Roskilde Universitet
Fysik + Filosofi Videnskabsteori

Hanna Træland Rostøl
Character Count: 173135

The Epistemology of Climate Models - An Investigation of Robustness in Climate Science

Supervisor:
Martin Niss

Co-supervisor:
Patrick Blackburn

April 4, 2023

Summary

This thesis aims to understand the construction and application of climate models in the climate scientific practice. It aims to analyse some of the mechanisms at play when generating scientific knowledge in this discipline, by examining the role of “model agreement”, or the “robustness” of model results, to support hypotheses about the climate. In particular, it investigates whether Jonah Shupbach’s framework of “explanatory robustness analysis” can be applied to results detected by multi-model ensembles, thus establishing this robustness as an epistemologically sound method.

In order to do so, firstly a number of important features of the climate system are reviewed, before describing relevant types of climate models that aim to represent this system. Their implementation as computer simulations is also discussed, along with a review of different ways of discretizing and parameterizing aspects of the climate system reviewed.

Secondly, the use of these models in the climate scientific practice is reviewed. In particular, with focus on the Intergovernmental Panel on Climate Change (IPCC) uses multi-model ensembles in detection and attribution, and projection studies. It will become apparent that despite of some of the difficulties and uncertainties relating to the model outputs, the IPCC still uses these outputs to make hypotheses about the climate with high levels of confidence. This then leads to the question of what kind of practices allow for this high level of certainty, and the robustness of the model outputs is highlighted as one such method in the climate scientific practice.

The epistemological power of this robustness is then analysed in terms of “robustness analysis” from the philosophy of science. Robustness analysis systematically considers the relations between different means of detection and exactly when a commonly detected result among these should increase confidence in a hypothesis supported by these results. Some of the classical conceptions of robustness analysis based on independence between the means of detection are reviewed, but shown to be unfeasible. However, Jonah Shupbach’s revised notion of robustness analysis, namely “explanatory robustness analysis”, is emphasised as an especially promising framework in the context of climate models. Eric Winsberg’s attempt at placing climate models within this framework is analysed, and is consequently revised to fully accommodate its application to multi-model ensembles. Shupbach’s framework is also, step-by-step, applied to three example models from IPPC’s modelling ensembles, demonstrating that it can be successfully applied.

The main finding of this thesis is that hypotheses supported by multi-model ensembles satisfy Shupbach's framework of robustness analysis, and that therefore, the robustness of climate model detections should increase our confidence in these hypotheses. This gives a normative epistemological power to this common practice in climate science. However, it is also highlighted that this does equate to full confirmation, nor does it imply that the multi-model ensembles are constructed in an ideal way to maximally increase our confidence in a corresponding hypothesis. In light of this, the possibility of explanatory robustness analysis acting as a guiding principle to construct better multi-model ensembles is briefly discussed. Revisiting detection and attribution studies, it is also noted how explanatory reasoning is characteristic of climate science, and extending ERA to other types of evidence is considered.

Preface

With a bachelor's degree and (hopefully) almost finished master's degree in physics and philosophy, I have by now grown used to the various reactions to the somewhat unfamiliar combination of these two disciplines. However, something has changed in the very recent years, especially when telling people about this thesis. In the light of the complexities of climate science, the role of this science in politics, social life and ultimately, for the way we view the role of humans in the world, this synthesis of science and philosophy seems to be growing less alien to many, and the reactions less baffled.

However, what seems to be ignored is that philosophy and science evolved together, and only diverged in the recent past. The extreme politicisation and ethical and social inertia of climate scientific knowledge only serve to highlight the artificiality of this separation. Furthermore, it also highlights some of the pitfalls that could be partially attributable to this bifurcation; namely people blindly trusting the science but keeping it at a distance, and thus not acting on the scientific knowledge, or people not understanding its methods and therefore thinking they are in a position to challenge them, thus denying the science.

This thesis will not try to provide an answer to these complex problems. It will, however, aim to shed some light on the interesting philosophical discussions present in the climate scientific discipline, and to show that there can be valuable reflections in the philosophy of science that in return can contribute to the scientific practice. Nonetheless, combining these two disciplines in the same thesis comes at a cost. It will be beyond the scope of this thesis to delve fully into the details of the climate system and the models describing it (however, such a task is arguably outside the scope of any master's thesis). Similarly, it will be beyond the scope of this thesis to create a fully comprehensive picture of the currents in the philosophy of science that has led to the interesting discussion of robustness analysis, nor to entertain further discussions of the relationship between theory, models and experiment (although arguably, this is also outside the scope of any master's thesis).

Instead, I aim to present a philosophical argument contributing to the question of what the value of model robustness in climate science is, heavily informed and inspired by fundamental climate science, an understanding of climate models, and the climate scientific practice. In general, I believe any philosophy of science should pay careful attention to the scientific details of the practice in question. Conversely, I also believe there should be space for

the philosophy of science to inspire reflection and revision in any scientific discipline.

Moving in between disciplines is never an easy task. I therefore want to thank my physics (although with a strong footing in the philosophy of science) main-supervisor Martin, and my philosophy (although with a strong footing in mathematics and science) co-supervisor Patrick, for helping me in this endeavour. I also want to thank them both for consistent guidance and support, for spending way more time on me than what was allocated, and for providing invaluable perspectives during this process.

Contents

1 Introduction

In his speech to the Conference of Parties in 2022 (COP27), Secretary-General of the United Nations António Guterres named climate change the "defining issue of our age". Further describing the situation, he also said that "[w]e are on a highway to climate hell with our foot still on the accelerator".[1]

The past eight years are the warmest on record. Regional temperature records all over the world keep reaching new extreme peaks. The summer of 2022 was the second warmest on record for Europe, with its second warmest June ever recorded at about 1.6°C above average and its warmest October, with temperatures nearly 2°C above average. Not to mention that both polar regions saw episodes of record temperatures during 2022, also with six months seeing record or near-record low Antarctic Sea ice extents for the corresponding month. Furthermore, in 2022 we also saw large areas of Pakistan flooded, with major economic losses and human casualties, as well as other extreme-weather related events such as wildfires across Europe. Record breaking heatwaves were also observed in China, North and South America. (Copernicus, 2023) And, perhaps most importantly, these dramatic records and extreme weather events appear to be caused by human actions. The CO2 levels in the atmosphere are the highest in 2 million years, and this is attributable to human burning of fossil fuels. (ibid) In light of these facts, one might begin to understand the choice of Guterres's dramatic words.

