form model for global carbon cycle–climate simulations. *Geoscientific model development (GMD)*, 11(5):1887–1908.

Sui, Y., Lang, X., and Jiang, D. (2014). Time of emergence of climate signals over china under the rcp4. 5 scenario. *Climatic change*, 125(2):265–276.

Sutton, R. T., Dong, B., and Gregory, J. M. (2007). Land/sea warming ratio in response to climate change: Ipcc ar4 model results and comparison with observations. *Geophysical Research Letters*, 34(2).

Tanaka, K. and Kriegler, E. (2007). *Aggregated carbon cycle, atmospheric chemistry, and climate model (ACC2): Description of the forward and inverse modes.* Max-Planck-Inst. für Meteorologie.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of cmip5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485–498.

Tebaldi, C. and Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical transactions of the royal society A: mathematical, physical and engineering sciences*, 365(1857):2053–2075.

Tett, S. F., Stott, P. A., Allen, M. R., Ingram, W. J., and Mitchell, J. F. (1999). Causes of twentieth-century temperature change near the earth's surface. *Nature*, 399(6736):569.

Thompson, M. V. and Randerson, J. T. (1999). Impulse response functions of terrestrial carbon cycle models: method and application. *Global Change Biology*, 5(4):371–394.

Thompson, T. M. (2018). Modeling the climate and carbon systems to estimate the social cost of carbon. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e532.

Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., Delgado-Arias, S., Bond-Lamberty, B., Wise, M. A., Clarke, L. E., et al. (2011). Rcp4. 5: a pathway for stabilization of radiative forcing by 2100. *Climatic change*, 109(1-2):77.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096.

Tol, R. S. (2003). Is the uncertainty about climate change too large for expected cost-benefit analysis? *Climatic change*, 56(3):265–289.

Urban, N. M., Holden, P. B., Edwards, N. R., Sriver, R. L., and Keller, K. (2014). Historical and future learning about climate sensitivity. *Geophysical Research Letters*, 41(7):2543–2552.

Valdes, P. (2011). Built for stability. *Nature Geoscience*, 4(7):414.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., et al. (2011a). The representative concentration pathways: an overview. *Climatic change*, 109(1-2):5.

van Vuuren, D. P., Lowe, J., Stehfest, E., Gohar, L., Hof, A. F., Hope, C., Warren, R., Meinshausen, M., and Plattner, G.-K. (2011b). How well do integrated assessment models simulate climate change? *Climatic change*, 104(2):255–285.

van Vuuren, D. P., Meinshausen, M., Plattner, G.-K., Joos, F., Strassmann, K. M., Smith, S. J., Wigley, T. M., Raper, S., Riahi, K., De La Chesnaye, F., et al. (2008). Temperature increase of 21st century mitigation scenarios. *Proceedings of the National Academy of Sciences*, 105(40):15258–15262.

Vose, R. S., Arndt, D., Banzon, V. F., Easterling, D. R., Gleason, B., Huang, B., Kearns, E., Lawrimore, J. H., Menne, M. J., Peterson, T. C., et al. (2012). Noaa's merged land–ocean surface temperature analysis. *Bulletin of the American Meteorological Society*, 93(11):1677–1685.

Walden, A. T., Percival, D. B., and McCoy, E. J. (1998). Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Transactions on Signal Processing*, 46(12):3153–3165.

Walter, K. M., Zimov, S., Chanton, J. P., Verbyla, D., and Chapin, F. S. (2006). Methane bubbling from siberian thaw lakes as a positive feedback to climate warming. *Nature*, 443(7107):71.

Wang, K., Zhang, T., Zhang, X., Clow, G. D., Jafarov, E. E., Overeem, I., Romanovsky, V., Peng, X., and Cao, B. (2017). Continuously amplified warming in the alaskan arctic: Implications for estimating global warming hiatus. *Geophysical Research Letters*, 44(17):9029–9038.

Way, R. G., Oliva, F., and Viau, A. E. (2017). Underestimated warming of northern canada in the berkeley earth temperature product. *International Journal of Climatology*, 37(4):1746–1757.

Weber, S. L. (2010). The utility of earth system models of intermediate complexity (emics). *Wiley Interdisciplinary Reviews: Climate Change*, 1(2):243–252.

Wigley, T., Ammann, C., Santer, B., and Raper, S. C. (2005). Effect of climate sensitivity on the response to volcanic forcing. *Journal of Geophysical Research: Atmospheres*, 110(D9).

Wigley, T. and Raper, S. (2002). Reasons for larger warming projections in the ipcc third assessment report. *Journal of Climate*, 15(20):2945–2952.

Wigley, T. M. and Raper, S. (1987). Thermal expansion of sea water associated with global warming. *Nature*, 330(6144):127.

Yang, Y., Smith, S. J., Wang, H., Mills, C. M., and Rasch, P. J. (2019). Variability, timescales, and nonlinearity in climate responses to black carbon emissions. *Atmospheric Chemistry and Physics*, 19(4):2405–2420.

Yiou, P., Baert, E., and Loutre, M.-F. (1996). Spectral analysis of climate data. *Surveys in Geophysics*, 17(6):619–663.

Zhu, F., Emile-Geay, J., McKay, N. P., Hakim, G. J., Khider, D., Ault, T. R., Steig, E. J., Dee, S., and Kirchner, J. W. (2019). Climate models can correctly simulate the continuum of global-average temperature variability. *Proceedings of the National Academy of Sciences*, 116(18):8728–8733.