

May 2015

Aggregating Evidence in Climate Science: Consilience, Robustness and the Wisdom of Multiple Models

Martin A. Vezér

The University of Western Ontario

Supervisor

Professor Wayne C. Myrvold

The University of Western Ontario

Graduate Program in Philosophy

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Martin A. Vezér 2015

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

Recommended Citation

Vezér, Martin A., "Aggregating Evidence in Climate Science: Consilience, Robustness and the Wisdom of Multiple Models" (2015).
Electronic Thesis and Dissertation Repository. 2837.

<https://ir.lib.uwo.ca/etd/2837>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca.

AGGREGATING EVIDENCE IN CLIMATE SCIENCE: CONSILIENCE,
ROBUSTNESS AND THE WISDOM OF MULTIPLE MODELS

by

Martin A. Vezer

Graduate Program in Philosophy

A thesis submitted
in partial fulfillment of the requirements for
the degree of PhD in Philosophy

The School of Graduate and Postdoctoral Studies

Western University

London, Ontario, Canada

©Martin A. Vezer 2015

Abstract

The goal of this dissertation is to contribute to the epistemology of science by addressing a set of related questions arising from current discussions in the philosophy and science of climate change: (1) Given the imperfection of computer models, how do they provide information about large and complex target systems? (2) What is the relationship between consilient reasoning and robust evidential support in the production of scientific knowledge? (3) Does taking the mean of a set of model outputs provide epistemic advantages over using the output of a single ‘best model’? Synthesizing research in philosophy and science, the thesis analyzes connections among consilient inductions, robustness analysis, and the aggregation of various sources of evidence, including computer simulations, by investigating case studies of climate change that exemplify the strength of consilient reasoning and the security of robust evidential support. It also explains the rationale and epistemic conditions for improving estimates by averaging multiple estimates, comparing a simple case of averaging estimates to practices in multi-model ensemble studies. I argue: (A) the concepts of consilience and robustness account for the strength and security of inferences that rely on imperfect computer modelling methods, (B) consilient reasoning is conducive to attaining robust evidential support, and (C) an analogy can explain why averaging the outputs of multiple models can improve estimates of a target system, given that conditions of model independence, skill and unequal weighting are taken into account.

Keywords: consilience, robustness, epistemology, climate science, models.

The Co-Authorship Statement

Chapter 4 of the dissertation is co-authored with Professor Wayne C. Myrvold, who formulated the mathematical content on pages 80–85.

To Despina

Acknowledgments

I wish to acknowledge and express my appreciation for the help provided by the members of my dissertation supervisory committee: Professor Wayne Myrvold, Dr. Eric Desjardins, Professor Stathis Psillos, Professor Gordon McBean, and Professor Jamie Voogt, in giving their feedback on earlier drafts of the thesis. Since this research focuses on topics at the intersection of epistemology and climate science, I was very fortunate to have worked closely with an interdisciplinary group of scholars from the departments of Philosophy and Geography at Western University. In particular, I am grateful to Professor Myrvold, who agreed to be my supervisor and mentor and in the process taught me a great deal about the philosophy of science in his graduate courses and inspired me during our ongoing collaborations. Dr. Desjardins provided leadership in preparation for two Rotman Institute conferences that helped shape my perspective on the philosophy of climate science: *Integrating Complexity* and *Knowledge and Climate Modelling*. Long before his move to London, Ontario, I had read the scholarly writings of Professor Psillos on the philosophy of science, so I consider myself privileged to have had the opportunity to interact with him since he joined the Department of Philosophy at Western. Professor McBean introduced me to interdisciplinary climate studies in his geography courses on environmental hazards and policy, and subsequent meetings with his climate change research group of natural and social scientists. Professor Voogt also welcomed me into his geography class on urban climatology, and recruited me to join his research network Environmental Prediction in Canadian Cities. These experiences have deepened my appreciation for the societal importance of climate science.

I am also indebted to other friends and colleagues who have brought valuable insights related to this thesis and supported my program of study. Many thanks to Dr. Gillian Barker, Dr. Reuven Brandt, Andy Chater, Dr. Radoslav Dimitrov, Colin Dobell, Dr. Nic Fillion, Professor Jim Fleming, Professor Roman Frigg, Professor Bill Harper, Dr. Dan Hicks, Professor Carl Hofer, Melissa Jacquart, Dr. Rob Moir, Professor Lisa Lloyd, Joshua Luczak, Greg Lusk, Dr. Carolyn McLeod, Professor Margie Morrison, Dr. Wendy Parker, Mr. Joseph Rotman, Professor Brent Sinclair, Professor Chris Smeenk, Karl Slezak, Professor Lenny Smith and participants in the *Vienna International Summer University on Climate Studies* (2013), the Rotman Institute conference *Knowledge and Climate Modelling* (2014) and the Rotman Institute *Epistemology of Scientific Modelling Reading Group* (2013-2014). For their efficient administrative assistance, I thank Susan Bock, Elisa Kilbourne, Meghan Magowan, Rob Read and Carol Suter.

Finally, a few heartfelt words of thanks the brevity of which belies their importance: I am grateful to my dear family for their unconditional support, and my partner Despina for her enduring love.

Contents

Abstract	ii
Co-Authorship Statement	iii
Dedication Page	iv
Acknowledgments	v
List of Figures	vii
List of Tables	viii
Abbreviations	x
1 Introduction	1
1.1 Introduction	1
1.2 Consilience	3
1.3 Robustness	6
1.4 Averaging Estimates	9
1.5 Climate Change	11
1.6 Overview	14
Bibliography	17
2 Computer Models and the Evidence of Anthropogenic Climate Change: Case Studies in Consilient and Robust Epistemology	27
2.1 Introduction	27
2.2 The Consilience of Global Warming	29
2.3 Climate Modelling, Robustness, and Anthropogenic Global Warming	33
2.4 Conclusion	42
Bibliography	43

3	Consilience, Robustness, and the Hockey Stick Hypothesis	47
3.1	Introduction	47
3.2	Consilience	49
3.3	Robustness	55
3.4	Conclusion	59
	Bibliography	60
4	Averaging Estimates: The Wisdom of Multiple Climate Models	67
4.1	Introduction	67
4.2	Climate Model Ensembles	69
4.2.1	Climate model ensemble evaluation	70
4.2.2	Climate change detection and attribution	71
4.2.3	Climate change projections	74
4.2.4	Model Independence	77
4.3	Averaging Multiple Estimates	80
4.3.1	Averaging measurements	80
4.3.2	Independent estimates with equal skill	81
4.3.3	Independent estimates with unequal skill	81
4.3.4	Estimates with dependencies	82
4.3.5	Comparing the simple case with CMIP ensemble method- ology	85
4.4	Conclusion	88
	Bibliography	90
5	Conclusion	96
	Bibliography	101
	Curriculum Vitae	102

List of Figures

1.1	The structure of consilience and robustness	7
2.1	Indicators of changing global climate	32
2.2	D&A Study indicating anthropogenic global warming	36
2.3	24 CIMP3 simulations of GMST over 20th century (2013)	38
3.1	Hockey stick graph, IPCC TAR Fig. 2.20	51
3.2	Hockey stick graph, IPCC AR5 Fig. 5.7	53
4.1	Climate model ensemble evaluations	72
4.2	Climate model ensemble in D&A study	75
4.3	Climate model ensemble projections	76
4.4	Hierarchical model clustering	79
5.1	Dissertation Overview	99

List of Tables

4.1 Comparing Cases of Combining Estimates 87

Abbreviations

- AGW** Anthropogenic Global Warming hypothesis: human activities caused more than half of the observed increase in GMST from 1951–2010
- AR4** Fourth Assessment Report of the IPCC
- AR5** Fifth Assessment Report of the IPCC
- BCE** Before Common Era
- Bern2.5D** An EMIC developed at the University of Bern, Switzerland
- CE** Common Era
- CO₂** Carbon dioxide: a naturally occurring chemical compound composed of two oxygen atoms each covalently double bonded to a single carbon atom. In the earth’s atmosphere, it is a trace gas at a concentration of 0.04 per cent (400 parts per million by volume of air), as of 2014
- CMIP** Coupled Model Intercomparison Project (number indicates phase): Under the World Climate Research Programme, the Working Group on Coupled Modelling established as a standard experimental protocol for studying the output of coupled atmosphere-ocean GCMs. CMIP provides a community-based infrastructure in support of climate model diagnosis, validation, intercomparison, documentation and data access. This framework enables a diverse community of scientists to analyze GCMs in a systematic fashion, a process which serves to facilitate model improvement
- CRU** The Climatic Research Unit (CRU) is an institution at the University of East Anglia, England concerned with the study of natural and anthropogenic climate change
- CRUTEM4** A gridded dataset of global historical near-surface air temperature anomalies over land. This specific version is CRUTEM4.2.0.0-2013-03, available for each month from January 1850 to March 2013, on a 5 degree grid. The dataset, which is updated on a monthly basis, is a collaborative product of the Climatic Research Unit at the University of East Anglia and the Met Office Hadley Centre . Public access to monthly, seasonal and annual data values and time series graphs is available online via Google Earth.

- D&A** Detection and Attribution: Detecting climate change involves a procedure of inferring that some climatic variable has changed in a defined statistical sense. Attributing causal force to a particular variable involves a procedure of determining the components of a target system responsible for the detected change.
- E** A given line of evidence. Numbers refer to lines of evidence in a set (E_1, \dots, E_n).
- EMIC** Earth System Models of Intermediate Complexity
- ENSO** El Niño Southern Oscillation: A basin-wide warming of the tropical Pacific Ocean east of the dateline. On time scales of two to about seven years, this oceanic event is associated with a fluctuation of a globalscale tropical and subtropical surface pressure pattern called the Southern Oscillation.
- ESM** Earth System Model: A coupled atmosphere-ocean GCM that includes a representation of the carbon cycle. Additional components (e.g., atmospheric chemistry, ice sheets, dynamic vegetation, nitrogen cycle) may also be included.
- FAR** First Assessment Report of the IPCC
- GCM** General Circulation Model: A mathematical model of the general circulation of a planetary atmosphere or ocean based on the NavierStokes equations on a rotating sphere with thermodynamic terms for various energy sources (radiation, latent heat). These equations are the basis for complex computer programs commonly used for simulating the atmosphere or ocean of the earth.
- GHG** Greenhouse Gas: Gaseous constituents of the atmosphere that absorb and emit radiation at specific wavelengths within the spectrum of terrestrial radiation emitted by the earth's surface, the atmosphere itself, and by clouds. Water vapour carbon dioxide, nitrous oxide, methane, and ozone are the primary greenhouse gases in the earth's atmosphere.
- GMST** Global Mean Surface Temperature
- GW** Global Warming hypothesis: GMST has increased since the late nineteenth century
- H** A given hypothesis

- HadCRUT4** A dataset of monthly instrumental temperature records formed by combining the sea surface temperature records compiled by the Hadley Centre of the UK Met Office and the land surface air temperature records compiled by the CRU of the University of East Anglia
- HSH** The Hockey Stick Hypothesis: NH mean surface temperatures were higher during the last few decades than during any comparable period over at least the preceding 400 years
- IPCC** Intergovernmental Panel on Climate Change
- LIA** Little Ice Age: An interval during the last millennium (1400 CE and 1900 CE) characterized by a number of extensive expansions of mountain glaciers and moderate retreats in between them, both in the Northern and Southern Hemispheres
- MBH** Michael Mann, Raymond Bradley, and Malcolm Hughes (number indicates particular publication, e.g., MBH98 refers to Mann et al., 1998)
- MM** Stephen McIntyre and Ross McKittrick (number indicates particular publication, e.g., MM05b refers to McIntyre and McKittrick, 2005b)
- MME** Multi-Model Ensemble: A collection of models characterizing a climate prediction or projection. Differences in initial conditions and model formulation result in different evolutions of the modelled system and may give information on uncertainty associated with model error and error in initial conditions in the case of climate forecasts and on uncertainty associated with model error and with internally generated climate variability in the case of climate projections.
- NH** Northern Hemisphere/Hemispheric
- NRC** United States National Research Council
- PAGES 2k** Past Global Changes: climate research project focusing on paleoclimate
- ppm** parts per million
- RCP** Representative Concentration Pathway: Model scenarios that include time series of emissions and concentrations of the full suite of GHGs and aerosols and chemically active gases, as well as land use/land cover.

- SAR** Second Assessment Report of the IPCC
- SB** Willie Soon and Sallie Baliunas (number indicates particular publication, e.g., SB03 refers to Soon and Baliunas, 2003 and Soon et al., 2003a)
- TAR** Third Assessment Report of the IPCC
- UKCP09** United Kingdom Climate Projections, launched for public use in 2009
- WA07** Eugene Wahl and Caspar Ammann
- WCRP** World Climate Research Program

Chapter 1

Introduction

1.1 Introduction

Human activities, primarily the burning of fossil fuels and clearing of forests, have increased the concentration of greenhouse gases (GHGs) in the atmosphere, intensifying the greenhouse effect and causing global warming. Global mean surface temperature (GMST) has risen by about 0.7°C over the period 1906–2005, and the rate of warming during the second half of this period was almost double that for the period as a whole (Solomon et al., 2007, p. 36). Furthermore, 1983–2012 was likely the warmest 30-year period of the last 800 to 1400 years (Masson-Delmotte et al., 2013, p. 386). 2014 was more likely than any other year in the instrumental record (since c. 1880) to be the warmest year on record, and at least 10 of the warmest years on record have occurred within the past two decades (Gillis, 2015). Atmospheric concentrations of the important GHG carbon dioxide (CO₂) have also increased since pre-industrial times (c. 1750) from 280 parts per million (ppm) to over 400 ppm today—a higher level than records show from between 800,000 to 15 million years ago (Tripathi et al., 2009; Thompson, 2015). The burning of coal,

natural gas, and other fossil fuels, releases CO₂ with a unique chemical ‘fingerprint,’ indicating that increasing concentrations are due to human emissions (Ghosh and Brand, 2003). Although global climate change has occurred many times over the course of our planet’s history due to natural causes such as fluctuations in the earth’s orbit, changes in the sun’s intensity, and volcanic activity, none of the known natural forcing mechanisms can account for the current rise in global temperatures.

An overwhelming majority of scientists agree that human activities are causing climate change (Oreskes, 2004; Doran and Zimmerman, 2009; Anderegg et al., 2010; Cook et al., 2013), but public opinion on this issue remains divided.¹ While climate science conveys the evidence of climate change, philosophy of science investigates how scientific methodology produces knowledge, which is the focus of this dissertation. My objective is to contribute to the epistemology of science by addressing a set of related questions arising from current discussions in the philosophy and science of climate change: (1) Given the imperfection of computer models, how do they provide information about large and complex target systems? (2) What is the relationship between consilient reasoning and robust evidential support in the production of scientific knowledge? (3) Does taking the mean of a set of model outputs provide epistemic advantages over using the output of a single ‘best model’? Synthesizing research in philosophy and science, the thesis analyzes connections among consilient inductions, robustness analysis, and the aggregation of various sources of evidence, including computer simulations, by investigating case studies of climate change that exemplify the strength of consilient reasoning and

¹For example, recent polls suggest that only 63% of Canadians believe the scientific evidence is conclusive that climate change is primarily caused by human activity (Environics/Suzuki), and only 45% of Americans attribute the warming of the past four decades to human causes (IPSOS/Reuters, 2015).

the security of robust evidential support. It also explains the rationale and epistemic conditions for improving estimates by averaging of multiple estimates, comparing a simple case of averaging estimates to practices in multi-model ensemble studies. I argue: (A) the concepts of consilience and robustness account for the strength and security of inferences that rely on imperfect computer modelling methods, (B) consilient reasoning is conducive to attaining robust evidential support, and (C) an analogy can explain why averaging the outputs of multiple models can improve estimates of a target system, given that conditions of model independence, skill and unequal weighting are taken into account.

During the middle of the nineteenth century, intellectual and environmental circumstances in the Northern Hemisphere (NH) gave rise to the four ideas that form the core topics of the dissertation: consilience, robustness, the averaging multiple estimates and anthropogenic global warming. The remainder of this chapter discusses these ideas and provides an overview of the chapters that follow.

1.2 Consilience

While some probabilities are objective, assigning a degree of credence to hypotheses about the causes, effects and trajectories of climate change involves incorporating expert judgment with respect to a wide variety of evidence (Mastrandrea et al., 2010). Consilience is a mode of reasoning that involves assigning a high degree of plausibility to a given hypothesis (H) when it is supported by a diverse set of independent lines of evidence, which would be unlikely to converge unless

H were correct.² With major developments in the fields of astronomy, biology, chemistry, and many other disciplines of ‘natural philosophy,’ this was an era of rapidly accelerating scientific discovery and invention (Weber, 2000). Developing an epistemology to account for the strength of the most reliable scientific inferences, William Whewell (1840, 1858, 1860) coined the term ‘consilience’³ (from the Latin, *con*, meaning ‘together,’ and *salire*, meaning ‘to leap’) to describe the “jumping together” of multiple independent lines of evidence in support of H. H receives strong evidential support when it successfully accounts for a set of diverse phenomena because, as Whewell notes:

No accident could give rise to such an extraordinary coincidence. No false supposition could, after being adjusted to one class of phenomena, exactly represent a different class, where the agreement was unforeseen and un contemplated. That rules springing from remote and unconnected quarters should thus leap to the same point, can only arise from *that* being the point where the truth resides (Whewell, 1858, p. 88).

Whewell compares the consilience of inductions to a case of deciphering an inscription, that is, “interpreting an unknown character, in which two different inscriptions, deciphered by different persons, had given the same alphabet. We should, in such a case, believe with great confidence that the alphabet was the true

²I, as others (e.g., Lloyd (2015)), am interested in agreeing inferences from multiple lines of evidence, which are often called ‘variety of evidence inferences.’ These may involve inductions, in Whewell’s sense, but may also refer simply to other sorts of inferences. I will speak of ‘consilience’ when inferences drawn from evidence stem from distinct areas of investigation supporting the same conclusion.

³Whewell coined several other terms in the English language, including the word ‘scientist’ (Snyder, 2012).

one” (Whewell, 1860, pp. 274–5).⁴ Analogously, scientific knowledge is strengthened by the agreement of multiple independent lines of evidence, which converge in support of hypotheses about the natural world. Whewell’s favourite example of consilience is Newton’s theory of universal gravitation, which unifies Kepler’s three laws of motion, accounting for a plethora of phenomena, including the revolutions of planets, the motions of satellites, the patterns of the tide, and the fall of a stone.

Aiming to explain how ‘historical sciences,’ such as paleontology and paleoclimatology, are epistemologically rigorous despite the fact that they employ methods of study outside the confines of traditional laboratory experiments, Stephen Jay Gould (1989, 2002, 2003) applies this idea to inquiries into the deep past. Gould refers to consilience as “the flower principle,” alluding to Charles Darwin’s gathering a wide range of evidence to support his hypotheses in the context of evolutionary biology. Darwin himself expressed the logic of consilience, noting that a “hypothesis may be tested . . . by trying whether it explains several large and independent classes of facts.” Natural selection, for instance, explains the “geological succession of organic beings, their distribution in past and present times, and their mutual affinities and homologies” (Darwin, 1868, vol. 1, p. 657).

⁴Incidentally, this metaphor parallels the following actual case that occurred around the time of Whewell’s writing. In the early 1850s, a handful of scholars claimed to have an approximate understanding of the hitherto undeciphered cuneiform texts of Mesopotamia (Wellard, 1972, p. 79). One of these scholars, William Talbot, attempted a translation of an Assyrian cylinder from the reign of King Tiglath-Pileser I (1116–1078 BCE), using a deciphering system developed by Henry Rawlinson (1846) and Edward Hincks (1846). Talbot sealed his translation in an envelope and sent it to the President of the Royal Asiatic Society, requesting that they test whether the interpretation methodology was reliable by having others translate the cylinder independently. The society agreed, accepting independent interpretations from Rawlinson, Hincks and Jules Oppert. Upon comparing the four translations, the consensus among them indicated the reliability of the method of interpretation (Tiglath Pileser I et al., 1857).

Since the nineteenth century, several philosophers of science have elaborated on the theory of consilience, relating it to more recent discourse in confirmation theory (Hesse, 1968; Laudan, 1971; Thagard, 1978; Cohen and Hesse, 1980; Fisch, 1985; Myrvold, 2003; Snyder, 2012), and applying it to various areas of scientific research, including astronomy (Forster, 1988; Harper, 1989), evolutionary biology (Lloyd, 1983; Gould, 2002), quantum mechanics (Forster, 2010), paleontology (Forber, 2011), historiography (Tucker, 2011), and other natural and social sciences (Wilson, 1998). As Naomi Oreskes (2007) notes, consilience can also account for the strength of inferences made in climate science—an important point upon which this dissertation expands.

1.3 Robustness

The mid-nineteenth century also generated precursors to the idea of robust reasoning. Responding to the epistemology of René Descartes, who held that knowledge could be secured by establishing a sequence of inferences anchored to an indubitable foundational premise, Charles Sanders Peirce (1868, p. 141) advocated for a robust epistemology:

Philosophy ought to imitate the successful sciences in its methods, so far as to proceed only from tangible premises which can be subjected to careful scrutiny, and to trust rather to the multitude and variety of its arguments than to the conclusiveness of any one. Its reasoning should not form a chain which is no stronger than its weakest link, but a cable whose fibers may be so slender, provided they are sufficiently numerous

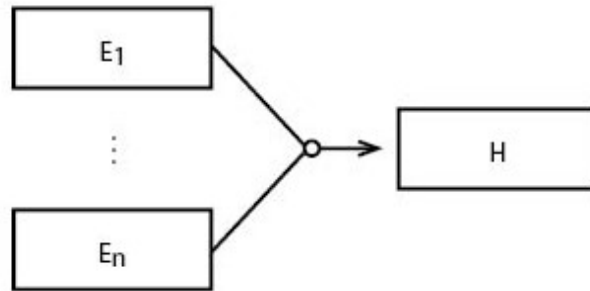


Figure 1.1: The structure of consilience and robustness. See text for details about the similarities and differences between these concepts

and intimately connected.