However, to anyone paying attention to the news, the facts above will perhaps not sound so unfamiliar. Yet, the familiarity with these climate scientific facts and hypotheses does not necessarily equate to an understanding of what climate science is, or how these facts come into being. Climate science is certainly not one of the sciences we were taught in school, and that even the non-scientist might remember experimentally testing hypotheses in, such as physics. Still, physics is certainly a central component of climate science. Applied mathematics and chemistry are two other important components that might also invoke more familiarity than the composite and perhaps mysterious "climate science". However, these are by far not the only disciplines important in climate science. In fact, as philosopher of science Eric Winsberg points out, climate science is really an incredibly interdisciplinary science, drawing on fields such as climatology, meteorology, atmospheric physics, atmospheric chemistry, solar physics, applied math-

[1] Speech available at https://www.un.org/sg/en/content/sg/speeches/2022-11-07/secretary-generals-remarks-high-level-opening-of-cop27

ematics and mathematical modelling, only to mention a few of his listed examples. (Winsberg, 2018)[p. 3]

This incredibly rich science is not just critical to the study of how the living conditions for all life on Earth is responding to anthropogenic influence; it also hosts a number of interesting philosophical discussions. As Winsberg points out, it contains all the conceptual, methodological and epistemological issues that preoccupy modern philosophers of science.[2] (ibid) One can find characteristic philosophical discussions such as how theory, data, models and scientists are connected in climate science (Lloyd, 2012), decision theory in light of uncertainty (Bradley and Steele, 2015), and the role of values in climate science (Bender et al., 2022), to mention only a few.

This thesis focuses on perhaps the most archetypal question of all in the philosophy of science, namely what makes scientific knowledge reliable. Although an age-old question, it takes on a more modern form in the face of climate modelling, because it concerns how complex computer simulations can be used to build confidence in climate scientific knowledge. In particular, this thesis investigates how the common climate scientific practice of referring to "model-agreement" and results "robust" across multi-model ensembles can be placed within a framework of "robustness analysis" in the philosophy. Achieving this would explain why this agreement should make the scientific knowledge more reliable. Specifically, it will inverstigate whether Jonah Shupbach's (2016) framework of explanatory robustness analysis can be applied to multi-model ensembles used to support hypotheses about the climate. Such a unification would explain exactly *why* and *how* the model agreement should carry the epistemological weight it seems to do in the climate scientific practice.

Before attempting to answer such questions however, and to understand why the discussion of robustness analysis is at all relevant in climate science, it is necessary to have an understanding of some of the mechanisms of the climate system, the models we have for representing it, and how these are implemented as computer simulations. This will be the topic of chapter 2. The focus will be on simple Energy Balance Models, which are discussed for pedagogical reasons, and Global Circulation Models, which are crucial for

[2] Winsberg further argues that climate science being loaded with philosophical issues does not mean that there is a corresponding overwhelming literature on the philosophy of climate science. He points to two potential reasons for this; firstly, that philosophers of science tend to cluster around a very narrow collection of scientific subjects and climate science has not been one of those; secondly, that the tradition of philosophy of science in the English-speaking world has been to stay clear of issues of social concern and science with strong political and social consequences. (Winsberg, 2018)[p. 3-4]

understanding the models used to make projections in the climate scientific practice. We will see how the models are based on some simple physical processes, but that the complexity of the Earth system, intrinsic limitations when implementing the models on a computer, as well as further biogeochemical processes, various simplifications and abstractions are necessary. In particular, the discretization of the models, and so-called "parameterizations", i.e. processes that cannot be resolved in the grid-scale chosen, will be discussed.

In chapter 3, the ways in which these models are used to construct and test hypotheses about the climate is discussed. The focus is on the practice of the Intergovernmental Panel on Climate Change (IPPC), and its Assessment Reports focusing on the physical scientific basis of climate change. The IPPC's assessment reports are the culmination of the most up-to-date climate scientific research from research groups across the world, and therefore represent the state-of-the-art of the discipline. Two main research uses for climate models are identified and reviewed, namely detection and attribution studies, and projection studies. The IPPC's use of multi-model ensembles will become apparent from this. We will also see how the IPCC makes strong statements about the behaviour of the climate system based on the output of these multi-model ensembles, and the robustness between the output of these. This naturally leads to the question of exactly what gives the scientists such a high confidence in hypotheses supported by the model outputs, especially in light of the many inaccuracies and imperfections in the models that became apparent in chapter 2.

Chapter 4 will then turn to the discussion of robustness analysis in the philosophy of science. Robustness analysis will be highlighted as one method capable of explaining why the agreement of various means of detection should increase our confidence in a hypothesis supported by a commonly detected result among these means. Closely following the analysis of Shupbach (2016), we will see how traditional conceptions of robustness analysis rooted in the independence between the different means is insufficient. I will also extend this to an example of climate modelling, illustrating how these traditional accounts cannot be applied in this case. Shupbach's Explanatory Robustness Analysis will then be reviewed, as will Winsberg's application of it to climate science. However, Winsberg argues that explanatory robustness analysis based on models alone is not necessarily confirmatory in the case of multi-model ensemble. It is contended that Winsberg's argument is based on an incomplete application of Shupbach's framework to the example in question, thus requiring revision.

This revision is the topic of chapter 5. It is shown that by rewriting the two alternative hypotheses used by Winsberg, Shupbach's weaker condition for the fulfillment of explanatory robustness analysis can be applied, which in turn shows how the framework really can be applied to the multi-model ensemble in Winsberg's example. The details of three examples of climate models will also be reviewed to see if Shupbach's framework can be applied to these real-world models. Some of the limitations of Explanatory Robustness Analysis are also discussed, as well as its potential for use in the climate scientific practice. Lastly, it is also extended beyond models to also include experimental evidence, by again looking at the example of detection and attribution studies.

2 The Climate System, Climate Models and Climate Simulations

Our climate is complex. So are the mathematical models of our climate. It is therefore no easy task either to accurately describe the climate, nor to exhaustively represent the models of it. In this section, however, I will attempt to give an account of the components of the climate system and climate modelling that are relevant for the discussions of this thesis. I will begin by defining the climate and describe some of the essential processes that constitute it, before I will move on to the models that attempt to represent all these processes. I will then give an account of some different types of climate models, starting from the most simple ones, namely Energy Balance Models, gradually building up to the models that are used the most for climate scientific research, namely Global Circulation Models and Earth System Models.

2.1 The Climate System

Intuitively, we might know that the "weather" describes the specific atmospheric conditions at a specific time, which will in turn define whether we get rain or snow, sun or hail, if it is windy or not. The "climate" also encompasses some of these familiar variables, but in a different way. However, exactly when does weather turn into climate? If we talk about the average weather over a week, is that the climate? What about a summer, or a year?

The IPCC defines climate in the following way:

> Climate in a narrow sense is usually defined as the average weather, or more rigorously, as the statistical description in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years. The relevant quantities are most often surface variables such as temperature, precipitation and wind. Classically the period for averaging these variables is 30 years, as defined by the World Meteorological Organization. (Cubasch et al., 2013)[p. 126]

This definition illustrates two important features. Firstly, it describes the climate as statistical description over a time interval, in terms of both the mean and the variability. In fact, the variability is essential when defining the climate, because perhaps more than in what remains unchanged, we are interested in the climatic conditions that change over time. Secondly, as the definition states, we can operate on many time-scales when defining the climate, but the the standard in the climate scientific practice today is an average over 30 years, defined by the World Meterological Organization. (Arguez and Vose, 2011) This means that when discussing the "climate", we are discussing the variability trends of certain variables over a period of 30 years.