The concepts of consilience and robustness are similar in that both pertain to the agreement of multiple lines of evidence with respect to the plausibility of H (Figure 1.1). These two modes of reasoning are distinct, however. While consilience suggests the occurrence of E_1, \dots, E_n are unlikely unless H is the case, robustness regards the plausibility of H in the face of some change to a subset of the body of evidence.

Many current discussions of robustness in the philosophy of science literature spring from the work of Richard Levins (1966) who, in examining methodologies in population biology, maintained that, although numerical models (or what he calls “theorems”) are at best incomplete approximations of reality, and the extent to which they represent reality are often ambiguous, one can still use them to secure the plausibility of H , if each model independently points to the same conclusion.

Finding inspiration in the work of Peirce, Levins, and several other proponents of multilateral evidential reasoning, William Wimsatt (1981, 1987) imputes this mode of inference to researchers in many disciplines over the history of human

thought, thus popularizing the notion of robustness analysis.⁵ Philosophers of science have since elaborated upon the idea of robustness, examining how it accounts for the security of inferences in several fields of scientific inquiry, including physics (Cartwright, 1991; Staley, 2004; Woodward, 2006), evolutionary biology (Weisberg, 2006), epidemiology (Nederbragt, 2012), cell biology (Trizio, 2012), mathematics (Krömer, 2012), astrophysics (Allamel-Raffin and Gangloff, 2012) and other disciplines (Soler, 2012).

Case studies about climate change are particularly interesting and informative for epistemological research into the logic of scientific inquiry because climate scientists have systematically scrutinized and reported on the methodologies and findings in this field of research. Accordingly, this dissertation expands on the analysis of Wendy Parker (2006, 2010, 2011) and Elisabeth Lloyd (2009, 2010, 2012, 2015), who investigate the role of robust inferences and variety-of-evidence reasoning in climate modelling studies.

⁵Wimsatt (1981) refers to robust and serial inferences as, respectively, the “Babylonian” and “Greek” methods, but this terminology is problematic. Explaining his reasoning for these historical labels, he cites Feynman (1967, p. 46), who uses the distinction metaphorically to describe two “ways of looking at mathematics.” In Babylonian schools, Feynman asserts, students learn by working on examples until they learn a rule; whereas the Greeks (i.e., Euclid) discovered how theorems can be proven on the basis of established axioms. Although Feynman’s discussion has a very different connotation than Wimsatt’s account of robustness, some Babylonian texts indicate an appreciation for some essential components of scientific inquiry. Brown (2000, Ch. 4) describes Mesopotamian texts from the last centuries of first millennium BCE that document systematic observations and mathematical and non-mathematical methods for predicting and retrodicting celestial phenomena, such as eclipses and planetary positions with a high degree of accuracy. He notes that different methods for predicting a particular phenomena, which were in use at the same time, were considered equally legitimate. As Wimsatt notes, however, the ancient Greeks, such as Aristotle, also “valued having multiple explanations of a phenomenon” (Wimsatt, 1981, p. 61).

1.4 Averaging Estimates

The dissertation will also concentrate on the epistemology of averaging the outputs of multi-model ensemble members, explaining the rationale of this methodology by comparing it to a simpler case of averaging estimates, which suggests crucial conditions that must be met for ensemble averages to improve estimates of a target system, including ensemble constituents' being relevantly skillful, independent, and unequally weighted to reflect differences in model evaluations and dependencies.

Many current lines of research into the causes, effects and trajectories of climate change rely on climate models. One way scientists use these tools to study the climate is by incorporating simulations into ensembles, comparing and combining results to explore and reduce uncertainties about various components of the target system (McGuffie and Henderson-Sellers, 2014). Intergovernmental Panel on Climate Change (IPCC) assess multi-model ensembles (MMEs) in climate model evaluation (Flato et al., 2013), paleoclimate reconstructions (Masson-Delmotte et al., 2013), detection and attribution studies (Bindoff et al., 2013), and climate change projection analysis (Collins et al., 2013). Some studies assume that averaging outputs from multiple simulation is a means of cancelling out random errors, thereby producing estimates of a target system more reliably than any individual simulation.

Statistical methods for averaging estimates developed in the beginning of the nineteenth century, particularly in the context of astronomy (Stigler, 1986). By the turn of the twentieth century, this approach was well-established, as illustrated by the following example. In 1906, Francis Galton attended the annual West of England Fat Stock and Poultry Exhibition in Plymouth, England. This exhibition featured a weight-estimating competition in which competitors, consisting of a mix

of experts (e.g., butchers, farmers) and non-experts, estimated how much an ox would weigh after it had been “slaughtered and dressed.” After the competition, Galton borrowed from the vendor the 800 tickets sold, allowing him to view the recorded estimates. Studying the tickets, he observed that the median estimate of 1207 lb. was only 9 lb. higher than the actual weight of 1,198 lb (Galton, 1907c), while the mean estimate of 1,197 lb. was only 1 lb. lower, notably closer to the actual weight than the best individual estimate, which was 7 lb. higher than the actual weight (Galton, 1907a). “This result,” he maintained, “is more creditable to the trust-worthiness of a democratic judgment than might have been expected” (Galton, 1907b). He called this success of aggregating multiple estimates “*vox populi*” or “the wisdom of crowds.” Since Galton, scholars in various disciplines, ranging from law and economics to political science and cognitive science, have drawn on the wisdom of aggregating multiple estimates to gain insights about their different targets of inquiry (Surowiecki, 2004).

Forming a key research program of the World Climate Research Program (WCRP), the Climate Model Intercomparison Project (CMIP) coordinates scientists from around the world to conduct computer simulation experiments, using different models with overlapping input variables to compile simulation output, which is provided by community members (Taylor et al., 2012). An underlying assumption of certain climate ensemble studies is that, by comparing multiple models, one may identify and reduce systematic errors in estimates because two or more plausible but relevantly independent models are unlikely to produce convergent results unless they accurately represent features of the target in question.

An important consideration for climate model ensemble methodology is the ex-

tent of systematic dependencies among models in a given ensemble. If models in an ensemble are in fact sampled from a ‘truth-centered distribution,’ then different models should on average have near-zero error correlations. In practice, however, analysis of CMIP ensembles demonstrates that such correlations are often strongly positive (Knutti et al., 2010). Although a consensus on how to proceed has yet to be established, some modellers propose adapting and correcting ensemble methods to adjust for systematic biases, determining the effective number of models in ensembles, and weighting models according to relevant degrees of skill and independence (Jun et al., 2008; Jewson and Hawkins, 2009; Pennell and Reichler, 2011). While modellers continue to refine procedures for improving ensemble estimates, the rationale for this approach has yet to be explored in full detail. Accordingly, the dissertation will examine MME methodology by comparing cases in climate modelling with a simpler, well-understood approach of averaging multiple measurements of a single quantity, explaining the ‘wisdom of multiple models.’

1.5 Climate Change

As well as marking the emergence of epistemological discussions about consilience, robustness, and averaging multiple estimates, the mid-nineteenth century marked the end of the Little Ice Age (LIA), a climate interval that started in the early fourteenth century CE, when mountain glaciers of the European Alps and other locations expanded, and mean annual temperatures across the NH declined by 0.6°C (when compared to the average temperature between 1000 and 2000 CE) (Jackson and Rafferty, 2014). From the early 1800s to the first half of 20th century, many

theories emerged about the causes of climate change, suggesting global temperature responses could be caused by changes in the earth's orbit, solar radiation, tides, elevation of land masses, atmospheric and oceanic circulation, continent-ocean distribution, atmospheric volcanic dust concentrations, cosmic dust, sunspots, polar migration and continental drift (Fleming, 1998, 109).

Analyzing the results of field experiments conducted by Horace-Bénédict de Saussure, Edme Marriotte, and others, Joseph Fourier (1837, p. 13), first postulated the greenhouse effect, observing that “the temperature [of the earth] can be augmented by the interposition of the atmosphere, because heat in the state of light finds less resistance in penetrating the air, than in repassing into the air when converted into non-luminous heat.” By the second half of the nineteenth century, John Tyndall (1861, 1863) conducted a series of controlled laboratory experiments, demonstrating how certain gases trap infrared radiation. In reflecting on his experimental findings, Tyndall suggested that changes in atmospheric concentration of some gases, such as water vapour and carbon dioxide (CO₂), could bring about climate change. Considering the implications of the greenhouse effect, Savante Arrhenius (1896) calculated that a doubling of atmospheric CO₂ would increase temperatures by 5 to 6°C.

By the first half of the twentieth century, the Industrial Revolution, which had begun in the mid-sixteenth century, had brought about a transition in manufacturing from hand production to machine methods, increasingly those driven by fossil fuel energy. During this time, scientists commonly thought that contemporary levels of CO₂ concentrations were already absorbing all available long-wave radiation, so increasing concentrations might affect plant growth but would be ineffective at

changing the earth's heat balance (Fleming, 1998).

In the second half of the twentieth century, however, as NH temperatures were rising, global warming was becoming an important public concern, especially with regard to rising sea levels, habitat loss, and changes to agricultural conditions (Fleming, 1998). When scientists investigated further the long-term effects of fossil fuel combustion, they understood that humankind was conducting a geophysical 'experiment' on the global climate (Callendar, 1939; Plass, 1956; Revelle and Suess, 1957). Today, the consensus among climate scientists is that global warming is occurring and its primary causes are anthropogenic.⁶

Among the findings in its latest report, the IPCC states that "For average annual NH temperatures, the period 1983–2012 was *very likely* the warmest 30-year period of the last 800 years" (Masson-Delmotte et al., 2013, p. 386). "Global Mean Surface Temperature has increased since the late 19th century" (Hartmann et al., 2013, p. 161). "It is *extremely likely* that human activities caused more than half of the

⁶National science societies and academies, including the Canadian Meteorological and Oceanographic Society, and the American Association for the Advancement of Science, have issued statements emphasizing their agreement with the conclusion that the evidence for anthropogenic climate change is compelling. The international consensus is articulated by assessment reports of the IPCC, which receives contributions from scientists from over 130 countries around the world. Literature reviews and surveys of scientists corroborate this consensus. Doran and Zimmerman (2009) invited 10,257 earth scientists to participate in a survey on the subject and found that overall, 90% of respondents said they think that, when compared with pre-1800s levels, mean global temperatures have generally risen, and 82% said that they think human activity is a significant contributing factor in changing mean global temperatures. Oreskes (2004) analyzed 928 abstracts of papers from refereed scientific journals between 1993 and 2003, finding that none of the papers disagreed with the consensus position. Anderegg et al. (2010) reviewed publication and citation data for 1,372 climate researchers and drew the following two conclusions: "(i) 97–98% of the climate researchers most actively publishing in the field support the tenets of ACC (Anthropogenic Climate Change) outlined by the Intergovernmental Panel on Climate Change, and (ii) the relative climate expertise and scientific prominence of the researchers unconvinced of ACC are substantially below that of the convinced researchers." Similarly, a review of 11,944 abstracts of scientific papers matching "global warming" or "global climate change," found 4,014 discussing the cause of recent global warming, 97.1% of which endorsed the consensus position (Cook et al., 2013).

observed increase in global mean surface temperature from 1951–2010” (Bindoff et al., 2013, p. 869). Exemplifying the relationship between consilience and robustness, this dissertation will focus on these and related findings, examining the ways in which this epistemology accounts for the strength and security of climate change studies that draw on multiple lines of evidence.

1.6 Overview

Chapter 2. Computer Models and the Evidence of Anthropogenic Climate Change: Case Studies in Consilient and Robust Epistemology

The global climate system is large and complex, with many causal factors interacting. To study climate change, scientists employ computer models, which approximate the target system with various levels of skill. Given the imperfection of climate models, how do scientists use simulations to generate knowledge about the causes of observed climate change? I address this question by developing an epistemology of evidential reasoning based on the ideas of consilient inductions and robustness analysis. Although philosophers have discussed these ideas with respect to the reliability of scientific knowledge in other contexts, the ways in which these concepts are distinct but related to each other, and their relation to climate modelling studies, have only begun to be examined (Lloyd, 2015). Applying this epistemology to case studies of climate change, I argue that, despite imperfections in climate models, consilience and robustness can account for the strength and security of climatologi-

cal inferences that employ computer simulations, including those used to infer that global warming is occurring and its primary causes are anthropogenic.

Chapter 3. Consilience, Robustness, and the Hockey Stick Hypothesis

Some theories of evidential reasoning that suggest the concepts of consilience and robustness can account for the epistemic strength and security of scientific inferences. The way in which these concepts relate to each other, however, has yet to be explored. Developing a conceptual framework to analyze connections between consilient reasoning and robust evidential support, this chapter investigates case studies in climate science supporting the claim that NH mean surface temperatures were higher during the last few decades than during any comparable period over at least the preceding 400 years—a proposition that, for reasons discussed below, is called the ‘hockey stick hypothesis’ (HSH). My thesis is that a consilience of inductions is conducive to its having robust evidential support. Examining studies of climate proxy indicators, which confer some degree of consilience to HSH, this chapter elaborates on the connection between consilience and robustness with respect to debates about global warming. This case also highlights the importance of epistemology as regards the ethics and politics of evidence-based decision-making.

Chapter 4. Averaging Estimates: The Wisdom of Multiple Climate Models

Many current climate studies average the outputs of multiple models, taking the mean to be the ‘best estimate’ of climatic quantities. Does this approach provide epistemic advantages over one that instead dedicates resources to developing a single ‘best model’? If so, what conditions must be met for the combined estimate of a quantity to be more skillful than an estimate from a single model? While modellers continue to refine procedures for improving ensemble estimates, the philosophical rationale for this approach has yet to be explained. Accordingly, this chapter examines ensemble methodology by comparing cases in climate modelling with a simpler, well-understood approach of averaging multiple measurements of a single quantity. It argues that climate change studies incorporating the results of a plurality of models can offer important epistemic advantages over employing only one model, if conditions of model skill, independence and unequal weighting are taken into account.

Chapter 5. Conclusion

The dissertation concludes with an overview of the main findings of the thesis, summarizes the arguments of the preceding chapters, identifies connections among them, and comments on the implications of this research for future work in the epistemology of science.

Bibliography

- Allamel-Raffin, C. and Gangloff, J. L. (2012). Scientific images and robustness. *Characterizing the Robustness of Science*, pages 169–188.
- Anderegg, W. R. L., Prall, J. W., Harold, J., and Schneider, S. H. (2010). Expert credibility in climate change. *Proceedings of the National Academy of Sciences*.
- Arrhenius, S. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science (Fifth Series)*, 41:237–276.
- Bindoff, N., Stott, P., AchutaRao, K., Allen, M., Gillett, N., Gutzler, D. and Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I., Overland, J., Perlwitz, J., Sebbari, R., and Zhang, X. (2013). *Detection and attribution of climate change: From global to regional. In: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Brown, D. (2000). *Mesopotamian planetary astronomy-astrology*, volume 18. Styx Groningen.
- Callendar, G. S. (1939). The composition of the atmosphere through the ages. *Meteorological Magazine*, 74:33–39.
- Cartwright, N. (1991). Replicability, Reproducibility, and Robustness: Comments on Harry Collins. *History of Political Economy*, 23(1):143–155.
- Cohen, L. J. and Hesse, M. B. (1980). *Applications of inductive logic: proceedings of a conference at the Queen's College, Oxford, August 21-24, 1978*. Clarendon Press ; Oxford University Press, Oxford ; New York.

- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, M. G., Shongwe, C. T., Weaver, A., and Wehner, M. (2013). *Long-term climate change: Projections, commitments and irreversibility. In: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs, P., and Skuce, A. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental Research Letters*, 8(2):024024.
- Darwin, C. (1868). *The variation of animals and plants under domestication*, volume 2. John Murray.
- Doran, P. T. and Zimmerman, M. K. (2009). Examining the scientific consensus on climate change. *Eos, Transactions American Geophysical Union*, 90(3):22.
- EnviroNics/Suzuki. Focus Canada 2014 Canadian public opinion about climate change. Technical report, The EnviroNics Institute, and the David Suzuki Foundation.
- Feynman, R. P. (1967). *The character of physical law*, volume 66. MIT press.
- Fisch, M. (1985). Whewell's consilience of inductions—An evaluation. *Philosophy of Science*, pages 239–255.
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M. (2013). *Evaluation of climate models. In: The Physical Science Basis. Contribution of Working Group*

- I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Fleming, J. R. (1998). *Historical Perspectives on Climate Change*. Oxford University Press.
- Forber, Patrick, E. G. (2011). Historical reconstruction: Gaining epistemic access to the deep past. *Philosophy & Theory in Biology*.
- Forster, M. R. (1988). Unification, explanation, and the composition of causes in newtonian mechanics. *Studies in History and Philosophy of Science Part A*, 19(1):55–101.
- Forster, M. R. (2010). Miraculous consilience of quantum mechanics. *The Place of Probability in Science*, pages 201–228.
- Fourier, J.-B. J. (1837). Ebeneser burgess’s translation of fourier (1824). *The American Journal of Science*, 32.
- Galton, F. (1907a). The ballot-box. *Nature*, 75:509–510.
- Galton, F. (1907b). One vote, one value. *Nature*, 75:414–414.
- Galton, F. (1907c). Vox populi. *Nature*, 75:450–451.
- Ghosh, P. and Brand, W. A. (2003). Stable isotope ratio mass spectrometry in global climate change research. *International Journal of Mass Spectrometry*, 228(1):1–33.
- Gillis, J. (2015). 2014 Breaks Heat Record, Challenging Global Warming Skeptics. *The New York Times*.
- Gould, S. J. (1989). *Wonderful life: the Burgess Shale and the nature of history*.
- Gould, S. J. (2002). *The structure of evolutionary theory*. Belknap Press.

- Gould, S. J. (2003). *The hedgehog, the fox, and the magister's pox: Mending the gap between science and the humanities*. Random House LLC.
- Harper, W. (1989). Consilience and natural kind reasoning. by *JR Brown and J. Mittelstrass*. Dordrecht: Kluwer.
- Hartmann, D., Tank, A. K., Rusticucci, M., Alexander, L., Bronnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., and Zhai, P. (2013). *Observations: Atmosphere and Surface*. In *Climate Change: The Physical Science Basis : Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge Univ Press.
- Hesse, M. (1968). Consilience of inductions. In Lakatos, I., editor, *Studies in Logic and the Foundations of Mathematics*, volume Volume 51, pages 232–257. Elsevier.
- Hincks, E. (1846). *On the First and Second Kinds of Persepolitan Writing*. M. H. Gill.
- IPSOS/Reuters (2015). Poll: Climate Change.
- Jackson, S. T. and Rafferty, J. (2014). Little ice age (LIA) geochronology. *Encyclopedia Britannica*. <http://www.britannica.com/EBchecked/topic/344106/Little-Ice-Age-LIA>.
- Jewson, S. and Hawkins, E. (2009). CMIP3 ensemble spread, model similarity, and climate prediction uncertainty. *arXiv preprint arXiv:0909.1890*.
- Jun, M., Knutti, R., and Nychka, D. W. (2008). Spatial analysis to quantify numerical model bias and dependence. *Journal of the American Statistical Association*,

- 103(483):934–947.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate*, 23(10):2739–2758.
- Krömer, R. (2012). Are we still babylonians? the structure of the foundations of mathematics from a wimsattian perspective. In *Characterizing the Robustness of Science*, pages 189–206. Springer.
- Laudan, L. (1971). William whewell on the consilience of inductions. *The Monist*, 55(3):368–391.
- Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54(4):421–431.
- Lloyd, E. (2009). I-Varieties of support and confirmation of climate models. In *Aristotelian Society Supplementary Volume*, volume 83, pages 213–232.
- Lloyd, E. (2010). Confirmation and robustness of climate models. *Philosophy of Science*, 77(5):971–984.
- Lloyd, E. A. (1983). The nature of darwin’s support for the theory of natural selection. *Philosophy of Science*, pages 112–129.
- Lloyd, E. A. (2012). The role of complex empiricism in the debates about satellite data and climate models. *Studies in History and Philosophy of Science Part A*, 43(2):390–401.
- Lloyd, E. A. (2015). Model robustness as a confirmatory virtue: The case of climate science. *Studies in History and Philosophy of Science Part A*, 49:58–68.
- Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Gon-

- zlez Rouco, J., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., and Timmermann, A. (2013). *Information from Paleoclimate Archives. In: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.
- Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., et al. (2010). Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. *Intergovernmental Panel on Climate Change (IPCC).*
- McGuffie, K. and Henderson-Sellers, A. (2014). *The Climate Modelling Primer.* John Wiley & Sons.
- Myrvold, W. C. (2003). A bayesian account of the virtue of unification. *Philosophy of Science*, 70(2):399–423.
- Nederbragt, H. (2012). Multiple derivability and the reliability and stabilization of theories. In Soler, L., Trizio, E., Nickles, T., and Wimsatt, W., editors, *Characterizing the Robustness of Science*, number 292 in Boston Studies in the Philosophy of Science, pages 121–145. Springer Netherlands.
- Oreskes, N. (2004). The scientific consensus on climate change. *Science*, 306(5702):1686–1686.
- Oreskes, N. (2007). *The scientific consensus on climate change: How do we know we're not wrong? In: Climate change: what it means for us, our children, and our grandchildren*, pages 66–99. MIT Press. pp. 66–99.

- Parker, W. (2006). Understanding pluralism in climate modeling. *Foundations of Science*, 11(4):349–368.
- Parker, W. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science*, 78(4):579–600.
- Parker, W. S. (2010). Comparative process tracing and climate change fingerprints. *Philosophy of Science*, 77(5):1083–1095. ArticleType: research-article / Full publication date: December 2010 / Copyright 2010 Philosophy of Science Association.
- Peirce, C. S. (1868). Some consequences of four incapacities. *The Journal of Speculative Philosophy*, 2(3):140–157.
- Pennell, C. and Reichler, T. (2011). On the effective number of climate models. *Journal of Climate*, 24(9):2358–2367.
- Plass, G. (1956). Carbon dioxide and the climate. *American Scientist*, 44(3):302–316.
- Rawlinson, H. C. (1846). The persian cuneiform inscription at behistun, decyphered and translated; with a memoir on persian cuneiform inscriptions in general, and on that of behistun in particular. *Journal of the Royal Asiatic Society*, 10.
- Revelle, R. and Suess, H. E. (1957). Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO₂ during the past decades. *Tellus*, 9(1):18–27.
- Snyder, L. (2012). William whewell (Stanford encyclopedia of philosophy). <http://plato.stanford.edu/entries/whewell/>.
- Soler, L. (2012). Characterizing the robustness of science after the practice turn in

- philosophy of science. *Springer eBooks*.
- Solomon, S., Qin, D., Manning, J., Alley, R., Berntsen, T., Bindoff, N. L., Chen, Z., Chidthaisong, A., Gregory, J., Hegerl, G. C., Heimann, M., Hewitson, B., Hoskins, B. J., Joos, F., Jouzel, J., Kattsov, V., Lohmann, U., Matsuno, T., Molina, M., Nicholls, N., Overpeck, J., Raga, G., Ramaswamy, V., Ren, J., Rusticucci, M., Somerville, R., Stocker, T. F., Whetton, P., Wood, R. A., and Wratt, D. (2007). *Technical Summary*. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, volume 4. Cambridge University Press.
- Staley, K. W. (2004). Robust evidence and secure evidence claims. *Philosophy of Science*, 71(4):467–488.
- Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Harvard University Press.
- Surowiecki, J. (2004). The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations*.
- 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.
- Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy*, pages 76–92.
- Thompson, A. (2015). 2015 Begins with CO2 above 400 PPM Mark.

- Tiglath Pileser I, K., Rawlinson, S. H. C., Talbot, W. H. F., Hincks, E., and Oppert, J. (1857). *Inscription of Tiglath Pileser, 1150 B.C., as translated by Sir H. Rawlinson, F. Talbot, Dr. Hincks, and Dr. Oppert.*
- Tripathi, A. K., Roberts, C. D., and Eagle, R. A. (2009). Coupling of CO₂ and Ice Sheet Stability Over Major Climate Transitions of the Last 20 Million Years. *Science*, 326(5958):1394–1397.
- Trizio, E. (2012). Achieving robustness to confirm controversial hypotheses: a case study in cell biology. In *Characterizing the Robustness of Science*, pages 105–120. Springer.
- Tucker, A. (2011). *A Companion to the Philosophy of History and Historiography*. John Wiley & Sons.
- Tyndall, J. (1861). The bakerian lecture: On the absorption and radiation of heat by gases and vapours, and on the physical connexion of radiation, absorption, and conduction. *Philosophical Transactions of the Royal Society of London*, 151:1–36.
- Tyndall, J. (1863). On radiation through the earth's atmosphere. *The Chemical news and journal of industrial science*, 7:101–103. Sir Henry Holland, Bart., M.D., D.C.L., F.R.S., Vice-President, in the Chair.
- Weber, A. S. (2000). *Nineteenth-Century Science: An Anthology*. Broadview Press.
- Weisberg, M. (2006). Robustness analysis. *Philosophy of Science*, 73(5):730–742.
- Wellard, J. H. (1972). *Babylon*. Saturday Review Press.
- Whewell, W. (1840). *The philosophy of the inductive sciences: founded upon their history*. London: John W. Parker, West Strand.

- Whewell, W. (1858). *Novum Organon Renovatum*. J.W. Parker and Son.
- Whewell, W. (1860). *On the philosophy of discovery, chapters historical and critical*;. London, J.W. Parker and son.
- Wilson, E. O. (1998). *Consilience: the unity of knowledge*. Knopf : Distributed by Random House, New York, 1st ed edition.
- Wimsatt, W. C. (1981). Robustness, reliability, and overdetermination (2012)). In *Characterizing the Robustness of Science*, pages 61–87. Springer.
- Wimsatt, W. C. (1987). False models as means to truer theories. *Neutral models in biology*, pages 23–55.
- Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology*, 13(2):219–240.

Chapter 2

Computer Models and the Evidence of Anthropogenic Climate Change: Case Studies in Consilient and Robust Epistemology

2.1 Introduction

The global climate system is large and complex, with many casual factors interacting. To study climate change, scientists employ computer models, which are very imperfect representations of the target system. The most detailed, high-resolution models of the global climate omit representations of major features of the planet that affect the climate (e.g., mountain ranges) and contain parameterizations that simplify complex climatic processes (e.g., cloud formation). Knowing whether a

given climate model provides insight into questions about a target system at various scales and about its responses to different perturbations can, therefore, be difficult to determine.

Among the findings of the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) is the conclusion that “[i]t is *extremely likely* that human activities caused more than half of the observed increase in global mean surface temperature [GMST] from 1951–2010” (Meehl et al., 2007, p. 869). An important source of evidence for this conclusion are the results of computer model simulations. political debates about climate change frequently characterize scientific methodology as a mode of reasoning that is only as strong as its weakest link.¹ Given that each of the modelled worlds used to study the climate is very different from the earth, how do scientists use computer simulations to generate knowledge about the causes of observed climate change?

I address this question by employing the ideas of consilient reasoning and robustness analysis, focusing on the epistemic advantages of drawing on a variety of evidence, and expanding upon related philosophical inquiries into this field of study (Weisberg, 2006; Lloyd, 2009, 2010; Parker, 2010; Edwards, 2010). Although philosophers have discussed the importance of multiple sources of evidence, the ways in which the concepts of consilience and robustness are distinct but related to each other, and relevant to climate modelling methodology, have yet to be explained. Applying this epistemology to case studies documented by the IPCC, I argue that the notions of consilience and robustness account for the strength and

¹For example, some have tried to discredit climate science by focusing on a few erroneous claims of IPCC, such as the suggestions that the Himalayan glaciers would melt by 2035, or that the Amazon rainforest may be at risk from droughts (Lewis, 2010).

security of climatological inferences that make use of imperfect computer models.

The approach of the paper is as follows. Section 2 illustrates the logic of consilience by describing a set of evidential pathways that converge in support of the global warming hypothesis. This example presents a clear case of consilient reasoning without having to address the question of the role of computer model simulations. Section 3 expands upon this case by explaining how multiple imperfect models contribute to the security of scientific knowledge because a variety of independent approximations of a target system provide alternative evidential pathways to support particular hypotheses. This point is exemplified by model agreement in studies of anthropogenic climate change. Section 4 concludes the paper by summarizing the main points of the argument.

2.2 The Consilience of Global Warming

Kent Staley (2004) notes that attaining multiple lines of evidence can increase the plausibility of a hypothesis (H) by enhancing its epistemic strength and security. While he draws a distinction between H's strength (i.e., the degree to which evidence indicates H) and its security (i.e., the insensitivity of H to changes in some evidential pathway), he sets aside the question of how one can increase the strength of an inference (p. 468). The idea of consilience, however, can account for the way in which multiple lines of evidence strengthen an inference. A consilience of inductions occurs when multiple independent evidential pathways indicate H such that it would be unlikely that these lines of evidence would occur if H were incorrect.

Climate science contains many examples of consilient reasoning, the case of

global warming providing a vivid illustration. In evaluating the Global Warming hypothesis (GW)—that is, the proposition that “Global Mean Surface Temperature has increased since the late 19th century” (Hartmann et al., 2013, p. 161)—the IPCC highlights several agreeing lines of evidence, colligating many facts which comprise a consilience of inductions. The evidence converging on GW results from multiple observations of different interconnected components of the climate system, the collection of which would be unlikely to occur if GW were incorrect. While land-surface weather stations provide the most direct evidential pathway supporting GW, a consilience emerges from other climate records, including measured changes in atmospheric and oceanic temperatures at various heights and depths; in glacier mass, snow coverage, and sea ice extent; in sea level; and in atmospheric water vapour content. Although these various measurements indicate different kinds of climate change, these findings are consilient with GW.

Figure 2.1 contains 10 graphs depicting this consilience. Since the atmosphere and hydrosphere are interconnected fluid bodies, a warming at the earth’s surface produces detectable effects at different levels of the atmosphere and ocean. Some of the energy absorbed by the climate system is stored in the oceans, and this energy uptake is detectable in global ocean heat content records going back to the 1950s. Another line of support is the change in the amount of water vapour in the atmosphere, i.e., its specific humidity, measurements of which indicate a positive change to this variable, both over the land and the oceans. Observed sea-level rise is another line of support; warming oceans result in water expansion, leading to rising sea levels, which are further heightened by additional water input from melting glaciers and ice sheets and changes to the storage and usage of water on land. The

cryosphere (i.e., the frozen parts of the planet) is also affected by changing temperatures. Snow cover, particularly during the spring, is sensitive to temperature changes. Since the 1950s, Northern Hemisphere spring snow cover has declined. Similarly, Arctic sea-ice losses are detectable in satellite records, particularly at the end of the annual melt in September, which is the time of its minimum extent. For at least the last 20 years, the amount of ice contained in glaciers globally has declined (Hartmann et al., 2013).

Since observations of these diverse phenomena are both consistent with GW and inconsistent with the claim that GW is incorrect, this body of evidence comprises a case of consilience. As Whewell would say, “[n]o accident could give rise to such an extraordinary coincidence” (Whewell, 1858, p. 88). With the detection of GW thus well established, the next question to consider is: What has been causing this warming?

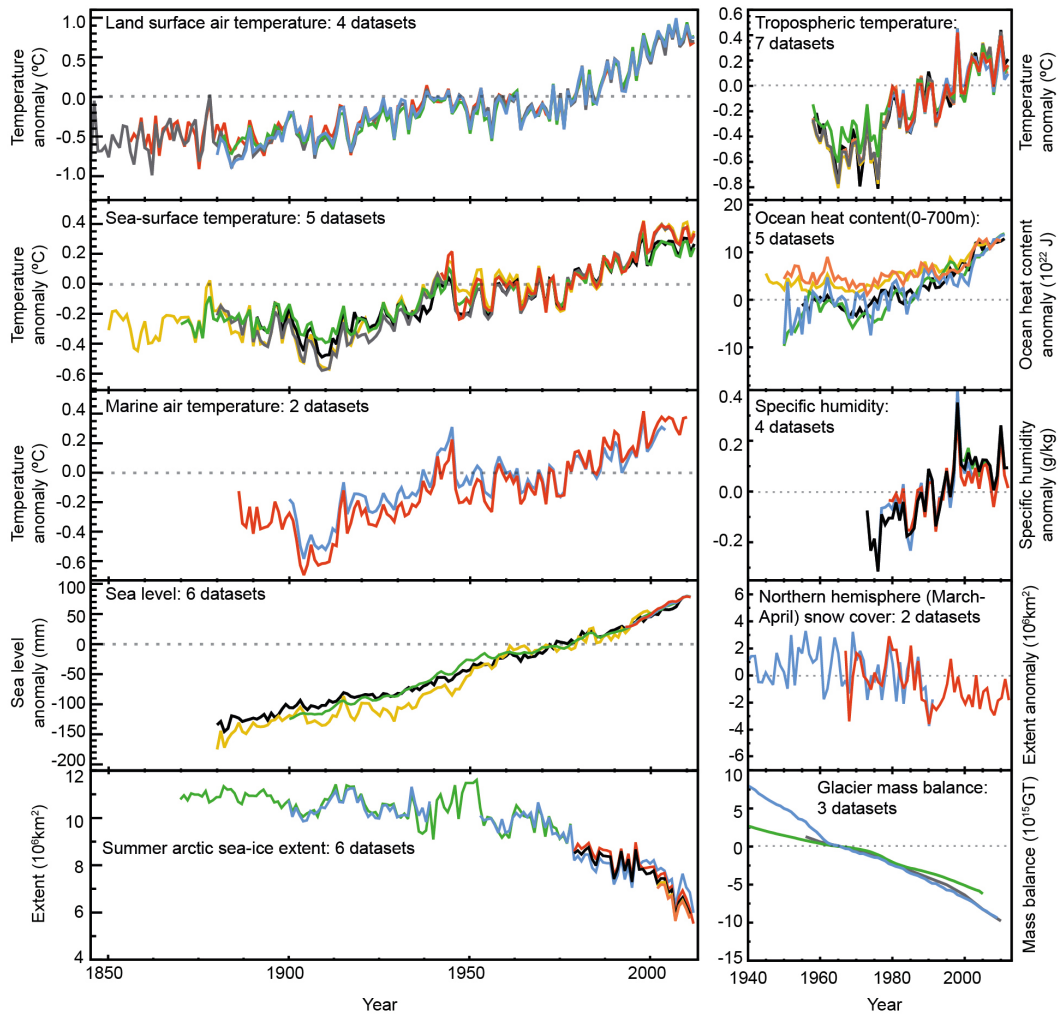


Figure 2.1: Multiple independent indicators of a changing global climate. Each line represents an independently derived estimate of change in the climate element. In each panel all data sets have been normalized to a common period of record (Hartmann et al., 2013, p. 199) ©Cambridge University Press.

2.3 Climate Modelling, Robustness, and Anthropogenic Global Warming

While consilience increases the strength of H by drawing on multiple lines of evidence, this mode of reasoning is distinct from the idea of evidential robustness, which also depends on multiple lines of evidence. Robustness analysis secures the plausibility of H by ensuring that it is insensitive to inaccuracies of a particular evidential pathway. The following case exemplifies how converging evidential pathways provide H with epistemic security, despite imperfections in particular lines of supporting evidence.

One of the earliest formulations of the thesis that humans are conducting a large scale ‘experiment’ on the global climate comes from the work of Guy Stewart Callendar who, in 1938, read a paper to the Royal Meteorological Society, arguing that CO₂ from fossil fuel consumption caused a measurable increase in the earth’s temperature. “The course of world temperatures during the next twenty years should,” he explained, “afford valuable evidence as to the accuracy of the calculated [warming] effect of atmospheric carbon dioxide” (Callendar, 1939, p. 236). Over a decade later, Gilbert Plass echoed this idea, writing that, “[i]f at the end of this century, measurements show that the carbon dioxide content of the atmosphere has risen appreciably and at the same time the temperature has continued to rise throughout the world, it will be firmly established that carbon dioxide is an important factor in causing climatic change” (Plass, 1956, p. 387). Perhaps the most well-known statement of this thesis is that of Roger Revelle and Hans Suess:

[H]uman beings are now carrying out a large scale geophysical exper-

iment of a kind that could not have happened in the past nor be reproduced in the future. Within a few centuries we are returning to the atmosphere and oceans the concentrated organic carbon stored in sedimentary rocks over hundreds of millions of years. This experiment, if adequately documented, may yield a far-reaching insight into the processes determining weather and climate (Revelle and Suess, 1957, pp. 19–20).²

The logic of climate change detection and attribution (D&A) studies involves a mode of induction that is more complex than simply correlating increasing anthropogenic greenhouse gas (GHG) emissions with increasing GMST. While detecting climate change involves a procedure of inferring that some climatic variable has changed in a defined statistical sense, attributing causal force to a particular variable involves a procedure of determining the components of a target system responsible for the detected change. D&A analyses investigate aspects of the climate over different scales of space (from a single grid point to the whole globe) and time (from minutes to millennia), using various methods to study the causes of climate change. Some focus on human activities that influence climatic variables such as surface air temperature, tropospheric temperature, depletion of stratospheric ozone, daily temperature extremes, upper ocean temperatures, glacier melting, ocean salinity, ocean acidification and oxygen depletion, and precipitation patterns.³ Attribution studies into the question of whether human activities are responsible for climate change employ computer modelling methods. An important modelling strategy is to examine

²For related historical accounts, see Fleming (1998) and Weart (2011).

³For an overview of D&A studies investigating these and other variables, see Bindoff et al. (2013, pp. 932–939).

whether a detected change is consistent with simulations (made by skillful⁴ models) that predict a given response to anthropogenic forcing while being inconsistent with alternative (counterfactual) simulations that exclude this forcing. Accordingly, this approach has been applied to investigate the causes of GW (Figure 2.2).

Although computer models and the simulations they produce are critical in scientific studies about climate change, even the most detailed, high-resolution models of the global climate omit important representations of major features of the planet that affect the climate and contain parameterizations that simplify complex climatic processes (McGuffie and Henderson-Sellers, 2014). While certain forms of scientific representations have unambiguous relations to reality, the extent of the representational relationship between numerical models and reality can be ambiguous. Richard Levins (1966) explains this point by comparing numerical models (in population biology) with traditional geographic maps. Map legends stipulate clear relations between the object system (i.e., the markings on a page) and the target system (i.e., the geographical space it depicts). Continuity of lines on a map correspond to contiguity of physical features of reality; relative distances on a map indicate relative distances in space; colours and symbols indicate topographic features, and so on. Unlike a geographical map, the representational limits of mathematical models are often less clear. Whether a given climate model provides insight into questions about a target system at various scales or about its responses to various perturbations can be difficult to determine.

Roman Frigg et al. (2013, p. 893), for example, maintain that systematic errors in models of the Climate Model Intercomparison Project (CMIP) Phase 3 “lead to

⁴The notion of model skill is explored in Chapter 4.

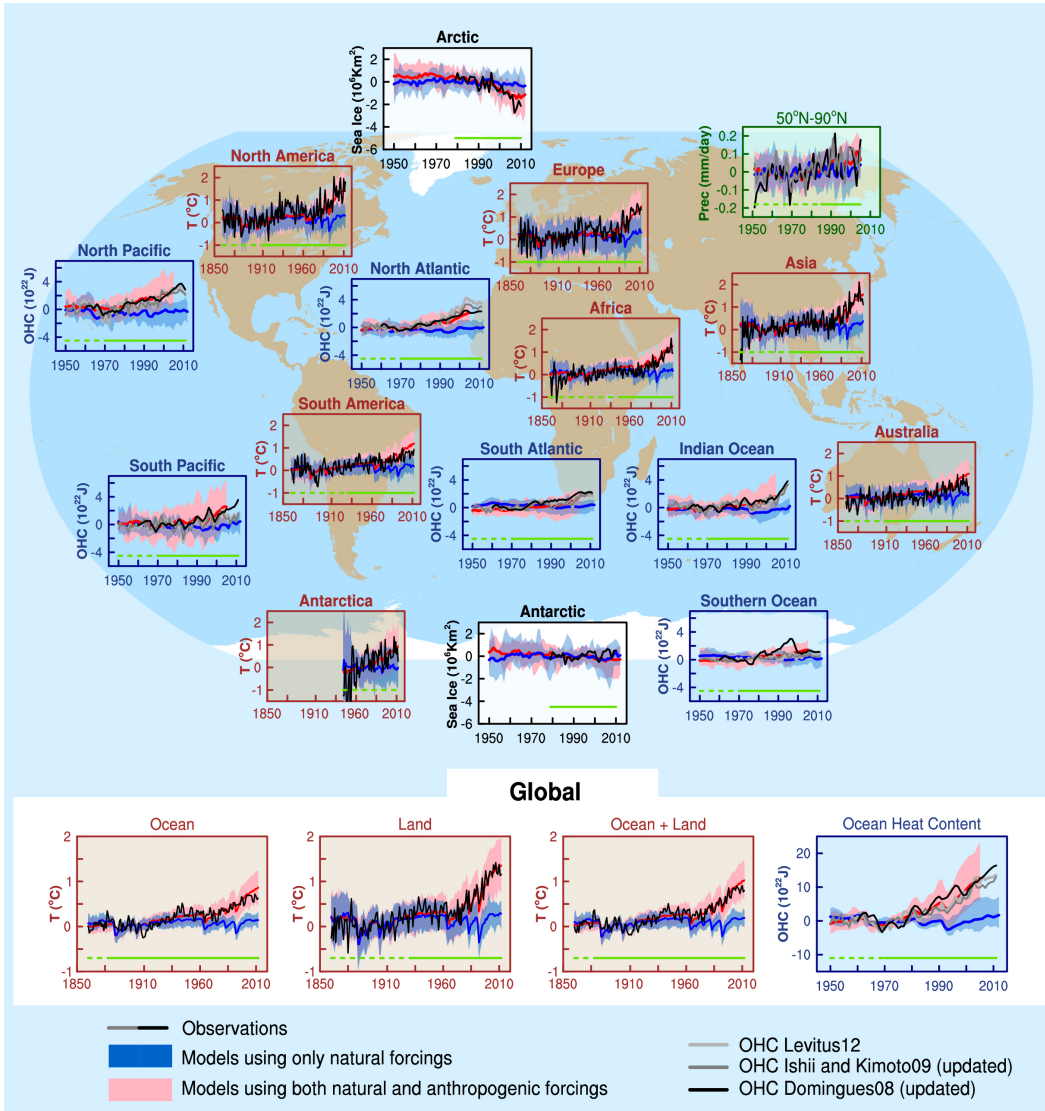


Figure 2.2: Observations are shown on each panel in black or black and shades of grey. Blue shading is the model time series for natural forcing simulations and pink shading is the combined natural and anthropogenic forcings. The dark blue and dark red lines are the ensemble means from the model simulations (Bindoff et al., 2013, p. 930) ©Cambridge University Press.

nontrivial macroscopic errors of simulation,” such that error in the GMST “in a hindcast of the last century casts significant doubt on the viability of the informativeness assumption on a 25-kilometer forecast to the end of this century,” particularly the downscaled regional projections of the 2009 United Kingdom Climate Projections (UKCP09) program. Figure 2.3 illustrates the differences among the 24 models of the CMIP3 in GMST over the twentieth century, raising concerns about the informativeness of regional forecasts that depend on downscaled global projections because of the wide range among the simulations. Whether errors in hindcasts of GMST cast doubt on the plausibility of regional projections, the spread of simulations and the difference between them and observed GMST raises the question of whether such models can be reliable for other purposes, such as that of attributing climate change to human activities. Should model results qualify as evidence of the causes of climate change even though they produce such a wide range of hindcasts with respect to absolute GMST change? What are the limits of how informative climate models can be about the world?