This answers the question of what time-scale we are operating with. However, what exactly is measured when studying the climate? As stated in the definition above, well known variables such as temperature, precipitation, and winds are certainly of importance, however, so are perhaps less-familiar variables such as surface radiation budget, properties of clouds, and ice sheet cover. Understanding what variables are important to monitor the climate and any climatic changes, require an understanding of the processes that constitute the climate. Let us therefore review some relevant components of the climate system.

Arguably the most defining feature of the climate system is that it is a dynamic system in transient balance. (McGuffie and Henderson-Sellers, 2005) Importantly, there is a balance between incoming energy from the sun, and the energy that is returned to space again. Any disturbance to this energy balance will therefore lead to a change in the energy contained in the Earth system. The Earth's radiation budget is a record of how much energy enters the Earth system, and how much energy goes back out, and a schematic of this is given in figure 1. If the energy balance is disturbed (and the radiation budget consequently changed) there is more energy contained within the climate system, and this extra energy needs to go somewhere.

There are four main ways energy in the climate system can be absorbed by its components. Firstly, it can go into the atmosphere, raising the air temperatures. Secondly, it can go into the land, increasing the land temperatures. Thirdly, it can go into the oceans, leading to more heat content in the oceans and increased sea temperatures. And lastly, it can go directly into the phase transition of snow and ice, melting glaciers and ice sheets. (Lee et al., 2021) This explains why we are so interested in monitoring the temperatures of the Earth system, namely because it is a good indicator of energy being added or taken away from the system due to an energy imbalance. Any factor that imposes a change on the planetary energy balance is called a climate forcing. (ibid) So what kind of climate forcings are there?

As we can see in figure 1, the main source of incoming energy to the climate system is radiation from the Sun. Therefore, it might not be surprising that the Earth's energy balance is very dependent on any changes to the amount of sunlight reaching the Earth. The most significant way in-

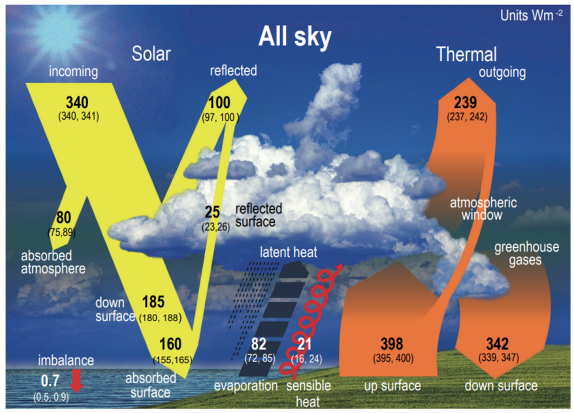

Figure 1: The radiation budget of the Earth's climate system. We can see how some components of the incoming light is directly reflected back to space, some are absorbed by the Earth and then re-released as infra-red waves, and correspondingly trapped in the climate system by greenhouse gases and clouds, or re-released into space. The figure is taken from (Lee et al., 2021).

solation to the Earth can change is if the orbit of the Earth around the Sun changes. There are three ways in which this can happen, namely changes in eccentricity (changes in the shape of the Earth's orbit from elliptical to more circular), changes in obliquity (the tilt of the Earth's axis of rotation), and changes in orbital precession (direction of the Earth's axis of rotation). The way these variables change are known as Milankovitch cycles, and have periodicities of 100,000/413,000 years, 41,000 years, and 19,000-23,000 years respectively. (Campsiano, 2012) Evidently, albeit significant, all these changes happen over periods too long to be of importance to our current society. Also remember from above that we are interested in climatic changes that happen over 30-year periods, and the Milankovitch variations would obviously go unnoticed over such a short period. Ergo, although Milankovich cycles are of massive importance when trying to understand past climate, these forcings are of diminishing importance when studying the climatic changes over the past few centuries.

Solar activity also undergo more frequent cyclical changes related to the sunspot cycle. This refers to the production of sunspots, i.e. dark areas, and faculae, i.e. bright areas that will accordingly change the emitted solar radiation, and consequently also the insolation to the Earth. (McGuffie and Henderson-Sellers, 2005) When looking at shorter-term climatic changes it is therefore relevant to consider the climate forcing due to the solar activity of that time. However, when looking at the observed climatic changes over the past centuries, solar activity cannot account for these changes. (Solanki and Krivova, 2003) Evidently, there are other climate forcings present in the system.

Perhaps the most well-known mechanism that regulates the energy contained within the climate system is the greenhouse effect. Giving a comprehensive outline of the greenhouse effect is beyond the present scope. However, let us summarize it as the mechanism by which sunlight, which is in the high-energy spectrum, passes through the atmosphere, and is then absorbed by the Earth and re-emitted at a much lower wavelength. The radiation at this wavelength, which lies in the infrared spectrum, does not pass back into space as easily. In particular, so-called greenhouse gases absorb certain wavelengths of the emitted radiation, trapping the heat inside the system instead of letting it pass back to space. (Le Treut et al., 2007) Greenhouse gases therefore act as climate forcings, where an increase in concentration will lead to an increase in the energy of the climate system. The most important forcings altering the greenhouse effect are human emissions of greenhouse gases, including carbon dioxide, methane, and nitrous oxide. However, volcanic eruptions can also release large quantities of greenhouse gases. (Lee et al., 2021)

Another type of climate forcing is troposheric aerosols. These are particles in the troposheric layer of the atmosphere, affecting the amount of incoming light. Typically these aerosols reflect radiation back into space which leads to a cooling effect, but there are also particles, such as soot, which do the opposite. Troposheric aerosols also affect how and how many clouds are formed. The effect of clouds in the climate system is two-fold, firstly the amount and type of cloud affect the incoming radiation because they reflect radiation back into space; secondly they also trap heat within the Earth system, affecting the amount of outgoing radiation. (McGuffie and Henderson-Sellers, 2005) Increased concentrations of aerosols can be a result of the combustion of fossil fuels, biomass and other sources of pollution, but are also released in large quantities in volcanic eruptions.

Other mechanisms that can change the energy balance include changes to the ozone layer, which again affects the total solar radiation entering the Earth system, and changes in surface reflectivity described by the albedo effect. Albedo describes ability of surfaces to reflect sunlight. Light surfaces reflect more light (high albedo) than dark-surfaces, which absorb the light. The albedo of the surface of the Earth will be affected by changes to the land-surface, for example through desertification, re- and deforestation and urbanization. It can also be affected by dark particles, for example those coming from the combustion of fossil fuels, sticking to surfaces that would otherwise have a high albedo, such as ice and snow. There also exists further climate forcings, such as aviation trails and cosmic rays; a more comprehensive overview can be found in (Lee et al., 2021).