On the one hand, as indicated in Figure 2.3, simulation estimates of absolute GMST change vary widely from one model to another, and between models and the instrumental record, which limits the informativeness of down-scaling general circulation models (GCMs) for regional forecasts. In D&A studies, on the other hand, rather than focusing on absolute GMST change or regional projections, scientists typically concentrate on GMST anomalies with respect to the mean over some baseline period, comparing models programmed with natural and anthropogenic forcing to those that contain natural forcing only (Figure 2.2). Models that include anthropogenic forcing are better than those that exclude it with respect to reproducing

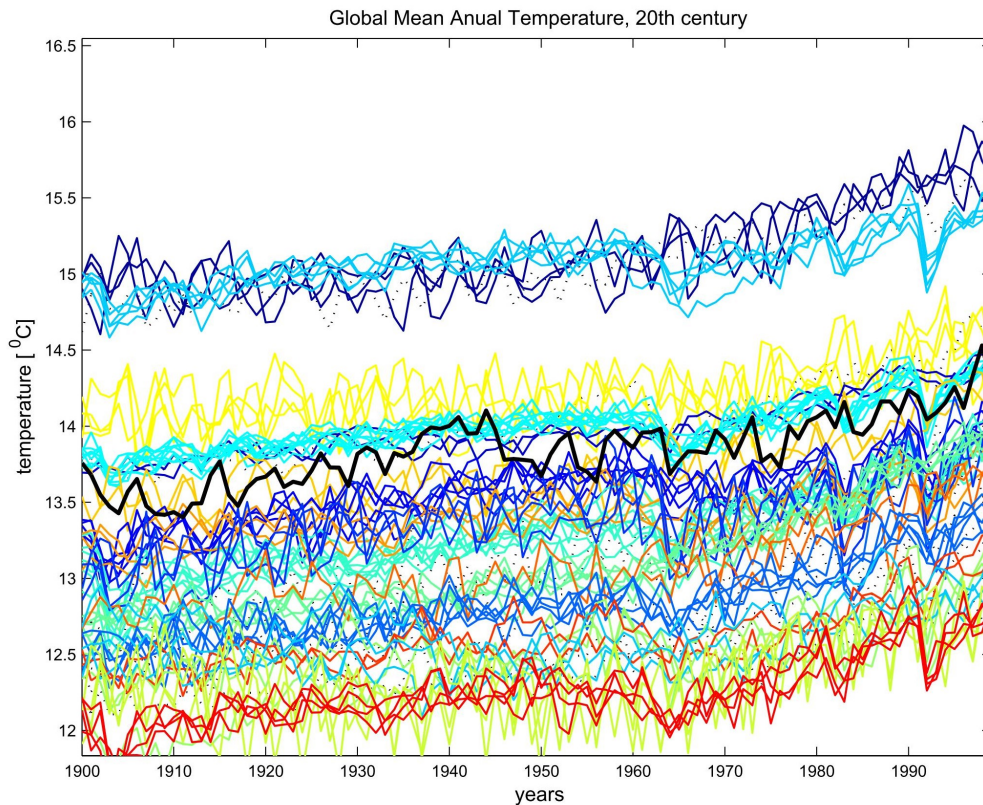


Figure 2.3: Output of Coupled Model Intercomparison Project Phase 3 models (24) simulating changes in annual global mean surface temperature over the twentieth century (coloured lines) and instrumental measurements (black line) (Frigg et al., 2013) ©Philosophy of Science Association.

observed estimates of GMST anomalies, which is a finding that constitutes evidence of the human influence on the climate.

The epistemology of consistent and robust reasoning is exemplified by the use of computer models in attribution studies supporting the anthropogenic global warming hypothesis (AGW)—that is, the proposition is that human activities caused more than half of the observed increase in GMST from 1951–2010 (Bindoff et al., 2013, p. 869). In testing AGW, scientists model the climate in various ways, using models of differing levels of complexity, including comparatively simple energy balance models, energy models of intermediate complexity, GCMs, and earth systems models that contain coupled interacting three-dimensional representations of different components of the climate system, integrating models of the atmosphere, the oceans, the carbon cycle, the cryosphere and the biosphere.

One can study these models and the simulations they produce individually and in climate model ensembles (Parker, 2006). Scientists use a range of methods to assess model-output, applying different statistical analyses to identify relationships between dependent and independent variables, such as GW and anthropogenic activities. AGW is a consistent finding because multiple modelling approaches of different components of the climate system are consistent with AGW and inconsistent with the idea that AGW is incorrect. The results of such modelling studies are also supported by a variety of other evidence, including observations of other changes in the atmosphere, the oceans and the cryosphere (Hegerl et al., 2007; Lloyd, 2009; Parker, 2010; Lloyd, 2010).

Given that each of the modelled worlds is very different from each other and from the earth with respect to their levels of complexity, components of the climate

represented, etc., how do simulations inform scientists about the causes of global warming? The answer to this question rests in the logic of robustness analysis. In cases in which a set of various lines of evidence rely on very rough approximations of a target, if different pieces of evidence, such as those attained using different modelling methods, independently point to the same conclusion, the epistemic security of H can be maintained. By treating the same problem with several alternative modelling methods, scientists have established independent evidential pathways supporting AGW such that, even though each climate model is an approximate representation of the target, each model is imperfect in a different way. If several independent models produce similar results, one can be more confident in the robustness of H because the agreement among these approximations is insensitive to alternative modes of representing a target system. “Hence,” Levins writes, “our truth is the intersection of independent lies” (Levins, 1966, p. 423).

The assumption that each model contributes positive evidential relevance, however, requires justification (Parker, 2011, p. 579). Accordingly, scientists evaluate models by comparing simulation output with observed estimates of the target system. Positive evaluations may justify the use of models in attribution studies, but details about the appropriate metrics for determining model skill continue to be a subject of debate (Flato et al., 2013).

The convergence of multiple lines of evidence supporting AGW exemplifies the way in which robustness analysis compensates for the imperfections in models that constitute evidential pathways. Gareth Jones et al. (2013), for example, investigate the causes of changes in near-surface temperatures from 1860 to 2010, utilizing the HadCRUT4 observational dataset and an ensemble of 8 coupled models from

CMIP5, finding that “calculations of attributable temperature trends based on optimal detection support previous conclusions that the human-induced greenhouse gases dominate observed global warming since the mid-20th century” (p. 4001). Similarly, examining an ensemble of simulations from an energy balance model of intermediate complexity (Bern2.5D), Markus Huber and Reto Knutti (2012, p. 31) conclude that “since the mid-twentieth century, greenhouse gases contributed 0.85°C of warming, . . . suggesting an even higher confidence that human-induced causes dominate the observed warming.”

Further support for AGW emerges from studies of interconnected components of the climate that carry the signal of anthropogenic activities. Jan Lavšovička et al. (2006, p. 1253) find that “the anthropogenic emissions of greenhouse gases influence the atmosphere at nearly all altitudes,” including the upper atmosphere, which has cooled and contracted—a predicted result of greenhouse warming. Comparing a dataset of observed ocean temperatures with simulations from two climate models (PMC and HadCM3), Tsuga Barnett et al. (2005) identify a signal of anthropogenic forcing in the oceans: “A warming signal has penetrated into the world’s oceans over the past 40 years. . . . [The signal] cannot be explained by natural internal climate variability or solar and volcanic forcing, but is well simulated by two anthropogenically forced climate models” (p. 284).

The epistemic security that results from robustness analysis contrasts with what William Wimsatt (1981) calls “serial chain” reasoning. Serial reasoning, on the one hand, involves drawing a conclusion by linking premises in a sequence that is only as strong as its weakest link. The problem with serial chain reasoning, both as a method of induction and as a description of scientific practice, is that, when

an error occurs in one line of evidence, any conclusion resting on it is undermined. Since reliable inferences about the natural world rely on many forms of inquiry, the fragile structure of serial reasoning is incongruent with scientific methodology. Robustness analysis, on the other hand, involves connecting a conclusion to several independent lines of evidence, providing multiple pathways of evidential support, the collection of which results in the insensitivity of H to changes in some subset of the various evidential pathways.

Robustness analysis accounts for the security of scientific inferences such as AGW because, when many lines of evidence converge in support of H, even when imperfections exist in particular lines of evidence, including various modelling methods, alternative evidential pathways provide independent lines supporting the plausibility of H. Even though each line of evidence contains imperfections, if these lines of evidence are imperfect in different ways, one line of imperfect evidence may compensate for the imperfections in another line of evidence. The case of climate change D&A illustrates how the collection of evidential pathways supporting AGW compensates for shortcomings in individual lines of inquiry. The logic of consilience and robust evidential reasoning thus explains how imperfect models can be informative in studies of the causes of climate change.

2.4 Conclusion

Studies of climate change have provided strong and secure support for GW and AGW. In these cases, the independence of different lines of evidence is due to different types of observations, datasets, models and methodology that scientists incor-

porate in their studies of climate change. The convergent lines of evidence supporting GW include a wide range of observations, such as measurements of changing surface atmospheric and oceanic temperatures, glaciers mass, snow cover, sea ice, sea level and atmospheric water vapour. Investigating the causes of climate change, scientists test AGW by using multiple computer modelling methods that, in conjunction with a variety of other lines of evidence, including observations of many components of the climate system, comprise as a consilience of inductions because these observations and modelling results would be unlikely if AGW were incorrect. Despite differences between the actual climate system and computer model representations of it, models provide important sources of information when model results are robust to variations in the way they approximate the climate.

Bibliography

- Barnett, T., Zwiers, F., Hegerl, G., Allen, M., Crowley, T., Gillett, N., Hasselmann, K., Jones, P., Santer, B., Schnur, R., et al. (2005). Detecting and attributing external influences on the climate system: A review of recent advances. *Journal of Climate*, 18(9):1291–1314.
- Bindoff, N., Stott, P., AchutaRao, K., Allen, M., Gillett, N., Gutzler, D. and Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I., Overland, J., Perlwitz, J., Sebbari, R., and Zhang, X. (2013). *Detection and attribution of climate change: From global to regional. In: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.

- Callendar, G. S. (1939). The composition of the atmosphere through the ages. *Meteorological Magazine*, 74:33–39.
- Edwards, P. N. (2010). *A vast machine*. MIT Press.
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M. (2013). *Evaluation of climate models. In: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Fleming, J. R. (1998). *Historical Perspectives on Climate Change*. Oxford University Press.
- Frigg, R., Smith, L. A., and Stainforth, D. A. (2013). The myopia of imperfect climate models: the case of UKCP09. *Philosophy of science*, 80(5):886–897.
- Hartmann, D., Tank, A. K., Rusticucci, M., Alexander, L., Bronnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., and Zhai, P. (2013). *Observations: Atmosphere and Surface. In Climate Change: The Physical Science Basis : Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge Univ Press.
- Hegerl, G. C., Crowley, T. J., Allen, M., Hyde, W. T., Pollack, H. N., Smerdon, J., and Zorita, E. (2007). Detection of human influence on a new, validated 1500-year temperature reconstruction. *Journal of Climate*, 20(4):650–666.
- Huber, M. and Knutti, R. (2012). Anthropogenic and natural warming inferred from changes in earth's energy balance. *Nature Geoscience*, 5(1):31–36.

- Jones, G. S., Stott, P. A., and Christidis, N. (2013). Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, 118(10):4001–4024.
- Lavštovička, J., Akmaev, R. A., Beig, G., Bremer, J., and Emmert, J. T. (2006). Global change in the upper atmosphere. *Science*, 314(5803):1253–1254.
- Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54(4):421–431.
- Lewis, S. L. (2010). How to beat the media in the climate street fight. *Nature*, 468(7320):7.
- Lloyd, E. (2009). I–Varieties of support and confirmation of climate models. In *Aristotelian Society Supplementary Volume*, volume 83, pages 213–232.
- Lloyd, E. (2010). Confirmation and robustness of climate models. *Philosophy of Science*, 77(5):971–984.
- McGuffie, K. and Henderson-Sellers, A. (2014). *The Climate Modelling Primer*. John Wiley & Sons.
- Meehl, G., Stocker, T.F., Collins, W., Friedlingstein, P., Gaye, A., Gregory, A., Kitoh, J., Knutti, R., Murphy, J., Noda, A., Raper, S., Watterson, I., Weaver, A., and Zhao, Z.-C. (2007). *Global Climate Projections*. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, volume 4. Cambridge University Press.
- Parker, W. (2006). Understanding pluralism in climate modeling. *Foundations of*

- Science*, 11(4):349–368.
- Parker, W. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science*, 78(4):579–600.
- Parker, W. S. (2010). Comparative process tracing and climate change fingerprints. *Philosophy of Science*, 77(5):1083–1095. ArticleType: research-article / Full publication date: December 2010 / Copyright 2010 Philosophy of Science Association.
- Plass, G. (1956). Carbon dioxide and the climate. *American Scientist*, 44(3):302–316.
- Revelle, R. and Suess, H. E. (1957). Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO₂ during the past decades. *Tellus*, 9(1):18–27.
- Staley, K. W. (2004). Robust evidence and secure evidence claims. *Philosophy of Science*, 71(4):467–488.
- Weart, S. R. (2011). *Discovery of Global Warming (rev. and Expanded Ed.)*. Diane Publishing Company.
- Weisberg, M. (2006). Robustness analysis. *Philosophy of Science*, 73(5):730–742.
- Whewell, W. (1858). *Novum Organon Renovatum*. J.W. Parker and Son.
- Wimsatt, W. C. (1981). Robustness, reliability, and overdetermination (2012)). In *Characterizing the Robustness of Science*, pages 61–87. Springer.

Chapter 3

Consilience, Robustness, and the Hockey Stick Hypothesis

3.1 Introduction

Some theories of evidential reasoning suggest that the concepts of consilience and robustness can account for the evidential strength and security of scientific inferences. In order to address the question of how consilience and robustness relate to each other, my account associates evidential strength with consilience and security with robustness. Analyzing connections between consilient reasoning and robust evidential support, this paper investigates case studies in climate science supporting the claim that Northern Hemispheric (NH) mean surface temperatures were higher during the last few decades than during any comparable period over at least the preceding 400 years—a proposition that, for reasons discussed below, I

call the ‘hockey stick hypothesis’ (HSH).¹ My thesis is that a consilience hypothesis (H) is conducive to its having robust evidential support. Examining studies of climate proxy indicators, which confer some degree of consilience to HSH, I elaborate on the connection between consilience and robustness with respect to debates about global warming. This case also highlights the importance of epistemology as regards the ethics and politics of evidence-based decision-making, because some opponents of climate change mitigation efforts have mischaracterized the epistemology of science in relation to HSH, which is clarified by the logic of consilience and robustness.

Tree-ring, glacier, borehole and other proxies qualify as relevantly independent sources of evidence supporting HSH because one can infer HSH from one source without relying on the same set of assumptions required to draw this conclusion from another source. The various evidential pathways supporting HSH result in a consilience of inductions because it would be unlikely that such different proxies would indicate HSH if HSH were incorrect. This case illustrates how consilience is conducive to robustness since the different evidential pathways that provide consilient support for HSH also provide epistemic security in the face of challenges to a subset of this evidence.

Consilient reasoning and robustness analysis involve connecting a conclusion to a set of several independent lines of evidence, providing multiple pillars of support, each of which contributes to the strength and security of H, contrasting with serial chain reasoning, which is the view that, when an error occurs in one line of

¹The paleoclimate reconstructions discussed below extend from about 400 to 2,000 years ago and, as will be noted, the extent of the current warming in relation to past climates differs from one study to another, depending on how far back a given reconstruction goes, and the baseline of comparison.

evidence, any conclusion resting on it is undermined (Wimsatt, 1981). The fragile structure of serial reasoning fails to capture the nature of scientific methodology because reliable inferences about the natural world rely on many forms of inquiry, which is a point exemplified by debates about HSH.

3.2 Consilience

Evidential strength is the degree to which data indicate the correctness of H (Staley, 2004; Parker, 2011). Consilience is a mode of reasoning that involves assigning a high degree of plausibility to H when it is supported by a diverse set of evidence, which contains lines of evidence that are, to some degree, relevantly independent, such that it would be unlikely to these lines of evidence to converge unless H were correct (Whewell, 1840, 1858, 1860). Mere agreement between different lines of evidence is insufficient for increasing the strength of H. A strong inference occurs when it is highly improbable that a given line of evidence (E_1) would be in agreement with another line of evidence (E_2) if H were incorrect. Accordingly, the success of this epistemology depends on each line of evidence being relevantly independent from each other.² Failures of relevant epistemic independence occur when both E_1 and E_2 rely on a single assumption, the failure of which undermines the reliability of H. In this respect, the epistemology of consilient reasoning helps explain how, and to what extent, scientists gain confidence in the plausibility of

²Relevant epistemic dependence is distinct from social dependence, as when members of a group relying on collaborative research projects, which Edward Wegman et al. (2006, p. 4) suggest is problematic: “authors in the area of paleoclimate studies are closely connected and thus ‘independent studies’ may not be as independent as they might appear on the surface.”

various claims, including those pertaining to the distant past.³

Consilience operates in the many disciplines of climate science, including paleoclimatology, which has accumulated various kinds of evidence pertaining to hypotheses about climate change, like HSH. This conclusion is associated with what have come to be known as ‘hockey stick temperature curves;’ a term used to describe the warming pattern that the Intergovernmental Panel on Climate Change (IPCC) first highlighted in its Third Assessment Report (TAR).⁴ In this report, the IPCC reproduces a graph from a 1999 paper authored by Micheal Mann, Raymond Bradley and Malcolm Hughes (MBH),⁵ showing a pattern of NH temperature change that slightly cools from about 1000 CE until about 1900 CE (the ‘shaft’ of the hockey stick) followed by a sharp warming that continues to the present day (the upward ‘blade’) (Figure 3.1). The work of MBH serves as an interesting case study in the epistemology of science because their results have been scrutinized extensively.

Exemplifying the idea that consilience adds strength to an inference, the IPCC qualifies its statements related to HSH with expressions of high confidence.⁶ Considering the work of MBH and others, the TAR notes that “[i]ndependent estimates of hemispheric and global ground temperature trends over the past five centuries

³For philosophical discussions about the confirmation of historical and prehistorical hypotheses, see Gould (1989, 2002, 2003); Cleland (2001, 2002, 2011); Turner (2004, 2007); Forber (2011); Tucker (2011).

⁴Although the climatologist Jerry Mahlman is reported to have coined this use of the term, it had previously been associated with similar climatological data-series patterns, such ozone depletion (Biello, 2012, p. 74).

⁵MBH09 refers to Mann et al. (1999), MBH08 refers to Mann et al. (1998), and MBH refers to both papers.

⁶For details about the IPCC’s likelihood and confidence terminology in the TAR, AR4 and AR5, see Schneider and Moss (1999), IPCC AR4 Appendix (2007) and Mastrandrea et al. (2010), respectively.

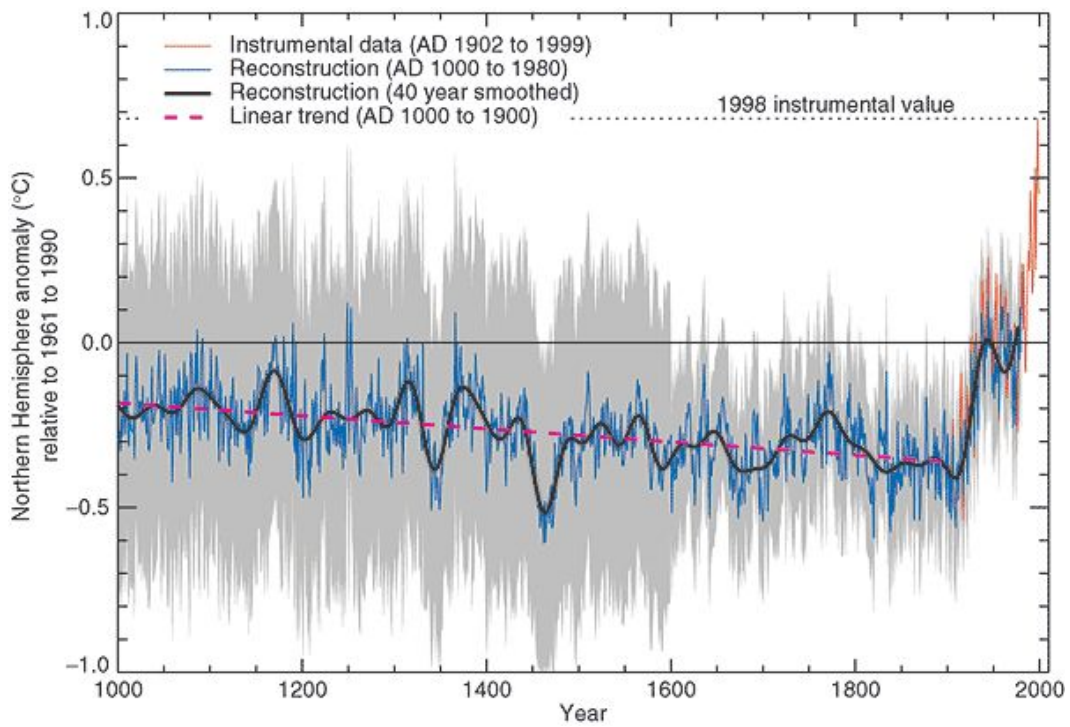


Figure 3.1: Millennial Northern Hemisphere (NH) temperature reconstruction (blue) and instrumental data (red) from AD 1000 to 1999, adapted from Mann et al. (1999). Smoother version of NH series (black), linear trend from AD 1000 to 1850 (purple-dashed) and two standard error limits (grey shaded) are shown. ©Cambridge University Press, (Folland et al., 2001, p. 134).

from sub-surface information contained in borehole data confirm the conclusion that the late 20th century warmth is anomalous in a longterm context (Folland et al., 2001, p. 133). In its Fourth Assessment Report (AR4), the IPCC states that “[i]t is *very likely* that average Northern Hemisphere temperatures during the second half of the 20th century were higher than for any other 50-year period in the last 500 years” (Jansen et al., 2007, p. 436). Similarly, in its Fifth Assessment Report (AR5), the IPCC concludes that “[f]or average annual NH temperatures, the period 1983–2012 was *very likely* the warmest 30-year period of the last 800 years” (Masson-Delmotte et al., 2013, p. 386). Figure 3.2 illustrates a convergence of multiple evidential pathways supporting the HSH, some of which make use of entirely different methodologies and proxies discussed below.

Since people only began taking instrumental measurements of climatic variables in the late seventeenth century (Fleming, 1998; Weart, 2011), scientists investigating climate conditions before this time rely on *proxy* climate indicators; i.e., preserved physical characteristics of the past that stand in for instrumental measurements. Interpreted according to physical and biophysical principles, proxies serve as a means of measuring historical and prehistorical phenomena in the absence of systematic instrumental records (Folland et al., 2001, p. 130).

The strength of the IPCC’s inferences related to HSH is due to a wide range of available supporting evidence. Scientists infer information about surface temperature change by, for example, measuring the width and density of tree rings (Jones et al., 1998; Briffa, 2000). Cores extracted from trees provide annually resolved time series of wood properties. Since, under certain conditions, temperature changes influence tree ring growth, trees from particular locations provide a means

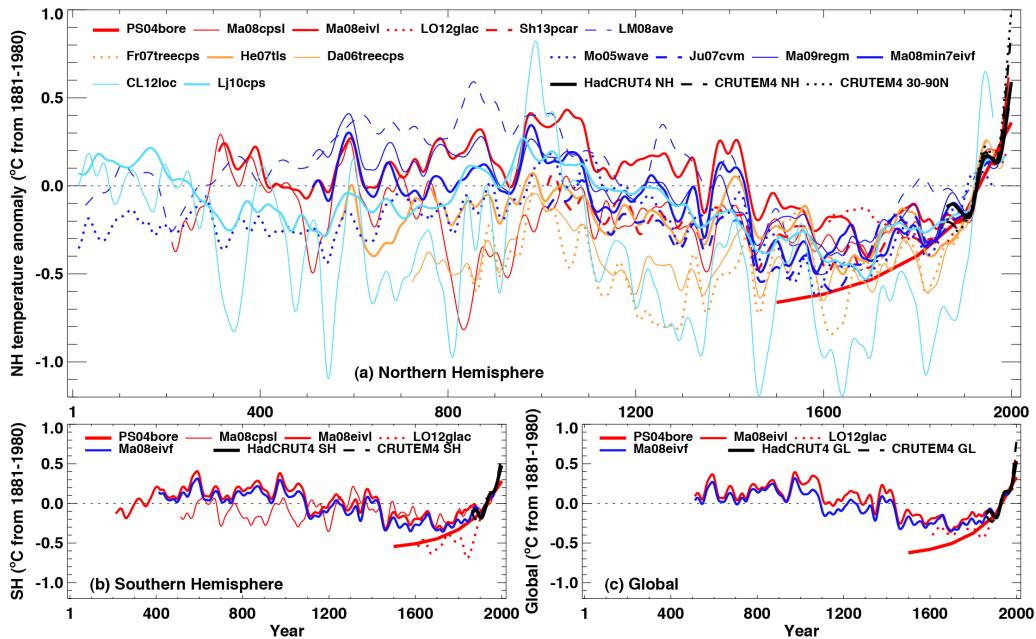


Figure 3.2: Reconstructed (a) Northern Hemisphere and (b) Southern Hemisphere, and (c) global annual temperatures during the last 2000 years. Individual reconstructions (see Appendix [IPCC AR5] 5.A.1 for further information about each one) are shown as indicated in the legends, grouped by colour according to their spatial representation (red: land-only all latitudes; orange: land-only extratropical latitudes; light blue: land and sea extra-tropical latitudes; dark blue: land and sea all latitudes) and instrumental temperatures shown in black (Hadley Centre/ Climatic Research Unit (CRU) gridded surface temperature-4 data set (HadCRUT4) land and sea, and CRU Gridded Dataset of Global Historical Near-Surface Air Temperature Anomalies Over Land version 4 (CRUTEM4) land-only; Morice et al., 2012). All series represent anomalies ($^{\circ}\text{C}$) from the 1881–1980 mean (horizontal dashed line) and have been smoothed with a filter that reduces variations on time scales less than about 50 years.

of reconstructing temperature change over time. Also, since earth surface temperature variations diffuse downward with time, scientists can estimate past surface temperatures by measuring the vertical temperature profile down boreholes drilled from rock, soil, and ice. Since the TAR, other borehole studies have arrived at similar conclusions, estimating that the NH average surface temperatures increased 1 K from CE 1500 to 2000 (Pollack and Smerdon, 2004). Furthermore, records of the lengths of mountain glaciers extending hundreds of years can be incorporated in models of glacier dynamics, which also draw on records of precipitation and variables, allowing scientists to relate changes in glacial extent to changes in local temperatures. Others reconstruct annual average surface temperatures of the past 400 years on hemispherical and global scales by studying fluctuations in glacier length over this time, an approach that “confirms the pronounced warming of the twentieth century, giving a global cumulative warming of 0.94–0.31 K over the period 1830–2000 and a cumulative warming of 0.84 ± 0.35 K over the period 1600–2000” (Leclercq and Oerlemans, 2012). Other studies combine several types of paleo-climate proxies. The PAGES2k project, for instance, reconstructs continental-scale temperature variability during the past two millennia using a data set that includes 511 time series of tree rings, pollen, corals, lake and marine sediments, glacier ice, speleothems and historical documents. This similarly study concludes that “[r]ecent warming reversed the long-term cooling; during the period AD 1971–2000, the area-weighted average reconstructed temperature was higher than any other time in nearly 1,400 years” (Ahmed et al., 2013, p. 339).

In the case of paleoclimate reconstructions, tree-ring, glacier, borehole and other proxies qualify as relevantly independent sources of evidence because in-

ferences drawn from one source can be maintained without relying on the same set of assumptions required to draw an inference from another source. The set of diverse evidential pathways supporting these inferences results in a consilience of inductions supporting HSH because it would be unlikely that these different proxies would carry a signal indicating HSH if HSH were incorrect. As the following section explains, the conducive relationship between consilience and robustness is exemplified by the epistemic security of HSH in the face of challenges to MBH.

3.3 Robustness

Epistemic security is the idea that the plausibility of H is insensitive to changes in a particular evidential pathway (Staley, 2004; Parker, 2011). Robustness analysis involves identifying a collection of independent lines of evidence supporting H such that the plausibility of H is insensitive to vulnerabilities in a subset of the total available evidence (Wimsatt, 1981).⁷ If different lines of evidence are sensitively dependent on a common assumption, such as a method of data calibration or a particular interpretation of physical principles, and that assumption were discovered to be fundamentally mistaken, the evidential network previously supporting the H could collapse. The more diverse a set of evidence is, however, the less likely it is that some new evidence will undermine the entire set. Furthermore, while a robust inference is insensitive to errors in a given evidential pathway, the individual plausibility and relevant independence of each line of evidence are critical for es-

⁷Brett Calcott (2011) identifies three senses of ‘robustness’ in Wimsatt’s account: “robust theorems,” “robust phenomena” and “robust detection.” The latter idea is the one most relevant to paper.

tablishing H. In the case of paleoclimate reconstructions, each line of evidence is limited by several factors. Confidence in earth surface temperature reconstructions, for example, weakens the further back in time one's claim extends because weathering destroys proxies (North et al., 2006). Interpreting proxies is also complicated by other environmental influences. Tree ring growth, for instance, is influenced by myriad of environmental variables, including local temperatures, precipitation, soil nutrients, and atmospheric CO₂ levels. Accordingly, Briffa (2000) notes that the "activities of humans may well be impacting on the 'natural' growth of trees in different ways, making the task of isolating a clear climate message subtly difficult."⁸ For this reason, it is important to test the robustness of each evidential pathway leading to H and consider as much evidence as possible.

The MBH99 hockey-stick reconstruction became iconic of the IPCC TAR. Its Working Group 1 highlights the graph several times, featuring it in the summary for policy makers (TAR-Fig. 1(b), p. 3), Technical Summary (TAR-Fig. 5, p. 29), Chapter 2: Observations (TAR-Fig. 2.20 p. 134), and in combinations with other reconstructions (e.g., Fig. 2.21, p. 134). In the midst of the publicity of the IPCC, several studies challenged the methods, proxies and findings of MBH, some of the most publicized and controversial of which were authored by Willie Soon and Sallie Baliunas (Soon and Baliunas, 2003; Soon et al., 2003) (SB03), who maintain that "the 20th century is probably not the warmest nor a uniquely extreme climatic period of the last millennium" (Soon and Baliunas, 2003, p. 89).⁹

⁸This issue has come to be known as the 'divergence' between instrumental and certain tree-ring proxies.

⁹The significance of SB03 to political debates about climate policy demonstrates the relevance of epistemology to moral and political decision-making: During a U.S. Senate debate on 28 July 2003, Senator James M. Inhofe referenced SB03 while opposing the Climate Stewardship Act, which proposed a policy for CO₂ restrictions. "With all of the hysteria," Inhofe concluded, "all of the

The political situation surrounding MBH led the U.S. Congress Representative and chairman of the U.S. House of Representatives Committee on Science Sherwood Boehlert to initiate a special committee assembled by the National Research Council (NRC) to investigate the science of paleoclimate reconstructions and produce a report on “Surface Temperature Reconstructions for the Past 2,000 Years” (the North Report). The committee examined the available evidence from instrumental and proxy records, including those consisting of data from tree rings, corals, boreholes, marine sediments, speleothems, ice cores, and climate models. Considering the wide variety of evidence regarding paleoclimate change, the investigation found that the basic conclusion of MBH “has subsequently been supported by an array of evidence that includes both additional large-scale surface temperature reconstructions and pronounced changes in a variety of local proxy indicators, such as melting on ice caps and the retreat of glaciers around the world, which in many cases appear to be unprecedented during at least the last 2,000 years” (North et al., 2006, p. 3)

Since reconstructions that use tree-ring, glacier, borehole and other proxies

fear, all of the phony science, could it be that man-made global warming is the greatest hoax ever perpetrated on the American people? It sure sounds like it” (Inhofe, 2003). On another occasion, Inhofe repeated this claim, stating “the more I checked into it, the things started with the United Nations, the International Panel on Climate Control [*sic.*], and they used one scientist. And his name was Michael Mann, the famous hockey stick—remember that. . . [I]n all of the recent science, as I’ve mentioned on your radio show, it confirms that I was right on this thing. This thing is a hoax.” (This video and transcript of Inhofe’s interview with Glenn Beck on CNN Headline News are available at <http://thinkprogress.org/politics/2006/07/21/6435/inhofe-gore/>.) Mann et al. (2003) responds to SB03 with three key points: (1) SB03 fail to assess the sensitivity of proxy data to past temperature variability, (2) they take regional temperature changes to indicate hemispheric changes across overly broad intervals of time (800–1300), without indicating whether warm anomalies in different regions were synchronous and (3) they take as their base period for comparison the mean temperatures over the whole of the 20th century, and they reconstruct past temperatures from proxies without resolving the data according to decadal trends. For details about the political debate about MBH, see Pearce (2010) and Mann (2012)

serve as independent lines of evidence supporting HSH, even if a reconstruction that uses one type of proxy is discovered to be fundamentally flawed, other studies provide independent support for this finding. So, despite a host of criticisms that have been made against MBH, the report concluded that “[i]t can be said with a high level of confidence that global mean surface temperature was higher during the last few decades of the 20th century than during any comparable period during the preceding four centuries” (North et al., 2006, p. 118).

Similar conclusions have been reached by other studies that investigate the proxies and methodology of MBH. Stephen McIntyre and Ross McKittrick (2005) (MM05b) maintain that MBH results for the fifteenth century exhibit a “pivotal dependence” on a relatively small subset of proxy records, particularly those derived from bristlecone pines, which “are widely doubted to be reliable temperature proxies” (p. 69). They hold that the hockey stick pattern in proxy principal component summaries for North America is disproportionately carried by the bristlecone pine proxies. Eugene Wahl and Caspar Ammann (2007) (WA07), however, demonstrate the robustness of MBH results with respect to the bristlecone proxies by noting the following two points. (1) They test whether excluding these proxies has a discernible effect on NH mean surface temperature reconstructions, especially in the fifteenth century as claimed by MM05b, concluding that, regardless of whether these records carry meaningful large-scale temperature information, their exclusion has relatively little effect on the magnitude and trajectory of NH temperature reconstructions over the last 600 years (pp. 43–44). So, whether these data inappropriately lead to the hockey-stick result “ceases to be of significant import.” By verifying the robustness of MBH against several criticisms of their methods and

proxies, WA07 provide evidence for HSH: “both the 20th century upward trend and high late-20th century hemispheric surface temperatures are anomalous over at least the last 600 years” (p. 33–34). (2) In response to MM05b’s suggestion that these proxies are problematic with respect to identifying local temperatures, WA07 note that the relationship between bristlecone-pines and NH temperatures can be interpreted according to more complex and indirect connections among climate parameters, such as those among tree-ring growth, precipitation, and El Niño Southern Oscillation (ENSO). “MBH do not claim that all proxies in their reconstruction are closely related to local-site variations in surface temperature. Rather, they invoke a less restrictive assumption that “whatever combination of local meteorological variables influence the proxy record, they find expression in one or more of the largest-scale patterns of annual climate variability” [Mann et al. 2000, p. 330] against which proxy records are calibrated in the reconstruction process” (Wahl and Ammann, 2007, p. 37).

These examples illustrate two ways of demonstrating the robustness of HSH. The first defense identifies alternative sources of primary evidence supporting HSH, which act as backup evidence, indicating that this conclusion is secure even if problems exist in the approach of MBH. The second defense responds directly to criticisms of MBH, providing evidence of HSH by reinforcing the approach of MBH.

3.4 Conclusion

Consilient reasoning is conducive to robust inferences because, when many lines of evidence converge to strengthen the plausibility of H, the epistemic security of

H is enhanced by these alternative evidential pathways. The consilience of HSH illustrates the connection between consilience and robustness because the converging lines of evidence that provide HSH with epistemic strength also provide it with epistemic security against criticisms that have targeted one of the many evidential pathways leading to this finding.

This epistemology also relates to ethical and political debates about climate change. Due to the social, political and moral implications of anthropogenic climate change, the results of paleoclimate studies related to global warming have become the subject of scientific scrutiny and political controversy. Having been targeted in political debates about climate change, Bradley (2011, p. 4) describes how the work of MBH was “chosen as a sacrificial lamb” by people and organizations attempting to discredit the science of global warming. “[A]ntagonists seemed to believe that if they could refute [MBH], the entire edifice of global warming science would crumble and fall.” While serial chain reasoning has led to the targeting individual lines of evidence and individual scientists, an epistemology of consilience and robustness analysis can provide strength and security to evidence-based ethical decisions-making.

Bibliography

Ahmed, M., Anchukaitis, K., Buckley, B. M., Braida, M., Borgaonkar, H. P., Asrat, A., Cook, E. R., Bntgen, U., Chase, B. M., Christie, D. A., and others (2013). Continental-scale temperature variability during the past two millennia: Supplementary information. *Nature geoscience*, 6(5).

- Biello, D. (2012). Hit them with the hockey stick. *Scientific American*, 306(3):72–75.
- Bradley, R. S. (2011). *Global warming and political intimidation: How politicians cracked down on scientists as the earth heated up*. Univ of Massachusetts Press.
- Briffa, K. R. (2000). Annual climate variability in the holocene: interpreting the message of ancient trees. *Quaternary Science Reviews*, 19(1–5):87–105.
- Calcott, B. (2011). Wimsatt and the robustness family: Review of Wimsatt’s Re-engineering Philosophy for Limited Beings. *Biology & Philosophy*, 26(2):281–293.
- Cleland, C. E. (2001). Historical science, experimental science, and the scientific method. *Geology*, 29(11):987.
- Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science*, 69(3):474–496.
- Cleland, C. E. (2011). Prediction and explanation in historical natural science. *The British Journal for the Philosophy of Science*, 62(3):551–582.
- Fleming, J. R. (1998). *Historical Perspectives on Climate Change*. Oxford University Press.
- Folland, C., , Karl, T., Christy, J., Clarke, R., Gruza, G., Jouzel, J., Mann, M., Oerlemans, J., Salinger, M. J., and Wang, S. W. (2001). *Climate Change 2001: The Scientific Basis*. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
- Forber, Patrick, E. G. (2011). Historical reconstruction: Gaining epistemic access

- to the deep past. *Philosophy & Theory in Biology*.
- Gould, S. J. (1989). *Wonderful life: the Burgess Shale and the nature of history*.
- Gould, S. J. (2002). *The structure of evolutionary theory*. Belknap Press.
- Gould, S. J. (2003). *The hedgehog, the fox, and the magister's pox: Mending the gap between science and the humanities*. Random House LLC.
- Inhofe, J. M. (2003). Sen. inhofe delivers major speech on the science of climate change; catastrophic global warming alarmism not based on objective science, u.s. senator for oklahoma.
- IPCC AR4 Appendix (2007). "Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties." In *Climate Change 2007: The Physical Science Basis : Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Jansen, E., Overpeck, J., Briffa, K., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W., Rahmstorf, S., Ramesh, R., Raynaud, D. Rind, D. S. O. V. R., and Zhang, D. (2007). *Paleoclimate. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, volume 4. Cambridge University Press.
- Jones, P. D., Briffa, K. R., Barnett, T. P., and Tett, S. F. B. (1998). High-resolution palaeoclimatic records for the last millennium: interpretation, integration and comparison with general circulation model control-run temperatures. *The Holocene*, 8(4):455–471.

- Leclercq, P. W. and Oerlemans, J. (2012). Global and hemispheric temperature reconstruction from glacier length fluctuations. *Climate Dynamics*, 38(5-6):1065–1079.
- Mann, M., Amman, C., Bradley, R., Briffa, K., Jones, P., Osborn, T., Crowley, T., Hughes, M., Oppenheimer, M., Overpeck, J., and others (2003). On past temperatures and anomalous late-20th-century warmth. *Eos, Transactions American Geophysical Union*, 84(27):256–256.
- Mann, M. E. (2012). *The Hockey Stick and the Climate Wars: Dispatches from the Front Lines*. Columbia University Press.
- Mann, M. E., Bradley, R. S., and Hughes, M. K. (1998). Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, 392(6678):779–787.
- Mann, M. E., Bradley, R. S., and Hughes, M. K. (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophysical Research Letters*, 26(6):759.
- Mann, M. E., Bradley, R. S., and Hughes, M. K. (2000). Long-term variability in the el nino southern oscillation and associated teleconnections. *El Niño and the Southern Oscillation: Multiscale variability and global and regional impacts*, pages 357–412.
- Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Gonzalez Rouco, J., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., and Timmermann, A. (2013). *Information from Paleoclimate Archives*. In: *The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Re-*

- port of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., et al. (2010). Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. *Intergovernmental Panel on Climate Change (IPCC)*.
- McIntyre, S. and McKittrick, R. (2005). The M&M critique of the MBH98 northern hemisphere climate index: Update and implications. *Energy & Environment*, 16(1):69–100.
- North, G. R., Biondi, F., Bloomfield, P., Christy, J., Cuffey, K., Dickinson, R., Druffel, E., Nychka, D., Otto-Bliesner, B., Roberts, N., et al. (2006). Surface temperature reconstructions for the last 2,000 years. *NRC Statement to Subcommittee on Oversight and Investigations, Committee on Energy and Commerce, US House of Representatives*, http://www7.nationalacademies.org/ocga/testimony/Surface_Temperature_Reconstructions.asp.
- Parker, W. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science*, 78(4):579–600.
- Pearce, F. (2010). *The climate files: the battle for the truth about global warming*. Guardian Books London.
- Pollack, H. N. and Smerdon, J. E. (2004). Borehole climate reconstructions: Spatial structure and hemispheric averages. *Journal of Geophysical Research: Atmospheres*, 109(D11):D11106.
- Schneider, S. and Moss, R. (1999). Uncertainties in the IPCC TAR. recommenda-

- tions to lead authors for more consistent assessment and reporting. *Unpublished document*.
- Soon, W. and Baliunas, S. (2003). Proxy climatic and environmental changes of the past 1000 years. *Climate Research*, 23(2):89–110.
- Soon, W., Baliunas, S., Idso, C., Idso, S., and Legates, D. R. (2003). Reconstructing climatic and environmental changes of the past 1000 years: a reappraisal. *Energy & Environment*, 14(2):233–296.
- Staley, K. W. (2004). Robust evidence and secure evidence claims. *Philosophy of Science*, 71(4):467–488.
- Tucker, A. (2011). *A Companion to the Philosophy of History and Historiography*. John Wiley & Sons.
- Turner, D. (2004). The past vs. the tiny: historical science and the abductive arguments for realism. *Studies In History and Philosophy of Science Part A*, 35(1):1–17.
- Turner, D. (2007). *Making prehistory: historical science and the scientific realism debate*. Cambridge University Press.
- Wahl, E. and Ammann, C. (2007). Robustness of the Mann, Bradley, Hughes reconstruction of northern hemisphere surface temperatures: Examination of criticisms based on the nature and processing of proxy climate evidence. *Climatic Change*, 85(1):33–69.
- Weart, S. R. (2011). *Discovery of Global Warming (rev. and Expanded Ed.)*. Diane Publishing Company.
- Wegman, E. J., Scott, D. W., and Said, Y. H. (2006). *Ad Hoc Committee Report on*

the 'Hockey Stick' Global Climate Reconstruction. United States House Committee on Energy and Commerce.

Whewell, W. (1840). *The philosophy of the inductive sciences: founded upon their history*. London: John W. Parker, West Strand.

Whewell, W. (1858). *Novum Organon Renovatum*. J.W. Parker and Son.

Whewell, W. (1860). *On the philosophy of discovery, chapters historical and critical*;. London, J.W. Parker and son.