It is possible to classify climate forcings into two main categories, namely "natural" and "anthropogenic". It is not difficult to understand that solar forcings are natural, as they are completely external to the Earth system and consequently also outside human influence.[3] Similarly, although internal to the climate system, volcanic eruptions are outside human influence and thus a natural forcing. Further emissions of greenhouse gases, aerosols from pollution, ozone depletion as a result of chlorofluorocarbons and land-surface changes on the other hand, all belong to the "anthropogenic" category.

So far, some climate forcings with potential of disturbing the energy balance and thus the climate system have been described. What remains then, is to describe how this system responds to the changes. As mentioned above, there are a few different ways in which additional energy can be absorbed in the system, namely by the atmosphere, land, oceans or cryosphere. However, after being absorbed by one of these, the temperature increase will lead to further reactions in the climate system. This is because there are processes internal to the climate system that will act to amplify or dampen the effect of any energy imbalance, namely feedback effects. Positive feedback effects amplify the effect of a perturbation to the system, meaning that a temperature increase for example will lead to a further temperature increase, without any more added energy. A negative feedback on the other hand, is one that opposes the effect of a perturbation to the system. In the case of a temperature increase, this would mean that the negative feedback effect would lead to a temperature decrease. (McGuffie and Henderson-Sellers, 2005)

There are a multitude of feedback effects of importance in the climate system. The most universal feedback mechanism is the Planck response, which is an effect due to the Planck blackbody radiation law, describing how the emitted radiation from a body is temperature dependent. Consequently, with a higher temperature, the Earth will also re-emit more radiation, and thus, this feedback is a strongly negative feedback which plays a crucial role

[3]The external forcings are at least outside human reach at the present moment. However, proposals of space-based solar geoengineering would break this distinction down. This includes proposals of so-called "sun-shields", which are constructions located in space between the Earth and the Sun and would block some of the incoming light from ever reaching the Earth. (Baum et al., 2022)

in stabilizing the climate. However, despite this strong negative feedback, the total effect due to feedback mechanisms in the climate system is believed to be positive. (Lee et al., 2021) This is because of the presence of a multitude of other, positive feedback mechanisms.

Positive feedback mechanisms in the Earth system include the ice-albedo feedback, which describes the contribution of melting of ice leading to a decrease in the reflectivity of the surface of the Earth because the highly reflective, white ice covers are replaced with darker substances such as land or water. The decrease in the albedo effect in turn leads to more radiation being absorbed, and thus a further temperature increase. However, although strong, this is a highly localised feedback mechanism, and therefore does not contribute accordingly to the totality of feedback effects. (Lee et al., 2021) Another important positive feedback mechanisms is water vapour feedback. This entails the enhanced greenhouse effect resulting from more water evaporating due to increased temperatures, which in turn works to increase the temperature further. Other feedback mechanisms include lapse rate feedback, clouds, and biogeophysical and biogeochemical processes. (ibid)

An important concept arising from this is that of climate "sensitivity". Generally, this is a measure of the response of the climate system when exposed to a change. Effectively, the climate sensitivity measures the combined effect of all the feedback and regulation mechanisms of the climate system, and is therefore an extremely important emergent feature of the climate system, that is essential to understand how the climate will change in the future. (McGuffie and Henderson-Sellers, 2005) However, different components of the climate system respond to changes on different time-scales. Water for example, because of its high heat capacity and high density compared air, has a much slower thermal inertia, meaning that the oceans respond to changes much slower than the atmosphere. There are therefore different ways of measuring this climate sensitivity, depending on how much time is allowed for the system to adjust. Equilibrium Climate Sensitivity, Transient Climate Response and Effective Climate Sensitivity are three methods to measure the sensitivity of the climate operated with by the IPCC. (Flato et al., 2013)

In this section, we have seen how the climate is described by a certain set of climatic variables over a defined time period. It was then demonstrated how it is the energy balance of climate system and any disturbances to it that ultimately defines the climate on Earth. This energy drives all the processes that constitute both the climate and the weather. The dynamics of the climate system, including winds and ocean currents, as well as precipitation patterns, are all dependent on the energy contained in the system. There are certain climate forcings, usually categorised as either natural or anthropogenic, that have the potential of changing the energy balance. The climate system will then respond accordingly.

This is an incredibly crude description of the climate system. However, for our purposes, this will be sufficient. Let us therefore move on to the models that aim to represent this climate system.

2.2 Climate Models

Climate modelling is the effort to represent the many processes that produce climate. Both physical, chemical and biological processes constitute the climate, and a mathematical model of the climate consists of equations expressing these different mechanisms. So climate models describe these processes mathematically, and implement the mathematical equations numerically on a computer, simulating the behaviour of the climate system. This can be done in varying degrees of complexity and detail, and with different components being focused on, or included at all. In this section we will look at some of the different choices that can be made when modelling the climate.

The discipline of climate modelling is relatively new, and although earlier analogue models exist, the first global circulation models appeared in the 1950s. (McGuffie and Henderson-Sellers, 2005) However, since these first attempts at modelling the climate system, there have been great advances in this field. The advances are both a result of better understanding of the physical processes that constitute the behaviour of the climate system, as well as a great increase in available computing power and simulation techniques. It is today the general consensus that these modelling efforts provide the most effective means for answering questions that require future climate and implications of future changes to the climate. (ibid)

However, as mentioned a few times already, the climate is an incredibly complex system, and it is not possible to represent it perfectly.[4] Although great advances has been made the last decades, we simply still do not have either the complete understanding of the processes in our climate, nor the computational power to do this, and it is questionable whether we ever will. This means that the models consists of major simplifications. First

[4]An interesting question is whether it is ever possible to represent any physical system perfectly by a mathematical model? However, what I mean here when saying that it is not possible to represent it perfectly is the fact that there are simplifications, idealisations and abstractions involved in models of the climate.

of all, it is necessary to make simplifications by idealising or approximating certain phenomena, known as parameterizations. These parameterizations will be revisited later in section 2.3.2, however it is important to note that it is possible to implement climate models based on the same fundamental equations very differently because of this.