Wimsatt, W. C. (1981). Robustness, reliability, and overdetermination (2012)). In *Characterizing the Robustness of Science*, pages 61–87. Springer.

Chapter 4

Averaging Estimates: The Wisdom of Multiple Climate Models

4.1 Introduction

Climate models are approximate representations of the climate system, expressed as differential equations based on known physical, chemical, and biological principles, discretized to run on a digital computer. Climate models vary in complexity, differing in the number of spatial dimensions they contain, and the extent to which various processes are explicitly represented. At the more comprehensive end of the spectrum, General Circulation Models (GCMs) are three-dimensional representations of interacting components of the climate system. Earth System Models (ESMs) couple GCMs of the earth's atmosphere with models of its oceans, the carbon cycle, the cryosphere, the biosphere, and other components of the system.

Many current lines of research into the causes, effects and trajectories of climate

change rely on climate models (McGuffie and Henderson-Sellers, 2014). One way scientists use these tools is by incorporating simulations into ensembles, comparing and combining results to explore and reduce uncertainties about various components of the target system. Forming a key source of information for Intergovernmental Panel on Climate Change (IPCC), the Coupled Model Intercomparison Project (CMIP) coordinates scientists from around the world to conduct computer simulation experiments and share their results. Using different models with overlapping input variables, CMIP compiles simulation output provided by community members (Taylor et al., 2012).

An underlying assumption of certain climate ensemble studies is that convergent results among relevantly independent models can reduce uncertainties regarding questions about climate change. Some studies also assume that averaging outputs from multiple simulations is a way of canceling out random errors, thereby detecting signals or producing projections more reliably than any individual simulation. In light of the extensive human, computational, and financial resources required to produce ensembles such as those of CMIP, a question to consider is: Would it be better for researchers to pool their resources in developing a single ‘best model’ rather than a plurality of different models? This chapter argues that incorporating the results of a plurality of models offers important epistemic advantages over employing only one model, insofar as ensemble members are weighted to reflect their skill and independence.

Our procedure is as follows. First, we describe several uses of climate model ensembles, focusing on cases in which multi-model averages improve the reliability of estimated climate change, particularly examples in climate model evaluation,

detection and attribution (D&A) and projection studies. We then compare climate ensemble approaches to a simple formulation of combining multiple estimates of a quantity. This comparison illustrates an important lesson: In forming best estimate averages, the optimal weights to be used are unequal weights if models are of unequal skill or exhibit correlated model errors.

4.2 Climate Model Ensembles

A climate model ensemble is a set of simulations representing aspects of the climate, allowing scientists to compare and combine model outputs and providing potential ways of enhancing modelling capabilities and scientific understanding of climatic phenomena. Different evolutions of a modelled system result from stochastic processes, variations in initial conditions and alternative model designs, so examining many different models or many runs of a single model may help to address questions about structural errors in a model or set of models and uncertainties associated with the internal variability of the climate. A multi-model ensemble (MME) is a collection of different models of similar structure and complexity which can be used to sample internal variability and structural uncertainty of the models comprising the set.¹ The IPCC draws on MME studies of CMIP. CMIP co-

¹The IPCC distinguishes between MMEs and perturbed physics ensembles (PPEs). A PPE is a set of multiple simulations, each of which is a run of a single base model containing different parameter values. Aiming to capture the effects of uncertainties in initial conditions and other details of climatic processes, the distributed computing project Climateprediction.net exemplifies this approach, carrying out a large number of model runs on volunteers' home computers. Instead of relying on a single supercomputer to carry out all these simulations, volunteers run a version of one model on their home computers and return the processed data to the project centre, which compiles these runs, providing a means of investigating features of the climate system such its sensitivity to increasing atmospheric GHG content (Stainforth et al., 2005). PPEs provide an organized and systematic way of examining representational uncertainty, that is, how changes to a model-parameter

ordinates scientists to conduct ‘benchmark experiments’ that compare models with each other and observed climatic quantities, including recent instrumental measurements from historical records, collected with *in situ* and remote sensing devices, and proxy estimates of paleo-climate conditions, reconstructed from environmental traces of climatic phenomena. These comparisons provide a means of examining a range of model behaviours under controlled conditions, allowing scientists to isolate strengths and weakness of different modelling methodologies. The IPCC draws on ensemble studies in addressing a range of questions, including whether a set of models is adequate for the purposes of accounting for observed estimates of climatic phenomena, whether climatic change can be detected and, if so, attributed to certain causal factors, and how the future climate will evolve, given various forcing parameters (Parker, 2009, 2011).

4.2.1 Climate model ensemble evaluation

Calibrating and evaluating a climate model by waiting to see if it can accurately predict future climatic conditions is practically impossible: climate model projections are on the order of decades, centuries and millennia, so making a comparison of model output with states of affairs in the distant future is infeasible. Why, then, should one think of climate models as reliable representational tools that offer sources of knowledge about the world? The IPCC’s rationale is twofold: (1) these models incorporate well-known physical principles, such as the conservation of mass, energy, and momentum, and (2) they demonstrate skill according to some metric of evaluation. Although scientists use various metrics of model evaluation,

representing some feature of the climate affect the overall simulation.

uation, a common approach is to compare model outputs with observed estimates of the target system, such as changes in global mean surface temperature (GMST), sea-ice, ocean heat content and precipitation. Scientists evaluate individual models, model-components and model ensembles by comparing their outputs with observed estimates of the target system. Satisfying the condition of adequately reproducing observed quantities alone, however, is insufficient for guaranteeing the reliability of a model because models can be ‘tuned’ to adhere to observational constraints by adjusting various combinations of climate model parameters (Flato et al., 2013, pp. 749–750)

The most common way of evaluating an MME is “to calculate the arithmetic mean of the individual model results, referred to as an unweighted multi-model mean” (*ibid*, p. 755). A result that holds in some studies of various climatic parameters is that taking the mean value of multiple model outputs often agrees more favourably with observations than any individual model output (*ibid* p. 767). Figure 4.1 illustrates climate model ensemble evaluation of two sets of models, comparing observed and simulated time series of anomalies in annual GMST.

An alternative to a ‘one model, one vote’ ensemble approach is to assign unequal weights to individual model estimates in order to reflect differences in model skill, independence and internal variability. We will return to this point after describing two other applications of averaging results of multiple climate models.

4.2.2 Climate change detection and attribution

Detecting climate change is the process of demonstrating that the target system has changed in some defined statistical sense. Attributing a change to some causal fac-

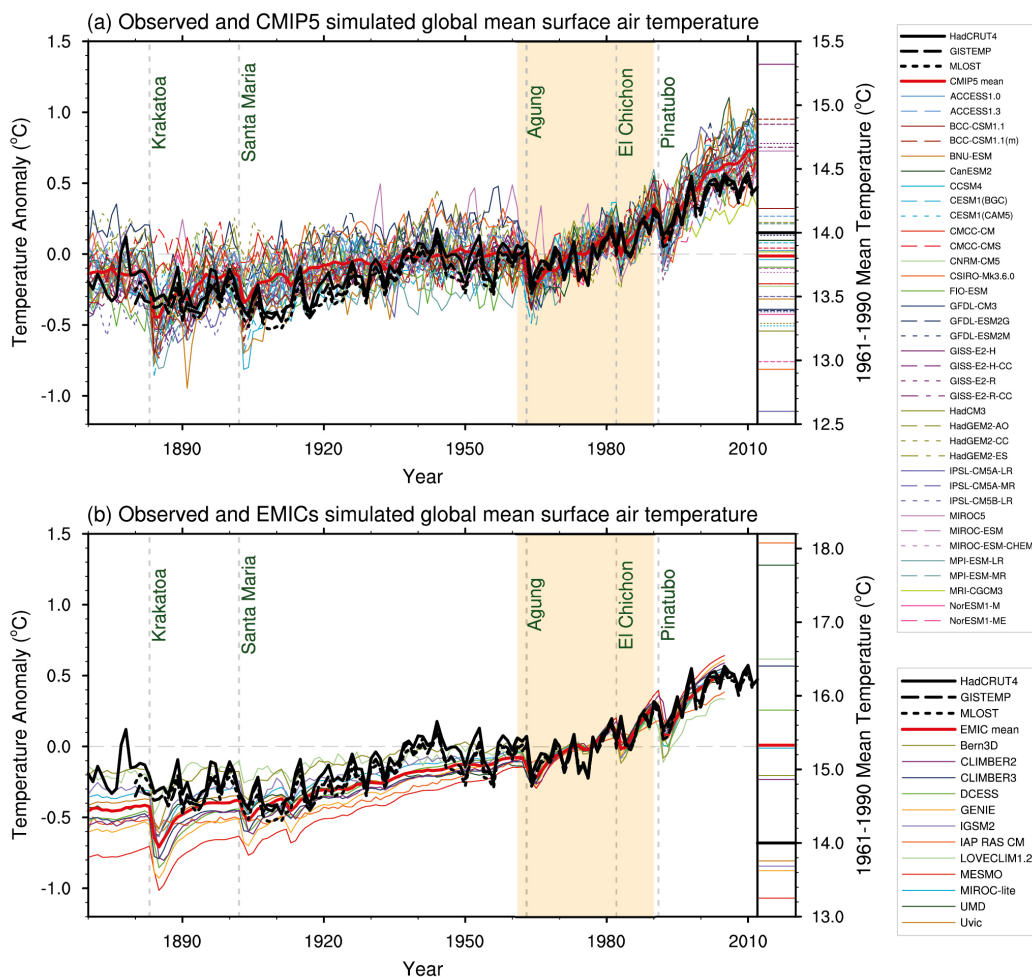


Figure 4.1: Climate model ensemble evaluations: anomalies relative to the 1961–1990 time-mean of each individual time series (yellow region). Vertical dashed grey lines mark major volcanic eruptions. Both graphs plot individual simulations (thin colour lines), multi-model mean (thick red line), and different observations (thick black lines). While coupled general circulation models (GCMs) from CMIP5 comprise the results in the top graph (a), the bottom graph (b) shows results from earth-system models of intermediate complexity (EMIC) (Flato et al., 2013, p. 768).

tors involves identifying the forcing mechanisms that led the system to evolve in a particular way (Hegerl et al., 2010). Model evaluation relates to D&A research because confidence in a model's ability to detect and attribute climate change depends on its ability to reproduce observed estimates of climatic quantities. Attribution studies typically involve comparing observed estimates with models that have been programmed to include or exclude different forcing mechanisms so as to demonstrate the efficacy of these mechanisms in producing observed changes to the system (Bindoff et al., 2013).

The IPCC assesses attribution studies that identify and quantify anthropogenic signals, finding the human 'fingerprint' in several climatic variables. Figure 4.2 illustrates the findings of a set of attribution studies indicating anthropogenic forcing at both regional and global scales, on land surface temperatures, precipitation, ocean heat content, and sea ice. Comparing observed estimates of changes in the atmosphere, ocean and cryosphere with outputs of model ensembles, these studies identify differences between models that include representations of natural causes alone and those that include representations of both natural and anthropogenic causes. The IPCC concludes that the "coherence of observed changes for the variables shown in Figure [4.2] with climate model simulations that include anthropogenic and natural forcing is remarkable... Both natural and anthropogenic forcings are required to understand fully the variability of the Earth system during the past 50 years" (*ibid* p. 927).

One of the main sources of Figure 4.2 is the publication of Gareth Jones et al. (2013), whose attribution study investigates the causes of near-surface temperature changes using observed estimates and CMIP3 and CMIP5 ensembles in an equal

weighting approach, which they acknowledge “may underestimate the uncertainty in the model spread” (p. 4003).² This point relates to the averaging approach in ensemble evaluation noted above, and one that modellers use to make climate projections, which we will outline below before addressing the epistemic question of when such averaging may provide an improved estimate of a given quantity.

4.2.3 Climate change projections

Unlike weather forecasts, which can be compared with observations on a weekly, daily and hourly bases, longer-term climate change projections aim to provide insights into how, given various contingencies such as the rate of anthropogenic greenhouse gas (GHG) emissions, climatic conditions will change in the decades and centuries ahead. The IPCC distinguishes projections from ‘predictions,’ explaining that the former term suggests a dependency on a range of alternative assumptions about forcing conditions, including the atmospheric concentrations of GHGs, while the latter term suggests a more definitive expectation about how a system will in fact evolve. The IPCC draws on CMIP climate projections, simulating a range of possible responses to different forcing scenarios. Representative Concentration Pathways (RCPs) describe climatic forcing resulting from variations in the emission rates and concentrations of GHGs and aerosols, and changes to landscapes (Moss et al., 2008, 2010). Each RCP represents one of many possible scenarios that could lead to particular radiative forcing characteristics. Figure 4.3

²Since some of the models in these ensembles contribute more than one simulation to the set, the authors calculate the “weighted” average of the simulations so as “to give equal weight to each model;” that is, they average the simulations of any model that contributes more than one simulation before averaging the output of the set of distinct models. So, while they assign unequal weights to simulations, they assign equal weight to each model in the ensemble.

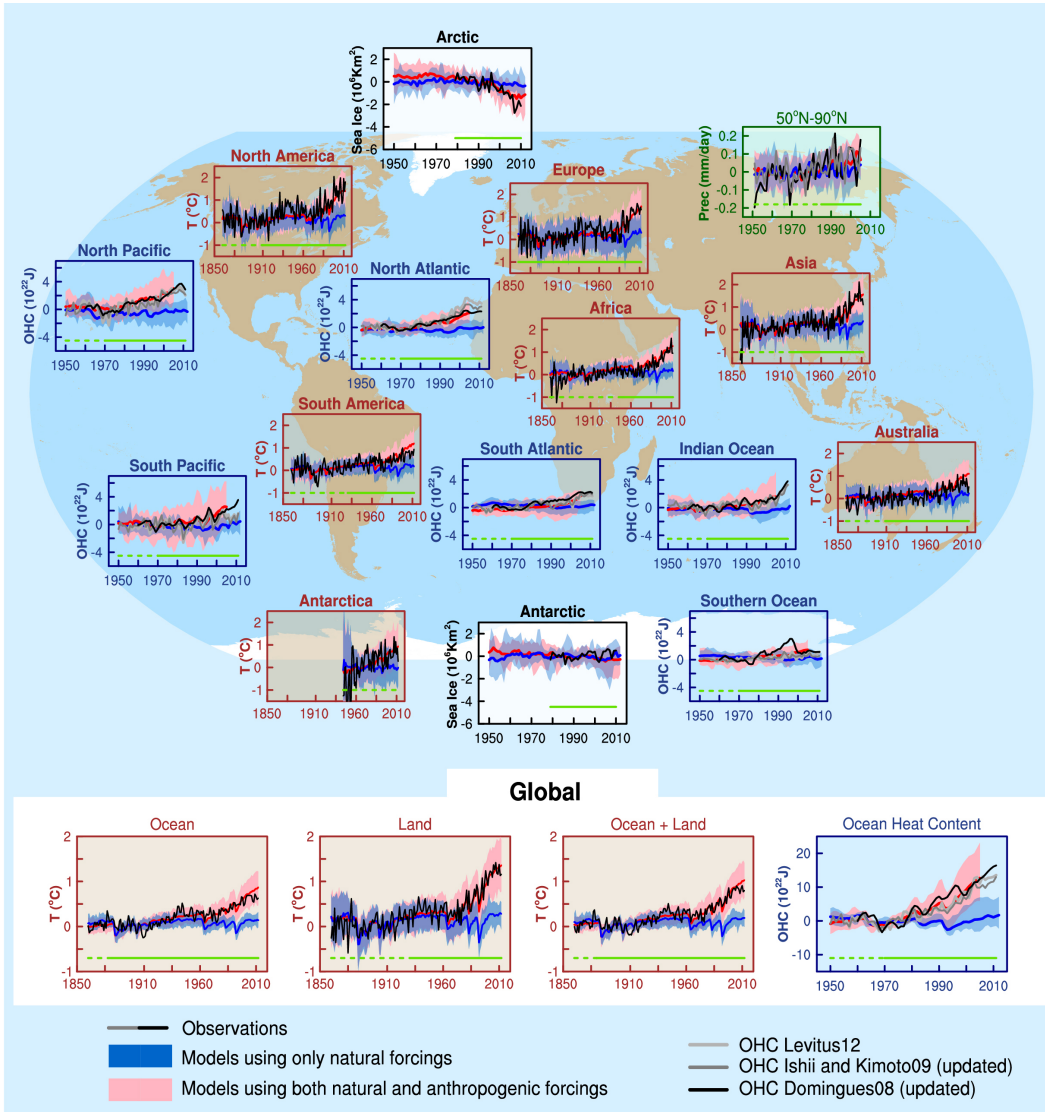


Figure 4.2: Climate change D&A studies using multi-model ensembles: For each graph, black and grey lines indicate observed estimates, light blue lines represent simulations that only include natural forcing, and pink lines represent simulations that combined natural and anthropogenic forcings. Dark blue and red lines show the ensemble means. The horizontal green lines mark the quality of the observed estimates across time (Bindoff et al., 2013, p. 930).

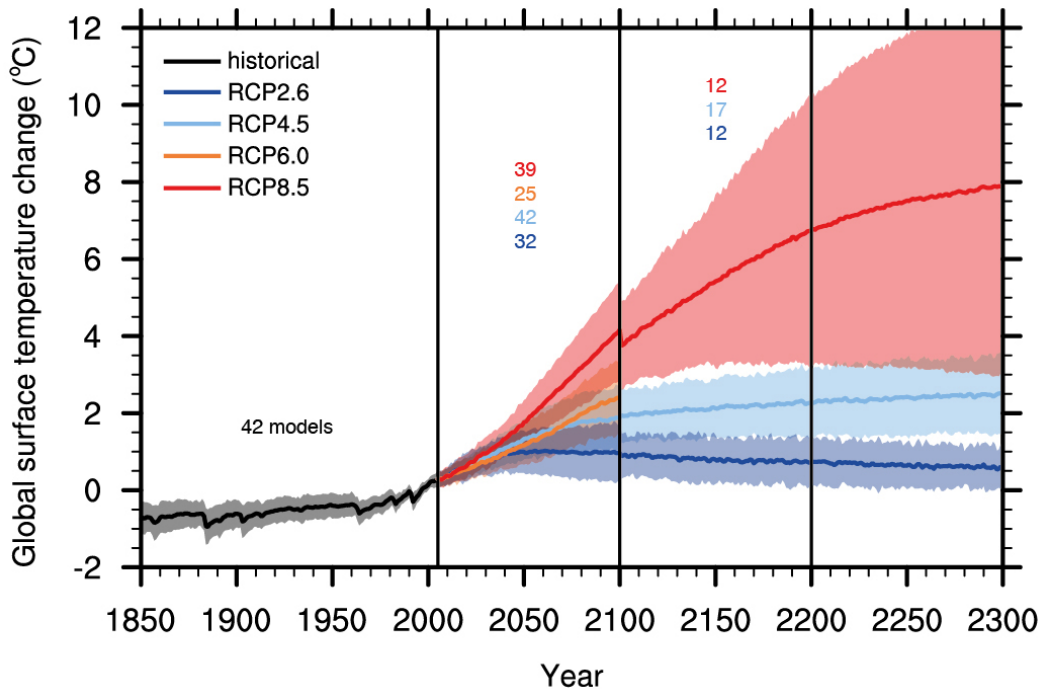


Figure 4.3: Time series of GMST anomalies (relative to 1986–2005) from CMIP5 concentration-driven experiments, showing RCP projections for the multi-model mean (solid lines) and the 5 to 95% range (± 1.64 standard deviation) across the distribution of individual models (shading). Numbers in the figure indicate the number of models contributing to the different time periods (Collins et al., 2013, p. 1054).

shows CMIP5 GMST projections for 4 RCPs.

Modellers often take MME projection averages to be the best estimates of future responses to various forcings. In its fourth assessment report, the IPCC states that, “[t]o the extent that simulation errors in different [coupled GCMs] are independent, the mean of the ensemble can be expected to outperform individual ensemble members, thus providing an improved best estimate forecast” (Meehl et al., 2007, p. 805).

The IPCC’s “Good Practice Guidance Paper on Assessing and Combining Multi

Model Climate Projections” outlines an epistemology of model ensembles, describing how quantifying uncertainties in model ensembles can assume that each model member is “sampled from a distribution centered around the truth” (Knutti et al., 2010a, p. 4).³ On this ‘truth-centered’ interpretation, one considers perfectly independent models in an ensemble as random draws from a distribution centered on reality such that uncertainties in predictions generated by the models tend towards zero as more samples are taken (Tebaldi and Knutti, 2007; Jun et al., 2008). Accordingly, model errors are equally likely to over estimate and under estimate quantities, allowing an ensemble mean to cancel errors as more models are included.

4.2.4 Model Independence

Models may be similar to each other because they describe the same features of system, but other similarities are due to common simplifications, numerical approximations, omissions of processes, parameterizations of processes, and other factors that may bias results. Discerning whether model agreement is due to their accuracy rather than than systematic biases complicates the interpretation of converging results. If models in an ensemble are in fact structurally independent and sampled from a truth-centered distribution, then different model errors should on average have near-zero pairwise correlations.

In practice, however, models share structural features that lead to correlated er-

³The IPCC also notes that one can assume that each member is “exchangeable” with the other members and the real system. This latter assumption considers observed estimates as “a single random draw from an imagined distribution of the space of all possible but equally credible climate models and all possible outcomes of Earth’s chaotic processes.” Under this interpretation, a given model drawn at random from a distribution of models is statistically indistinguishable from observed estimates, implying that uncertainties will “converge to a value related to the size of the distribution of all outcomes.”

rors (Tebaldi and Knutti, 2007; Jun et al., 2008; Pirtle et al., 2010; Knutti et al., 2010b; Masson and Knutti, 2011; Knutti et al., 2013). For example, investigating the extent of independence among 24 models in the CMIP3 ensemble, Christopher Pennell and Thomas Reichler (2011) apply two methods of quantifying the “effective” (as opposed to “actual” or stipulated) number of models in an ensemble, finding that the amount of unique information produced by each model implies that the effective number of models in the ensemble (7.5–9) is smaller than the number stipulated (24). “As more models are included in an ensemble,” they conclude, “the amount of new information diminishes in proportion” (Pennell and Reichler, 2011, p. 2358). Using an alternative method, Reto Knutti et al. (2013) arrive at a similar conclusion with respect to CMIP5 models, finding that ensemble members are “neither independent of each other nor independent of the earlier [CMIP3] generation.” Figure 4.4 illustrates the clustering of models according to similarities in their output with respect to temperature and precipitation projections. Models appearing in the same branch of the trees are close to each other according to a distance metric that considers the spatial field of monthly values in a control simulation without external forcing (see *ibid* Supporting Information for details).