In addition to making simplifications to the processes involved in climate models, it is also necessary to make another type of simplification: namely the model resolution, both in time and space. It seems evident that a finer spatial resolution would also give us more accurate model output. However, computational time and data availability put limitations to the spatial resolution of our models. In terms of the temporal resolution, there are also similar considerations for the balance between computational cost and accuracy. Most computational procedures require 'timestepping'. This means that the processes are allowed to act for a certain length of time before new conditions are calculated, then the process is iterated until the required length of time has been achieved. Ideally, we would want the temporal resolution to be able to capture changes relevant to all the processes we are interested in. However, to do this, we would need to have timesteps of a few seconds, which when we want to simulate the climate system for more than a hundred years, obviously is extremely computationally expensive. Instead, modern climate models usually have timesteps of around 30 minutes. (McGuffie and Henderson-Sellers, 2005)

So, it is evident that there are numerous considerations that need to be made when building a climate model. In the first part of this section we will look at some of the components of a climate model, and what different types of climate models look like. Specifically, the focus will be on energy balance models, which for pedagogical reasons will be treated in quite some detail, and Global Circulation Models (GCMs), which are the basis for the models used in current climate scientific research. The second part of this section will consider climate simulations, i.e. the output of climate models when run on a computer. Once again, the focus is on GCMs. This will allow us to understand the considerations required in translating the model into a computer implementation.

2.2.1 Types of Climate Models

When constructing a climate model, we can identify a few central components that need to be accounted for. Following the framework of McGuffie and Henderson-Sellers (2005), those are

1. Radiation - this entails the input and absorption of solar radiation in

the atmosphere and the surface of the earth, and the corresponding emission of infrared radiation;

2. Dynamics - this implies the circulation of energy around the globe, both horizontally in terms of wind and ocean currents, and vertically in terms of small-scale turbulence, convection, and deep-water formation;

3. Surface processes - the effects of sea and land ice, snow, vegetation and the resultant change in albedo, emissivity and surface-atmosphere energy and moisture interchanges are all a part of these processes.

4. Chemistry - this specifies the chemical composition of the atmosphere and the interactions with other components, such as the carbon cycle (i.e. the exchange of carbon between ocean, land and atmosphere).

Recalling the importance of the energy contained in the climate system discussed in section 2.1, it is easy to understand the importance of the first component. The difference between the incoming and outgoing radiation will define the amount of energy in the climate system, which is the driving force of all the other processes in the climate. Energy balance models, that will be discussed shortly, is a type of model focusing solely on this aspect of the climate. The remaining three components respectively deal with what happens to the energy inside the climate system, surface processes that affect the amount of energy absorbed, and chemical processes that affect the concentration of radiatively active species, i.e. greenhouse gases and aerosols. The different approaches to treating these four components and the interplay between them, as well as choices for spatial and temporal resolution as mentioned above, is what separates different types of climate models from each other. Some commonly distinguished types of climate models are:

1. Energy balance models (EBMs) of zero or one dimension. These models predict the surface (or strictly, the sea-level) temperature as a function of the energy balance of the Earth. This means it does not take large-scale wind and atmospheric circulation systems, ocean currents, convective features, or other essential components of the climate system into account. In the zero-dimensional case, the Earth is simply treated as a single, mathematical point in space. It then balances the incoming and outgoing radiation. Simplified relationships are used to calculate the terms contributing to the energy balance in each latitude zone in the one-dimensional case.

2. Dimensionally-constrained models that represent either two horizontal dimensions or the vertical plus one horizontal dimension. The general circulation is assumed to be composed mainly of cellular flow between latitudes, which is defined using a combination of empirical and theoretical formulations. Models that belong to this category include the early 'statistical dynamical (SD) models, and Earth Models with Intermediate Complexity (EMICs).

3. General Circulation Models (GCMs). These can either include the atmosphere (including land surfaces) only, with prescribed sea surface temperatures and sea ice. Simulations with such atmosphere only models are used intensively to validate climate models against observations - often based on remote sensing. Or, they can model the oceans with sea ice, and prescribe the atmospheric conditions (i.e. wind, precipitation, downward radiation). Lastly, they can combine the atmospheric and land components and the sea and ice components into a fully coupled GCM. These models do not however, include the carbon cycle or advanced online atmospheric aerosol chemistry. Several of the models the IPCC operates with are of this type, because of the further computational demands to include the aforementioned components.

4. Earth system models (ESM). These can either be earth system models of intermediate complexity (known as EMICs), which include important Earth system components such as interactive vegetation, ice sheets, oceanic and terrestial carbon cycles, and so on. However, at the expense of these factors, the dynamics of the atmosphere and oceans are simplified to allow for faster computations. This means that the atmospheric behaviour from timestep to timestep (i.e. the weather) is not modelled, and instead, the climatic impact of weather is parameterized or simplified. ESMs can also be fully coupled, dynamic Earth System Models which in principle include all components important to climate. These share some common features with GCMs as they are based on the same atmospheric and ocean components, difference being that ESMs include more components. However, at the present moment, they usually do not include features such as interactive ice sheet model components or long-term geological processes such as weathering, as it is impossible with present computer technology to perform simulations that cover sufficiently long periods of time.

It is straightforward to see that these different types of models make different choices in terms of exactly which of the components mentioned above (radiation, dynamics, surface components, chemistry), and in how much detail they are to be included. Let us now review EBMs and GCMs in more detail, before briefly discussing ESMs, which are the most common type of model used to make projections in the climate scientific practice.

2.2.2 Energy Balance Climate Models

Despite their naive simplicity seen in light of the complex coupled circulation models, energy balance models have been and still are instrumental in increasing our understanding of the climate system. Their main advantage is that they are easy to program and computationally light to run. This makes them suitable for developing parameterizations for more complicated models, as simulations can be made quickly so the statistics of the solution fields are readily available. The computational efficiency also mean that they have applications in paleoclimate studies, where the climate for very long time periods is modelled. (North and Stevens, 2006)

As mentioned above, zero-dimensional EBMs simply treat the Earth as a single, mathematical point in space, in which the incoming, short-wave radiation from the sun and outgoing long-wave radiation from the Earth are balanced. Let us now look to how such a model can be derived.

We start by assuming that the amount of shortwave radiation absorbed by the Earth simply is the expression

$$(1 - \alpha) \cdot S/4 \tag{1}$$

where S is the solar constant, i.e. the flux density measuring the total solar irradiance per unit area, measured at the distance from the Sun to the Earth[5]. We divide the solar constant by four as this is the ratio of the entire surface of the Earth ($4\pi r^2$) and the cross sectional area of the Earth (πr^2), which is the actual surface area that 'sees' the Earth at any given time. Recall from section 2.1 that the amount of light being reflected is the planetary albedo, which is represented α, and $(1 - \alpha)$-term correspondingly represents the amount of radiation that is absorbed. The expression above then represents the total amount of radiation absorbed by the surface of the Earth. (McGuffie and Henderson-Sellers, 2005)[pp. 82-84]

As mentioned before, an energy balance model balances the incoming and outgoing radiation. The expression above gives us the incoming radiation,

[5] Strictly speaking, it is measured on a surface perpendicular to the rays, one astronomical unit (au) from the Sun (roughly the distance from the Sun to the Earth). However, for most purposes this difference is insignificant.

so the We now have an expression for the incoming radiation, and therefore the next step is to derive an expression for the outgoing radiation. This can be done by treating the Earth as a black body, i.e. a body that absorbs all radiation that is incident upon it. The Stefan-Boltzmann law then gives the amount of radiation produced by a blackbody per unit surface area per unit time, j, as

$$j = \sigma \cdot {T_S}^4 \tag{2}$$

where σ is the constant of proportionality, known as the Stefan-Boltzmann constant, and T is the temperature of the blackbody. If the body is not a perfect blackbody, this relation still holds if the emissivity of an object, ϵ, is also inserted into the equation. Since ϵ is always between 0 and 1, where an emissivity of 1 means a perfect blackbody, this effectively reduces the radiation produced by a so-called 'grey-body'. However, the emissivity of the Earth is in fact very close to 1, and for most purposes Earth is simply assumed to be a perfect black body.

Now, let's approximate this T_S as the average temperature over the entire planet. Seeing that this layer is 70% ocean water, the thermal properties of the ocean are obviously very important. Now, we treat the ocean as a so-called 'mixed-layer', meaning that there are no ocean currents, no temperature variations or any other differences; the ocean is simply a stagnant layer. This allows us to approximate the thermodynamic effect of this mixed layer ocean in terms on an effective heat capacity of the entire earth system, and this heat capacity can be estimated to be $C = 2.08 \cdot 10^8 JK^{-1}m^{-2}$.

We are then in a position to write an expression for the energy balance. The change in the internal energy per unit area per time can be written as an expression of the heat capacity of the Earth system and the rate of change of surface temperature:

$$\Delta U = C\frac{dT_S}{dt} \tag{3}$$

This must in turn balance the rate of net heating. The rate of net heating is going to be the difference between the incoming radiation from the sun, and the outgoing radiation from the surface. From this we get the following expression of energy balance:

$$C\frac{dT_S}{dt} = \frac{(1-\alpha)\cdot S}{4} - \sigma \cdot {T_S}^4 \tag{4}$$

From this we can easily recognise the fact that if the incoming radiation is more than the outgoing radiation, that will lead to a rise in surface temperature of the Earth system, just like it was established in 2.1. That means that

T_S is increasing, which in turn will increase the outgoing radiation, eventually bringing the two terms on the right-hand side into balance. When these terms are balanced, there is no longer any change in surface temperature, meaning the left-hand side is zero. This means the two terms on the right-hand side are completely balanced, i.e.

$$\sigma \cdot {T_S}^4 = \frac{(1-\alpha) \cdot S}{4} \tag{5}$$

The value of the solar constant S can be approximated to be $1370Wm^{-2}$, and the value of the planetary albedo can be approximated as $\alpha = 0.3$. σ is just a constant whose value is $5.67 \cdot 10^{-8}Wm^{-2}K^{-4}$. (ibid) We can therefore solve the equation above for T_S:

$$T_S = \left(\frac{(1-0.3) \cdot 1370Wm^{-2}}{4 \cdot 5.67 \cdot 10^{-8}Wm^{-2}K^{-4}} \right)^{\frac{1}{4}} \approx 255K \approx -18°C \tag{6}$$

So in this model we get that the average surface temperature of the entire Earth should be minus 18 degrees, which we know very well is not the case. This is clearly because the greenhouse effect is not accounted for in this model. It is possible to change the simplistic model discussed here into an EBM where the greenhouse effect is also accounted for, consequently getting more realistic values for the average surface temperature, see for example the simple revision described in (McGuffie and Henderson-Sellers, 2005)[pp. 83-85].

Furthermore, it is also possible to add more complexity to EBMs by leaving this zero-dimensional view. Most EBMs are in fact one-dimensional, where the dimension of the Earth's latitude is taken into account. Since the energy balance is allowed to vary from latitude to latitude, a horizontal energy transfer term must be introduced, and we get the following general expression for the energy balance at each latitude, ϕ:

$$C\frac{dT_S(\phi)}{dt} = R_{in}(\phi) - R_{out}(\phi) + nettransportintozone\phi \tag{7}$$

where $R_{in}(\phi)$ and $R_{out}(\phi)$ is the incoming and outgoing radiation respectively, and can be substituted by the explicit expressions above, and the net transport into zone ϕ is the amount of energy that is transported into latitude zone ϕ from other latitude zones. So one-dimensional EBM allows for horizontal energy transfer across the Earth, making it slightly more realistic than the zero-dimensional EBM. However, still there are no atmospheric dynamics modelled in an EBM; this is simply approximated by having all heat transfer modelled by latitudinal heat diffusion. (ibid)

As mentioned a few times already, perhaps the most defining characteristic of the climate system is its energy balance and disturbances to it. This is exactly what EBMs try to model. We have seen how some of the features of the climate system discussed before can be translated into mathematical equations, and we solved the very simplest case analytically. It is easy to imagine that slightly more complicated versions of this could also without much problem be implemented on a computer. Although intuitive and easy to understand and run on a computer, EBMs are not able to simulate other climatic variables than the temperature, which greatly limits their potential use. Neither are they good at modelling temperature changes over longer periods, as they do not take the dynamics or biogeochemical processes such as the carbon cycle into account. However, as we will see in the following section, taking these further components of the climate system into account greatly complicates things.

2.2.3 General Circulation Models

In contrast to EBMs, which are either zero- or one-dimensional, the aim of GCMs is the calculation of the full three-dimensional character of the atmosphere and/or ocean. They are based on fundamental physical laws governing the temporal and spatial evolution of atmospheric and oceanic flow and thermodynamics. These equations are well known from other physical contexts, and include the equation of state, Newton's equation of motion, the first equation of thermodynamics, and continuity equations for air or water mass, water vapour, liquid water (cloud droplets), ice (ice crystals in clouds), and other atmospheric or oceanic traces such as chemical compounds and aerosols, or salinity in ocean water. (McGuffie and Henderson-Sellers, 2005)

That means that in effect, we have a set of four equations that define the dynamics of the whole climate system. This might raise the question as to why climate models are so complex, and so computationally expensive. However, let us first have a look at these equations in turn, and then return to the implementation, and the problems of implementing, these equations later on.