Although a consensus on how to proceed has yet to be established, some modellers propose adapting and correcting ensemble methods to adjust for systematic biases, determining the effective number of models in ensembles, and weighting models according to relevant degrees of skill and independence.

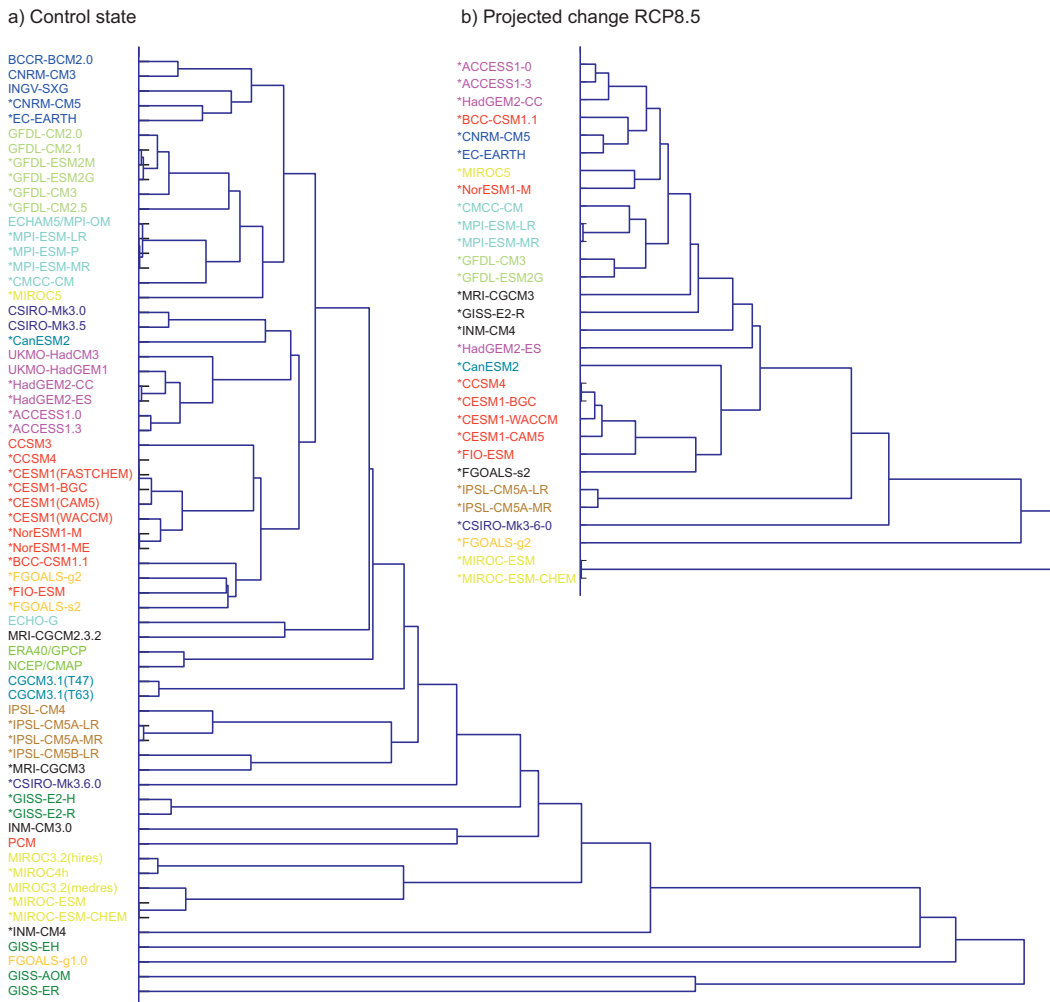


Figure 4.4: (a) The model “family tree” from CMIP3 and CMIP5 (marked with asterisks) control climate plus observations (ERA40/GPCP and NCEP/CMAP), shown as a dendrogram (a hierarchical clustering of the pairwise distance matrix for temperature and precipitation fields, see text). Some of the models with obvious similarities in code or produced by the same institution are marked with the same color. Models appearing in the same branch are close, and similarity is larger the more to the left the branches separate (for a detailed description of the method, see Masson and Knutti (2011)). (b) Same but based on the predicted change in temperature and precipitation fields for the end of the 21st century in the RCP8.5 scenario relative to the control (Knutti et al., 2013) ©American Geophysical Union 2013

4.3 Averaging Multiple Estimates

4.3.1 Averaging measurements

The rationale for averaging climate model ensemble output can be clarified by considering the following simpler, well-understood approach of averaging multiple measurements of a single quantity arising from different measuring methods.

We begin by denoting that each measurement or source of data $X_i = \text{True Value} + \text{Random Error}$:

$$X_i = x^* + e_i. \quad (4.1)$$

Note that, in general, X_i and X_j are not identically distributed because they arise from different measuring methods.

Then suppose that we take a weighted average of the measurement results:

$$\bar{X} = \sum_{i=1}^n w_i X_i, \quad \text{where } \sum_{i=1}^n w_i = 1. \quad (4.2)$$

This gives us an estimator of the quantity being measured. We choose, as a measure of ‘skill’ of the estimator, the variance:

$$V(\bar{X}) = \langle (\bar{X} - x^*)^2 \rangle, \quad (4.3)$$

noting that lower variance \Leftrightarrow better estimator.⁴ We can choose the weights w_i to minimize the variance of our estimator.

⁴The \langle and \rangle symbols enclose the expectation value.

4.3.2 Independent estimates with equal skill

Suppose that the errors associated with distinct sources are probabilistically independent. Then the variance of our estimator is given by

$$V(\bar{X}) = \sum_{i=1}^n w_i^2 V(X_i). \quad (4.4)$$

Now suppose that the variances of each source are the same: $V(X_i) = V$, for each $i = 1, \dots, n$. In this case the optimal weights are equal; and the variance of our estimator is

$$V(\bar{X}) = V/n. \quad (4.5)$$

Note that the variance of our estimator tends to zero as the number n of sources increases.

4.3.3 Independent estimates with unequal skill

In this case the optimal weights give more weight to sources with smaller variance. The optimal weights are those that have the weight accorded to a given source be inversely proportional to the variance of that source. That is,

$$w_i \propto 1/V(X_i). \quad (4.6)$$

To see this, first define the quantity κ by

$$1/\kappa^2 = \sum_{i=1}^n 1/V(X_i). \quad (4.7)$$

With some algebraic manipulation, we get

$$V(\bar{X}) = \kappa^2 + \sum_{i=1}^n \left(w_i - \kappa^2/V(X_i) \right)^2. \quad (4.8)$$

We minimize by taking

$$w_i = \kappa^2/V(X_i). \quad (4.9)$$

If, for example, we have two measurements of the width of a desk, one with a ruler, with an precision of ± 1 mm, and the other with a tape-measure, with precision ± 100 mm, then an unweighted mean of the two measurements could produce a less accurate measurement than taking a single estimate. One could just use the estimate from the more precise source, but excluding the other would be to ignore some of the information at our disposal. Instead of excluding relevant data, we can improve the performance of our estimate by taking a weighted mean, weighting the better source of data more strongly.

4.3.4 Estimates with dependencies

Suppose that, rather than being independent, the errors of our sources exhibit correlations between them. One might be tempted to think that the best way to proceed in such a case would be to pare down the ensemble until we have an independent set of sources. This approach, however, would be to discard potentially useful information; the optimal strategy is to form a weighted mean in which correlated sources are accorded less weight.

We define the *covariance matrix*:

$$C_{ij} = \langle (X_i - x^*)(X_j - x^*) \rangle. \quad (4.10)$$

Assuming this matrix is invertible, and letting C^{-1} be its inverse, one can show that optimal weights are given by

$$w_i \propto \sum_{j=1}^n C^{-1}_{ij}. \quad (4.11)$$

The variance of the estimator \bar{X} that we get,⁵ using these weights, is

$$V(\bar{X}) = 1 / \sum_{i,j=1}^n C^{-1}_{ij}. \quad (4.12)$$

If we suppose we have a number of sources, with the same variance V , we can compare (4.5) and (4.12), showing that these are as useful as an *effective number* (n_e) of independent sources, where

$$n_e = \sum_{i,j=1}^n C^{-1}_{ij} / V. \quad (4.13)$$

Example 1. Suppose that we have three data sources. Two of them, X_1 and X_2 , have the same variance σ^2 and have correlated errors, with correlation coefficient r , defined by

$$r = C_{12} / \sigma^2. \quad (4.14)$$

The third, X_3 , has errors independent of the other two, and variance τ^2 . The covari-

⁵cf. Dickinson (1973)

ance matrix is given by

$$C = \begin{pmatrix} \sigma^2 & r\sigma^2 & 0 \\ r\sigma^2 & \sigma^2 & 0 \\ 0 & 0 & \tau^2 \end{pmatrix} \quad (4.15)$$

Its inverse is

$$C^{-1} = \begin{pmatrix} 1/(1-r^2)\sigma^2 & -r/(1-r^2)\sigma^2 & 0 \\ -r/(1-r^2)\sigma^2 & 1/(1-r^2)\sigma^2 & 0 \\ 0 & 0 & 1/\tau^2 \end{pmatrix} \quad (4.16)$$

This yields optimal weights that satisfy

$$\frac{w_1^{opt}}{w_3^{opt}} = \frac{w_2^{opt}}{w_3^{opt}} = \frac{1}{(1+r)} \frac{\tau^2}{\sigma^2}. \quad (4.17)$$

That is, for positive r , the correlated estimates are accorded less weight than they would be if they were independent, and the weight accorded to each by the optimal weighting decreases as r increases.

If $\sigma^2 = \tau^2$, we have

$$n_e = \frac{3+r}{1+r}. \quad (4.18)$$

This is, as one would expect, equal to 3 when $r = 0$, and goes to 2 as r approaches 1; two perfectly correlated sources are no better than one.

Example 2. Suppose that we have n sources of data with the same variance σ^2 , and that each pair has the same mutual correlation r , so that

$$C_{ij} = \begin{cases} \sigma^2, & , i = j, \\ r\sigma^2, & , i \neq j. \end{cases} \quad (4.19)$$

Since none of the data sources is better than any of the others, the optimal weights are equal. We find that

$$V(\bar{X}) = (1 - r)V/n + rV. \quad (4.20)$$

Though this decreases with increasing n (so that it is always better to have more sources of data than fewer), if r is nonzero, there is a residual error rV that is not eliminated by increasing the number of data sources.

The effective number of sources is

$$n_e = \frac{n}{1 + (n - 1)r}. \quad (4.21)$$

Although this increases with increasing n , it does not increase without limit, and is never greater than $1/r$, no matter how large n is.

4.3.5 Comparing the simple case with CMIP ensemble methodology

The simple case of combining measurements suggests what is needed for effective use of multimodel ensembles. In both cases, methods for averaging results aim to track the truth about some target, and correlations of errors in estimates suggest methodological interdependencies which can influence the number of effective estimates in a set. In the simple case of combining measurements, the goal of the procedure is to identify sources with small variance. In the case of combining climate models, the goal is to improve the reliability of methods estimating climate parameter quantities. While in the simple case, sources of data will differ in vari-

ance, in the case of climate model ensembles, model skill varies from one ensemble member to another, depending on different metrics of evaluation. In the simple case, combining independent sources of equal variance involves taking an equally weighted mean of each source, suggesting an assumption similar to that of the one ‘model, one vote’ strategy, which is that an equally weighted mean can provide the best estimate even though the models in question demonstrate different degrees of skill and independence. The simple case of combining independent sources of unequal variance weights estimators with small variance more strongly, suggesting that averaging independent climate models of unequal skill should weight more skillful models more strongly than models that perform less well. Similarly, the simple case deals with non-independent sources by using covariance as a measure of dependence, adjusting weights accordingly, noting that positive correlations indicating the non-independence of sources detracts from the usefulness of combining measurements. In dealing with climate model ensembles, modellers use statistical techniques to measure systematic dependencies (Doblas-Reyes et al., 2005; Hagedorn et al., 2005). Table 4.3.5 summarizes these points of comparison. An important difference between our simple example and climate model ensembles is that climate models often estimate several different parameters. Some climate models are more skillful in addressing questions about some variables than others, so modellers assign different measures of success for different purposes (Flato et al., 2013, p. 766).

While the success of averaging ensembles is evident in many cases, a caveat should be noted. The benefit of aggregated performance is more significant when exploring multiple variables represented by many models. In estimating a single

Combining Measurements	Combining Models
Goal: To infer estimators with small variance	Goal: to improve reliability of estimates of quantities in climate studies
Sources of data differ in variance	Models may vary in skill
Independent sources, equal variance: Take unweighted mean	Independent models, equal skill: Take unweighted mean
Independent sources, unequal variance: Weight sources with small variance more strongly	Independent models, unequal skill: Weight models with more skill more strongly
Non-independent sources: Use covariance as measure of dependence, adjust weights accordingly. Positive correlation detracts from usefulness of combining measurements	Non-independent models: Use statistical method to measure systematic dependencies and weights accordingly. Positive error correlations detract from usefulness of combining models

Table 4.1: Comparing Cases of Combining Estimates

variable, one model can perform better than the equally-weighted multi-model average (Hagedorn et al., 2005). However, an unequally weighted average may provide further improvements. Our comparison with the simple example suggests that a best estimate from models should assign unequal weights to models in order to account for differences in their skill and relevant dependencies.

An alternative to assigning ‘one model, one vote’ is to assign model weights based on skill (e.g., model error in relation to observed estimates), independence

(error correlation), and internal variability (unpredictable noise), which has improved performance in some cases (Van Den Dool and Rukhovets, 1994; Robertson et al., 2004; Min and Hense, 2006; Weigel et al., 2008). Andreas Weigel et al. (2010) weight models according to these three factors, finding that, while on average equally weighted MME produce more accurate projections than do individual models, projection errors can be further reduced with optimum weights that are a function of these factors. However, if model weights inappropriately reflect model uncertainties, the weighted estimate may be less reliable than the unweighted one. “In fact, if the noise is of comparable or even larger magnitude than the model errors, then equal weighting essentially becomes the optimum way to construct a multimodel, at least if the models to be combined have similar internal variability” (Weigel et al., 2010, pp. 4189).

4.4 Conclusion

The aggregation of outputs of multiple numerical models has advantages for many fields of study, including those related to human health (Thomson et al., 2006), agriculture (Cantelaube and Terress, 2005), economics (Bauer et al., 2003), sports game predictions (Servan-Schreiber et al., 2004), and political polling (Silver, 2012). Scientific inquiries into questions about atmospheric phenomena have demonstrated that, according to various performance metrics, averaged estimates from MMEs can out-perform estimates from any individual model member—a conclusion exemplified by weather and seasonal forecasts (Krishnamurti et al., 1999), studies of mean climate (Gleckler et al., 2008; Reichler and Kim, 2008), climate change

attribution (Zhang et al., 2007), statistics of variability (Pierce et al., 2009), long-term simulations estimating the present-day climate (Lambert and Boer, 2001), and predictions of ENSO (Palmer et al., 2005).

This paper discussed cases in which one attains a best estimate of a quantity by averaging multiple estimates of various degrees of independence and skill. An assumption of the climate MME approaches described above is that individual model biases will partly cancel when averaged. Applying the ‘one model, one vote’ approach to ‘ensembles of opportunity’ takes the average estimate from a set of equally weighted, unequally skillful models with various degrees of systematic dependence and internal variability. Our comparison between climate modelling case studies and the simpler case of combining estimates highlights the importance of assigning unequal weights to ensemble members when estimates of model skill and independence can be adequately assessed. While correlated model errors can detract from the accuracy of MME estimates, efforts have been made to reduce the degree of these correlations. Some degree of model correlation is expected and acceptable because model weight can be adjusted to reflect model independence. Our analysis recommends that climate modellers combine outputs of models with different assumptions to reduce these correlations. Although improving model skill is an important part of MME methodology, quantifying degrees of relevant dependencies is also key for improving estimates of climate system variables.

Bibliography

- Bauer, A., Eisenbeis, R. A., Waggoner, D. F., and Zha, T. (2003). Forecast evaluation with cross-sectional data: The blue chip surveys. *Economic Review-Federal Reserve Bank of Atlanta*, 88(2):17–32.
- Bindoff, N., Stott, P., AchutaRao, K., Allen, M., Gillett, N., Gutzler, D., and Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I., Overland, J., Perlwitz, J., Sebbari, R., and Zhang, X. (2013). *Detection and attribution of climate change: From global to regional. In: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.
- Cantelaube, P. and Terress, J.-M. (2005). Seasonal weather forecasts for crop yield modelling in europe. *Tellus A*, 57(3):476–487.
- Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, M. G., Shongwe, C. T., Weaver, A., and Wehner, M. (2013). *Long-term climate change: Projections, commitments and irreversibility. In: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.
- Dickinson, J. P. (1973). Some statistical results in the combination of forecasts. *Operational Research Quarterly*, 24:253–260.
- Doblas-Reyes, F. J., Hagedorn, R., and Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting—II. calibration and combination. *Tellus A*, 57(3).

- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M. (2013). *Evaluation of climate models. In: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- Gleckler, P. J., Taylor, K. E., and Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research*, 113(D6):D06104.
- Hagedorn, R., Doblas-Reyes, F. J., and Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting—i. basic concept. *Tellus A*, 57(3):219–233.
- Hegerl, G. C., Hoegh-Guldberg, O., Casassa, G., Hoerling, M. P., Kovats, R. S., Parmesan, C., Pierce, D. W., and Stott, P. A. (2010). Good practice guidance paper on detection and attribution related to anthropogenic climate change. In *Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Detection and Attribution of Anthropogenic Climate Change*, page 8. Citeseer.
- Jones, G. S., Stott, P. A., and Christidis, N. (2013). Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, 118(10):4001–4024.
- Jun, M., Knutti, R., and Nychka, D. W. (2008). Spatial analysis to quantify numerical model bias and dependence. *Journal of the American Statistical Association*, 103(483):934–947.

- Knutti, R., Abramowitz, G., Collins, M., Eyring, V., Gleckler, P. J., Hewitson, B., and Mear, L. (2010a). Good practice guidance paper on assessing and combining multi model climate projections. in: Meeting report of the intergovernmental panel on climate change expert meeting on assessing and combining multi model climate projections. In Stocker, T.F., D. Q. G.-K. P. M. T. and Midgale, P., editors, *IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections*. IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A. (2010b). Challenges in combining projections from multiple climate models. *Journal of Climate*, 23:2739–2758.
- Knutti, R., Masson, D., and Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, 40:1184–1199.
- Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S. (1999). Improved weather and seasonal climate forecasts from multimodel superensemble. *Science*, 285(5433):1548–1550.
- Lambert, S. J. and Boer, G. J. (2001). CMIP1 evaluation and intercomparison of coupled climate models. *Climate Dynamics*, 17(2-3):83–106.
- Masson, D. and Knutti, R. (2011). Climate model genealogy. *Geophysical Research Letters*, 38:L08703.
- McGuffie, K. and Henderson-Sellers, A. (2014). *The Climate Modelling Primer*. John Wiley & Sons.

- Meehl, G., Stocker, T.F., Collins, W., Friedlingstein, P., Gaye, A., Gregory, A., Kitoh, J., Knutti, R., Murphy, J., Noda, A., Raper, S., Watterson, I., Weaver, A., and Zhao, Z.-C. (2007). *Global Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, volume 4. Cambridge University Press.
- Min, S.-K. and Hense, A. (2006). A bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. *Geophysical Research Letters*, 33(8):L08708.
- Moss, R. H., Babiker, M., Brinkman, S., Calvo, E., Carter, T., Edmonds, J. A., Elgizouli, I., Emori, S., Lin, E., and Hibbard, K. (2008). Towards new scenarios for analysis of emissions, climate change, impacts, and response strategies. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US).
- 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., and others (2010). The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282):747–756.
- Palmer, T. N., Doblus-Reyes, F. J., Hagedorn, R., and Weisheimer, A. (2005). Probabilistic prediction of climate using multi-model ensembles: from basics to applications. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463):1991.
- Parker, W. (2009). II—Confirmation and Adequacy-for-Purpose in climate modelling. In *Aristotelian Society Supplementary Volume*, volume 83, pages 233–

249.

- Parker, W. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science*, 78(4):579–600.
- Pennell, C. and Reichler, T. (2011). On the effective number of climate models. *Journal of Climate*, 24(9):2358–2367.
- Pierce, D. W., Barnett, T. P., Santer, B. D., and Gleckler, P. J. (2009). Selecting global climate models for regional climate change studies. *Proceedings of the National Academy of Sciences*, 106(21):8441–8446.
- Pirtle, Z., Meyer, R., and Hamilton, A. (2010). What does it mean when climate models agree? a case for assessing independence among general circulation models. *Environmental Science & Policy*, 13(5):351–361.
- Reichler, T. and Kim, J. (2008). How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, 89(3).
- Robertson, A. W., Kirshner, S., and Smyth, P. (2004). Downscaling of daily rainfall occurrence over northeast brazil using a hidden markov model. *Journal of climate*, 17(22):4407–4424.
- Servan-Schreiber, E., Wolfers, J., Pennock, D. M., and Galebach, B. (2004). Prediction markets: Does money matter? *Electronic Markets*, 14(3):243–251.
- Silver, N. (2012). *The signal and the noise: Why so many predictions fail-but some don't*. Penguin.
- Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., and Allen, M. R. (2005). Un-

- certainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, 433(7024):403–406.
- 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.
- Thomson, M. C., Doblas-Reyes, F. J., Mason, S. J., Hagedorn, R., Connor, S. J., Phindela, T., Morse, A. P., and Palmer, T. N. (2006). Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, 439(7076):576–579.
- Van Den Dool, H. M. and Rukhovets, L. (1994). On the weights for an ensemble-averaged 610-day forecast. *Weather and Forecasting*, 9(3):457–465.
- Weigel, A. P., Knutti, R., Liniger, M. A., and Appenzeller, C. (2010). Risks of model weighting in multimodel climate projections. *Journal of Climate*, 23(15):4175–4191.
- Weigel, A. P., Liniger, M. A., and Appenzeller, C. (2008). Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, 134(630):241–260.
- Zhang, X., Zwiers, F. W., Hegerl, G. C., Lambert, F. H., Gillett, N. P., Solomon, S., Stott, P. A., and Nozawa, T. (2007). Detection of human influence on twentieth-century precipitation trends. *Nature*, 448(7152):461–465.

Chapter 5

Conclusion

This dissertation addressed a set of related questions regarding the epistemology of science, including how consilient reasoning and robust evidential support account for the use of simulations in studies of large and complex target systems, how the notions of consilience and robustness relate to each other, and what the conditions are under which taking the mean of a set of multiple estimates provides epistemic advantages over using a single estimate. While philosophers have engaged with epistemology of science different contexts, this research provided a systematic examination of climate science methodology, explaining the distinctions and relations among different forms of evidence aggregation, and accounting for how scientific methods help to account for the high levels of confidence scientists have with respect to certain climate change hypotheses. The thesis argued that consilient reasoning and robust evidential support account for the strength and security of inferences that rely on imperfect computer modelling methods, that consilience is conducive to robustness, and that averaging the outputs of multiple models can provide im-

proved estimates of a target system when notions of relevant independence, skill and unequal weights are taken into account.

A salient theme among these chapters is a persistent emphasis on the importance of aggregating multiple lines of evidence in order to establish the plausibility of hypotheses. This theme encompassed general philosophical discussions about consilience, robustness and averaging estimates, and particular scientific cases of paleoclimate reconstructions, detection and attribution (D&A) studies, and climate change projections.

The methodology of individual case studies presented throughout the thesis also overlap in several ways. In order to increase their confidence in the results of D&A studies, for example, scientists evaluate models by comparing outputs with observed estimates of climatic phenomena, including patterns determined by paleoclimate reconstructions. Similarly, climate modelling studies inform certain paleo-climate reconstructions by calculating the potential amplitude of temperature changes over past centuries that are consistent with proxy data that serve as samples of climatic variables. The plausibility of climate model projections also depends on the skill with which such models can reproduce observed estimates of past and present climatic conditions. So, in several respects, paleo-climate reconstructions, climate model evaluations, D&A studies and climate projections draw on each other as sources of information. This exchange of information among different lines of inquiry expands the evidential resources available to each study, and demonstrate the complexity of research programs aimed at producing knowledge of the climate. In focusing on the importance of multiple lines of evidence, each chapter also highlighted the importance of identifying degrees of methodological

skill, evidential independence. The point of weighting evidence unequally is also relevant to each of the case studies, as in the principle component analysis of the paleo-climate reconstructions, and the model ensembles used in evaluations, D&A studies, and climate projections. These overlapping themes are depicted in Figure 5.1 by the dashed red line encompassing the common elements of the different chapters.

Each chapter exemplified how scientific inferences can be contrasted with ‘serial chain’ reasoning, which is the idea that an inference is only as strong as its weakest link. In this respect, the epistemology of this thesis provides a defense against arguments made in political debates about climate change that suggest errors in a given study can debunk a scientific consensus that has emerged from multiple independent lines of inquiry.

A related area of study for future research which has yet to receive sufficient attention is the intersection of epistemology and ethics in the context of ethical decision-making with respect to the science of climate change. Public debates about the hockey stick hypothesis (HSH) is one of many examples in which clarifying the epistemology of climate science is important for public policy. Paul Edwards (2010, p. 415) documents other cases in which problematic assumptions about climate modelling methodology have lead to “an ill-formed political debate based on a fundamental misrepresentation of the nature of climate knowledge.” Accordingly, clarifying the epistemology of climate science is important for understanding how scientific inferences relate to moral and political decision-making (Tuana, 2013).

The IPCC summarizes several observed effects of climate change, including changes to natural and human systems on all continents and across the oceans.

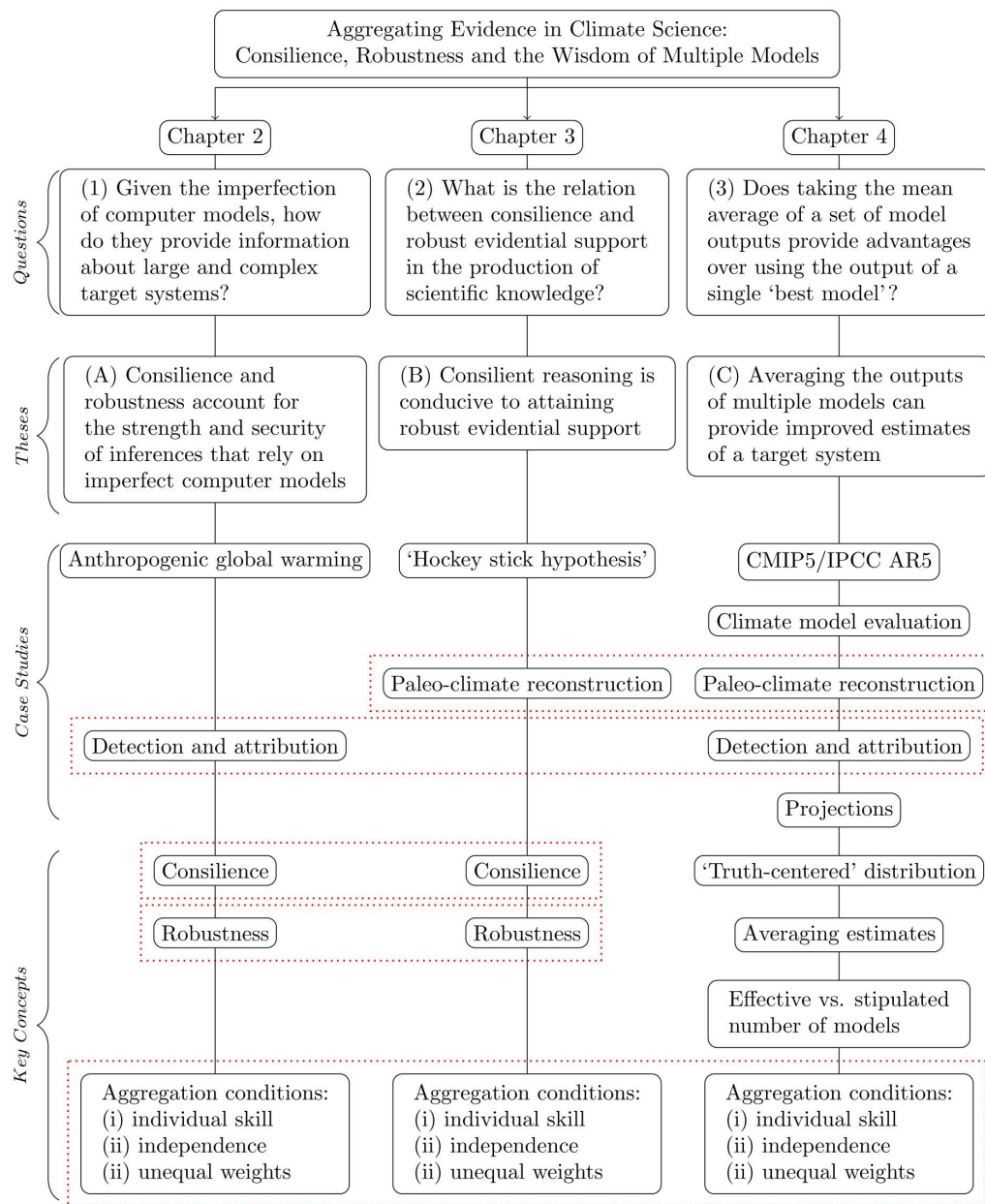


Figure 5.1: Dissertation overview. See text for description.

According to its latest assessment, “risks of harmful impacts on ecosystems and human systems increase with the rates and magnitudes of warming, ocean acidification, sea-level rise and other dimensions of climate change” (IPCC, 2014, p. 71). Further risks are indicated by observing ecosystem shifts and species extinctions, on land and in the oceans, during the past millions of years, caused by natural global climate change, including ones that occurred at rates lower than current anthropogenic climate change. Reasons for concern arise from projections regarding unique and threatened systems, including ecosystems and cultures already at risk from climate change, such as those associated with the Arctic sea and coral reefs. Additional warming also increases risks of extensive biodiversity loss, and related threats to ecosystem goods and services upon which people depend. Warming feedback can also lead to ‘tipping points’ that exacerbate risks of abrupt and irreversible changes to physical and ecological systems. Other risks that increase with further warming pertain to extreme events, such as heat waves, heavy precipitation and coastal flooding. Since such risks are distributed unevenly among different regions and people, disadvantaged people are the most vulnerable to the effects of climate change (IPCC, 2014).

The limitations in climate modelling projection methods make precautionary climate change mitigation and adaptation measures even more urgent because these epistemic limits affect our ability to plan for what might be needed to prepare for future environmental change. Although scientists face challenges in evaluating climate model reliability and they express high levels of uncertainty in their projections of regional scale climate change, adaptation and mitigation decisions need to be made in order to improve the well-being of people and ecosystems in the decades

ahead. Timely investments in climate change mitigation and adaptation efforts offer the advantage of reducing humanitarian, ecological and economic costs in the future. Estimating the costs and benefits of such measures is, however, contingent on the reliability of scientific knowledge. Accordingly, further inquiry drawing on climate science, epistemology of science, moral and political philosophy and environmental philosophy may improve our ability to navigate the waters of ethical decision-making under conditions of scientific uncertainty with respect to climate change.

Bibliography

Edwards, P. N. (2010). *A vast machine*. MIT Press.

IPCC (2014). *Climate Change Synthesis Report*, page 132. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Tuana, N. (2013). Embedding philosophers in the practices of science: bringing humanities to the sciences. *Synthese*, 190(11):1955–1973.

Curriculum Vitae

Martin A. Vezér

EDUCATION

PhD, Philosophy
Western University
London, Ontario, Canada.
2015

MSc, Philosophy of the Social Sciences
London School of Economics and Political Science
London, England.
2008

BA (Hons), Philosophy
York University,
Toronto, Ontario, Canada.
2007

RESEARCH

Area of Specialization
Epistemology of climate science.

Areas of Competence
General philosophy of natural and social sciences; moral and political philosophy; environmental philosophy.

EXPERIENCE

The Rotman Institute of Philosophy
2010–2015

- Steering Committee Student Representative, elected (Fall 2014–Fall 2015).
- Collaborator for Rotman Fall Conference *Knowledge and Climate Modelling: Philosophical, Historical and Scientific Perspectives* (23–26 October 2014).
- Chair of Epistemology of Scientific Modelling Reading Group (Fall 2013–present).
- Research Assistant to Professor Wayne Myrvold, studying epistemology of climate model ensembles (Summer 2014).
- Co-producer (with Rueven Brandt) of Livestream internet videos (Fall 2013–Fall 2014).
- Research Assistant to Professor Wayne Myrvold, studying epistemology of robustness analysis in climate science (Summer 2013).
- Chair of climate change conference brainstorming project (Summer 2013–Winter 2013).
- Chair of Internal Outreach for *Rotman Events Planning and Outreach (REPOC)*, developing outreach activities between the institute and other organizations in the university (Fall 2012–Fall 2013).

- Research Assistant to Professor Chris Smeenk, studying the epistemology of historical inquiry in natural science (Fall 2010).
- Research Assistant to Professor Gillian Barker, Professor Eric Desjardins and Professor Chris Smeenk, developing a reading group on the epistemology of historical inquiry, and organizing the *International Society for History, Philosophy, and Social Studies of Biology (ISHPSSB)* Off-Year Workshop, *Integrating Complexity: Environment and History* (Winter 2010–Summer 2010).

The Vienna International Summer University (VISU) Graduate Program on Climate Studies Summer 2013

- Certificate Program led by Professors James Fleming, Roman Frigg and Wendy Parker, hosted by Professors Friedrich Stadler, Karoly Kokai and at the *Institute Vienna Circle (IVC)*, the University of Vienna, and Sabine Koch at the *Center for Interdisciplinary Research (CIR)*.

Department of Philosophy, Western University

Winter 2013

- Research Assistant to Professor Devin Henry, studying precursors to evolutionary theory in antiquity.

Environmental Predictions in Canadian Cities (EPiCC)

Summer 2010–2012

- Outreach Manager and Research Assistant to Professor Jamie Voogt (UWO Geography), managing outreach program for *EPiCC* network of urban climatologists at UWO, University of British Columbia, McGill University, Environment Canada, King's College (University of London, England), and Meto-France. Duties include editing, accounting, progress and final reports; organizing meetings logistics; designing scientific tables and diagrams; designing and implementing off-line database; assisting in the development and management of network website. (<http://www.epicc.ca/>)

Department of Geography, Western University Ontario

Summer 2010–Fall 2010

- Graduate Research Assistant to Professor Gordon McBean, working on a series of papers comparing aspects of Canadian and American policies, public perspectives, and inter-governmental relations regarding climate change. Research funded in part by *ArcticNet*.

Department of Biology, Western University

Fall 2009

- Volunteer Lab Assistant to Professor Brent Sinclair and colleagues, assisting with laboratory experiments focused on insect responses to variations in micro-climate, in *The Biotron Experimental Climate Change Research Centre*.

Department of Philosophy, Logic and Scientific Methods at the London School of Economics and Political Science (England)

2007–2008

- Chair of LSE Postgraduate Philosophy Staff-Student Forums.
- Editor of LSE student philosophy journal, *Rerum Causae*.
- Sustainability Officer at *Shaftesbury Student Housing: Lillian Knowles House*.

Department of Environmental Studies, York University

Summer 2000 & Summer 2004

- Volunteer Field Assistant to Professor Howard Daugherty and colleagues, participating in student research field trips in Costa Rica, developing an environmentally and socially sustainable

shade-grown coffee program, *Las Nubes*.

WORKS IN PROGRESS

- [1] Myrvold, W. C. and Vezér, M. A. “On Potential Benefits and Risks of Weighted Averaging in Multimodel Ensembles”
- [2] Vezér, M. A. “Models and the Evidence of Anthropogenic Climate Change: Case Studies in Consilient and Robust Epistemology”
- [3] Vezér, M. A. and Myrvold, W. C. “Aggregating Evidence and Averaging Estimates: The Wisdom of Multiple Climate Model”
- [4] Vezér, M. A. “Consilience, Robustness and the ‘Hockey Stick Hypothesis”

PUBLICATIONS

- [1] Vezér, M. A. (2014). “Book Notice: Dale Jamieson, *Reason in a Dark Time* (Oxford, 2014).” *Metascience*. 10.1007/s11016-014-9940-5.

CONFERENCE PRESENTATIONS

- [1] Vezér, M. A. and Myrvold, W. C. (26 May 2014). “The Wisdom of Multiple Models: An Epistemology of Climate Model Ensemble Methods.” Talk presented at the *Canadian Society for the History and Philosophy of Science (CSHPS)* Meeting at Brock University, St. Cathrines, Ontario.
- [2] Vezér, M. A. and Myrvold, W. C. (4 December 2012). “Climate Model Ensembles: Rationale and Challenges.” Poster presented at the *American Geophysical Union* Fall Meeting, San Francisco, California.
- [3] Vezér, M. A. (17 July 2011). “Comparing Methodologies of Classical, Natural, Field and Computer Experiments Deployed in Climate Change Studies.” Talk presented at *The 14th Congress of Logic, Methodology and Philosophy of Science*, Nancy, France.
- [4] Vezér, M. A. (2 April 2011). “Towards A Taxonomy of ‘Experiments.’” Poster presented at *Epistemology of Models & Computer Simulations: National Conference*, Pittsburgh, PA.
- [5] Vezér, M. A. (15 December 2010). “Ontological and Epistemological Issues Regarding Climate Models and Computer Experiments.” Poster presented at the *American Geophysical Union* Fall Meeting, San Francisco, California.
- [6] Vezér, M. A. (23 October 2010). “On Turner’s Account of Gould, Cleland and the Role of Background Theories in Historical Inquiry.” Talk presented at *Field Science: the 26th Boulder Conference on the History and Philosophy of Science* at the University of Colorado at Boulder.
- [7] Vezér, M. A. “Paleoclimatology and Historical Methodology” (30 May 2010). Talk presented at the *Canadian Society for the History and Philosophy of Science (CSHPS)* Conference at Concordia University in Montreal, Quebec.
- [8] Vezér, M. A. (14–18 December 2009). “A Philosophical Defense of the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) Bayesian Methodology.” Poster presented at the *Philosophy Graduate Student Association* at UWO; and at the *American Geophysical Union* Fall Meeting in San Francisco, California.

INVITED TALKS

- [1] Vezér, M. A. (22 December 2011). “Climate Science: Epistemology, Policy and Outreach.” Talk presented to the *Toronto and Region Conservation Authority (TRCA)* Research and Development Group at TRCA Archetype House, Toronto, Ontario.

TEACHING

Department of Philosophy, Western University

Winter 2014, Lecturer: *Terrorism* (PH 2083G). New course design included topics in just war theory, ethics of government surveillance, ethics of drones, ‘eco-terrorism’ and other current geo-political issues, with particular focus on Canadian issues. (Teaching Honour Roll 2014.)

Winter 2011, Teaching Assistant to Professor Chris Smeenk: *Theory of Knowledge* (PH 2250).

Fall 2009–Winter 2010, Teaching Assistant to Professor Samantha Brennan and Professor John Nicholas: *Advanced Introduction to Philosophy* (1022E).

Winter 2009, Teaching Assistant to Dr. Steven Bland: *Big Ideas* (1130G).

OTHER ACADEMIC CONTRIBUTIONS

Chair, ‘Conceptualizing Organism-Environment Interactions’ (Oct. 7), ‘Climate Science and Sustainability’ (Oct. 9), ‘Panel : Perspectives on the Historical Sciences’ (Oct. 10), at the ISHPSSB Off-Year Workshop, *Integrating Complexity: Environment and History*, 7–10 October, 2010.

Chair, ‘Conceptions of Autonomy’ session at *XIV IAPh Symposium: Feminism, Science and Values*, at UWO, June 2010.

HONOURS & AWARDS

The University Student Council (USC) Teaching Honour Roll (2014)

Social Sciences and Humanities Research Council (SSHRC), Connection Project Grant. Role: Collaborator for Rotman Institute 2014 Annual Conference, *Knowledge and Climate Science: Philosophical, Historical and Scientific Perspectives*. Role: Collaborator.

Graduate Research Scholarship (GRS), Department of Philosophy, Western University, Fall 2011, Winter 2012, Fall 2012.

Arts and Humanities Graduate Thesis Research Award, awarded by the Dean of Arts and Humanities at UWO, Fall 2010, Fall 2011, and Fall 2012.

National Science Foundation, Student Travel Grant, awarded for the 14th Congress of Logic, Methodology and Philosophy of Science in Nancy, France, July 2011. Value:

American Geophysical Union Outreach Student Travel Grant, awarded for Fall Meeting in San Francisco, December 2009.

Social Sciences and Humanities Research Council (SSHRC), Canada Graduate Scholarship: Master’s Scholarship for Social Sciences and Humanities Research Council of Canada 2007.

York University’s Faculty of Arts Sessional Academic Achievement List for Outstanding Performance, 2005–2006, and 2006–2007.

GRADUATE COURSEWORK

PhD Philosophy of Science (UWO)

2013 Fall Term

Survey of Philosophy of Science (audited) with Professor Wayne Myrvold.

2012 Winter Term

Prospectus Course: Epistemology of Climate Science (enrolled) with Professor Wayne Myrvold.

2011 Winter Term

Epistemology of Computer Simulations (enrolled) with Professor Margaret Morrison (GRD-STUDY 998; University of Toronto).

2010 Fall Term

Empiricism in the Philosophy of Science (enrolled) with Professor Chris Smeenk (PHILOSOP 9209).

Moore's Ethics (enrolled) with Professor Anthony Skelton (PHILOSOP 9012).

2010 Winter Term

Environmental Philosophy (enrolled) with Professor Wayne Myrvold (PHILOSOP 9889).

Game Theory (enrolled) with Professor Bill Harper (PHILOSOP 9895).

Environmental Hazards (audited) with Professor Gordon McBean (GEOGRAPH 9332).

2009 Fall Term

Epistemology of Traditional and Computer-Simulation Experiments (reading course; enrolled) with Professor Brent Sinclair and the Sinclair Lab (PHILOSOP 9927).

Environmental Policy (enrolled) with Professor Gordon McBean (GEOGRAPH 9334).

Environmental Modelling (audited) with Professor Marco Van De Wiel (GEOGRAPH 9105).

2009 Summer Term

Epistemology of Historical Methodology (reading course; enrolled) with Professor Wayne Myrvold (PHILOSOP 9931)

2009 Winter Term

Philosophy of Perception Seminar (enrolled) with Professor John Nicholas (PHILOSOP 9205).

Research Seminar, Bayesian Probabilities in the Intergovernmental Panel on Climate Change (enrolled) with Professor Bill Harper (PHILOSOP 9900).

MSc Philosophy of the Social Sciences (LSE)

2007–2008

Philosophy of the Social Sciences (enrolled) with Professor Jason McKenzie Alexander (PH405).

Philosophy of Science (enrolled) with Professor Miklós Rédei and Professor Roman Frigg (PH400).

Dissertation Seminar (enrolled) with Professor Jason McKenzie Alexander (PH422).

Dissertation supervised by Miklós Rédei (PH499).

REFERENCES

References available upon request.