Equation of state

The equation of state describes the state of the system in terms of its state variables. For the atmosphere, the equation of state, ρ is given by

$$\rho = f(p, T, C) \tag{8}$$

where f is some function, p is pressure, T is temperature, and C is the actual composition of the atmosphere in terms of various gases, including water vapour. Since the atmosphere can be considered a mixed ideal gas, the equation of state can be simplified by the help of the ideal gas law, yielding

$$\rho = \frac{p}{RT} \tag{9}$$

where R is the gas constant for air. For the ocean, this is

$$\rho = f(p, T, S) \tag{10}$$

where S is salinity. The oceanic equation of state is a more complicated non-linear function that can be estimated empirically. (**GCM˙eqn**)

Momentum equation (Newton's second law/Navier-Stokes' equation)

Newton's second law of motion the rate of change of momentum of a body to the forces acting on that body. Expressed per unit mass of air or water, it can be expressed as

$$\frac{D\boldsymbol{U}}{Dt} = -2\boldsymbol{\Omega} \times \boldsymbol{U} \times -\frac{1}{\rho}\nabla p + \boldsymbol{g} + \boldsymbol{F}_r \tag{11}$$

where $\boldsymbol{U}$ is the three-dimensional velocity vector relative to the rotating Earth, i.e. the wind or ocean current vector, t is time, $\boldsymbol{\Omega}$ is the angular velocity of the Earth, $\boldsymbol{g}$ is the centrifugal and gravitational forces combined, and $\boldsymbol{F}_r$ are the molecular frictional forces. Expressed in this form, the equation is also known as Navier-Stokes' equation (in vector form).

Just as in the perhaps more familiar generic form of Newton's equation, $\boldsymbol{a} = \boldsymbol{F}/m$, the left-hand side of the equation above represents the acceleration, in this case of an air or water parcel. And correspondingly, the right-hand side represents the forces per unit mass, i.e. per kilogram of air or water. The first term on this side, involving the cross product of Earth's angular velocity $\boldsymbol{\Omega}$ and the velocity vector $\boldsymbol{U}$, corresponds to Coriolis force; a force apparent in any system whose velocity is relative to a rotating reference frame. (McGuffie and Henderson-Sellers, 2005)

Thermodynamic equation
The first law of thermodynamics describes the relationship between internal energy, heat and work in a thermodynamic system. For a unit mass air parcel it can be expressed as

$$c_v \frac{DT}{Dt} + p \frac{D\alpha}{Dt} = J \tag{12}$$

where c_v is the specific heat of air at constant volume, T is the temperature, p is the pressure, α the specific volume (the volume of one unit mass of air is $\alpha = 1/p$, and J is heat per unit mass added or extracted to or from the air parcel. The equation above therefore states that if heat is added to an air parcel it will go into increasing the internal energy (the first term on the left hand side), or to do work on the surroundings (the second term on the left hand side).

By inserting for the equation of state for atmospheric air above, it is possible to rewrite this equation as

$$c_p \frac{DT}{Dt} - \alpha \frac{Dp}{Dt} = J \tag{13}$$

where c_p is the specific heat of air at constant pressure.

The heat term J represents condensation of water vapour into cloud droplets or ice crystals, evaporation of falling precipitation or of cloud droplets/ice crystals, and radiation, and in some models also the net heating effect of molecular friction taking place at microscopic scales, i.e. heat from radiation, and heat from latent heat and phase transitions. (**GCM˙eqn**)

Continuity equations
Continuity equations describe the transport of some quantity. In our case they are used to describe the transport of mass, and since mass is a conserved quantity, the equations take on a particularly simple form. For dry air, the continuity equation can be written as

$$\frac{D\rho_d}{Dt} = -\rho_d \nabla \cdot \boldsymbol{U}, \tag{14}$$

and it states that the dry air density of an air parcel decreases proportionally if the flow is divergent, because in this case the volume of the air parcel increases and therefore density must decrease.

There are also continuity equations for other quantities, such as water. Since water can condensate and evaporate, it is necessary to include 'source' and 'sink' terms which account for these respective processes. For any tracer partial density i, the continuity equations are

$$\frac{D\rho_i}{Dt} = \rho_i \nabla \cdot \boldsymbol{U} + ss_i \tag{15}$$

where $i = 1, ...$ and ss_i is source or sink terms. For example, if water vapour condenses onto liquid cloud droplets the water vapour density will decrease (i.e. ss will be negative), while the liquid water density in the same air parcel will increase accordingly. **GCM˙eqn**

Although GCMs formulated in this way have the potential to closely approach the real oceanic and atmospheric situation, at present there are a number of practical and theoretical limitations In principle, these equations fully describe the dynamics of the climate system. However, biogeochemical cycles, such as the carbon cycle, are not modelled, neither is land physics, dynamic vegetation, nor ice sheets. GCMs are just descriptions of the dynamics of the atmosphere and oceans. This is however, exactly what ESMs aim to do. The ESMs used in most reasearch for the IPCC are fully coupled ESMs, which are at their core based on GCMs, but the coupled to other components that model these processes. (McGuffie and Henderson-Sellers, 2005) A review of the details of this is not within the scope of this thesis. However, because of their many similarities, looking at GCMs allow us to understand essential features of ESMs as well. Let us therefore continue to focus on the GCMs, but now shift the focus to how they can be implemented on a computer.

2.3 Climate Simulations

In the previous section, we encountered the general circulation models that are widely used to make climate predictions in climate science today. We also encountered the simple laws governing the behaviour of the atmosphere and ocean, namely Newton's laws of motion, the conservation of mass, and the first law of thermodynamics. However, what we get out of these laws is a coupled set of non-linear partial differential equations for which there is no closed-form solution. (McGuffie and Henderson-Sellers, 2005) That means that we cannot solve them exactly and analytically; rather, we have to get a numerical approximation of how a system obeying these laws should behave.

This process of discretization is done by approximating the original, continuous differential equations into discrete difference equations. A computer can then solve these difference equations step-by-step over discrete intervals of time for discrete points in space. Instead of a continuous function that provides us with the values of the variables of interest, such as temperature, for any point in time and space, we end up with numerical approximations for these variables on a discrete space-time grid. This means that the computer has to calculate the difference equations at each point in time and space. (ibid) When we then consider the resolution in both time and space and the timescale we are interested in to capture an accurate description of the behaviour of the climate system, we can begin to understand why such simulations are so computationally costly.

Despite the complexity of simulating the atmosphere alone, this is only a part of the climate models used today. We saw that some of the equations in the previous section also apply to the ocean, and we also saw that in more complex models, the complete hydrosphere (including not only oceans, but also rivers, lakes and other water components), the cryosphere, the land surfaces and the biosphere are also included, each with their own sets of equations. Full ESMs also tracks sources and sinks of carbon and other biogeochemical processes. To make simulations of such models then, the computer need to solve for all these equations, and account for the interactions between them, at each step in time and point in the space-grid.

This serves to illustrate a point that has been frequently repeated so far; namely that state-of-the-art climate models are complex and their simulations immensely computationally costly. There is a reason why massive resources are being put into making supercomputers that can run these simulations, and even using our best supercomputers, it takes months to make climate simulations of a couple of hundreds of years. The labour behind producing simulation results also provides an understanding for the motivation behind the close collaborative effort of the IPCC combining the efforts of different climate simulation groups around the world.

It should therefore come as no surprise that a complete overview of the simulation process of the most complex climate models is a lengthy task, one which is far beyond the scope of this thesis. Instead, we will settle on the more modest goal of describing two relevant ingredients in the long recipe for making climate model simulations, namely the discretization and parameterizations.

2.3.1 Discretization

Discretization refers to the process of turning continuous quantities into something discrete that can be solved by a computer. For climate models, this process refers to several sub-processes. Firstly, the Earth itself needs to be divided up into discrete points. The climate phenomena of interest, such as temperature and wind, cannot really be assigned to discrete points, but we need to somehow represent this continuous data discretely. Secondly, we need to estimate the spatial derivatives of the fluid equations encountered before of the discrete data. Thirdly, we also had time derivatives in these equations, and these also need to be estimated discretely. I will consider two main methods for performing these processes, namely finite-grid methods and spectral methods.

Finite-grid methods involve, as the name suggests, dividing the Earth up into a finite grid. So first of all, let us consider exactly how we can represent the Earth as a discrete grid. This process involves dividing the Earth up into grid-boxes, i.e. boxes with a certain extension in three dimensions. These boxes cover the surface of the Earth all around, and also extend vertically to include the oceans and the atmosphere. The simplest way to do this is to just divide the Earth's surface up into rectangular boxes, in a so-called 'regular latitude-longitude grid. This method has the advantages of the faces all being regular, and the coordinate lines being orthogonal to each other, making it easy to define spatial derivative operators. However, since the Earth is round and not square, this means that towards the poles the grid boxes will get smaller and smaller, eventually converging. This is known as "polar singularities". (S. N. Collins et al., 2013) As we will see below, the size of the grid-box is also intimately connected with the length of the time step required, meaning that a very high temporal resolution is needed for these small grid-boxes.

To avoid the polar singularity, other methods of discretizing the Earth have been utilized. The Cubed-Sphere grid for example, is obtained by placing a cube inside a sphere and 'inflating' it to occupy the total volume of the sphere. This means that it consists of the eight panels of the cube instead. Effectively, this means that the two polar singularities from the regular latitude-longitude grid are split up into eight weaker singularities at the corners of the cube. However, at the panel edges there are some kinks, meaning that it is difficult to construct differential operators here. (ibid)

Many global models also use the icosahedral geodesic grid. This grid can be viewed as an arrangement of triangular tiles covering the sphere. Each vertex on this shape is a model grid point. There are no polar singularities on this grid, and it has the most uniform element spacing, but the grid is largely unstructured, meaning that it can be difficult to compute high order differential operators. (ibid) The three types of grids mentioned here are just a few of many, and there are numerous alternatives to how to divide the Earth up into grid boxes. (S. N. Collins et al., 2013)

Vertically, the norm for all the grids is that they are divided up into a fixed amount of pressure levels, which typically follow the terrain by so-called "sigma co-ordinates", which makes mountain ranges easier to handle. Unlike the horizontal resolution, the vertical resolution is not usually uniform, and there is often finer vertical resolution near areas of particular interest, such as the tropopause and the surface. (McGuffie and Henderson-Sellers, 2005)

After choosing how to divide the Earth up into a grid, it is also necessary to define how and where the data in the individual grid-cells should be stored. For the rectangular grid boxes discussed above, the data can either be stored in discrete points, which is what is done in finite-difference (FD) models, or as averages over the whole cell, which is what is known as finite-volume (FV) models. The FV-method use an integrated form of the governing equations to track the flux of energy, mass and momentum at the cell boundaries. This leads to a conservative relationship between the fluxes and the enclosed entity reminiscent to that of Gauss' divergence theorem, and therefore the conservation of these quantities is an automatic consequence of this method. (Peiro and Sherwin, 2005)

In the FD-method, it is also possible to store the data either in an unstaggered way (Arakawa-A grid), where the scalar quantities such as density and temperature, and the vector quantities such as wind are stored in the same point in the cell, or in a staggered way (Arakawa-B, -C, -D grids[6]) where these quantities are stored in different locations. (ibid)

[6]Whereas Awakawa-C and -D grids are commonly used in different climate models, the type B grid is not implemented because of unphysical behaviour.

When the grid-boxes and how they store the data is defined, it further is necessary to define how to advance the difference equations in time. In finite grid methods, at every point the variables for a particular location in the grid is simply moved into computer memory and new calculations undertaken for the next step in time. The length of this timestep is chosen based on the timescale of the physical processes involved in the simulations, and in fact different processes are calculated at different timesteps, with the dynamics of the simulation usually having the finest temporal resolution. However, there is another important constraint on the timestep arising from numerical considerations. That is, it has to be short enough so that the maximum speed of propagation of information does not permit any transfer from one grid point to another within one timestep. This means that the timestep Δt has to obey the condition that

$$\Delta t \leq \Delta x / c \tag{16}$$

where Δx is the grid spacing and c is the fastest propagation velocity, which in GCMs is the speed of gravity waves. (McGuffie and Henderson-Sellers, 2005)[p. 170] This is what was hinted towards before, when it was mentioned that the time steps would lead to issues for the regular grid; when the grid boxes become very small towards the poles, Δx also become very small, meaning that the timesteps must be very short. As mentioned before, it is possible to use filters to diminish this effect, specifically the filters are used to overcome the instability caused by not matching the requirement above. (ibid) Conversely, it is possible to overcome this problem by using so-called "Langrangian time-stepping", which is a semi-implicit timestepping scheme which treats the motion of gravity waves, which effectively is what restricts the timestep, different so that longer timesteps become available. (El amrani and Seaid, 2008)

There is much more to say about finite-grid methods and the details of how these are implemented to produce climate simulations, but going into further detail about this is again beyond the current scope. Nevertheless, the point I want to make in this section is that firstly, simulating climate models is a complex process, and secondly, this complexity gives rise to alternative mathematical ways to simulate climate models. The spectral method of the following section will provide us with yet another mathematical alternative to the simulation of climate models.

The spectral method is based on a spherical coordinate system, making it particularly suitable for modelling the atmosphere, which effectively is a continuous spherical shell of air. Similarly to finite grid GCMs, spectral GCMs also divide the atmosphere into grid cells. However, fundamentally different is how the atmospheric fields are held and manipulated: namely in the form of waves. This makes it easier to calculate the gradients from the difference equations, thus making it computationally faster. (McGuffie and Henderson-Sellers, 2005)

The basic idea of spectral methods is based upon Fourier's theorem. This states that any ordered sequence of numbers can be represented as a sum of sine and cosine waves. In our context, the variation of any quantity around a latitude zone can be represented as a sum of a number of waves. The relationship between these waves and the original climate data is that they are the Fourier transform of this data. All the original information in the initial data series is still contained in this Fourier formulation, but the advantage is that some computations are much more readily done. (ibid)

However, this does not mean that all parts of the model are formulated by waves; a rectangular grid is used for vertical transfers, radiative transfer and surface processes are simulated in this grid space. This means that the spectral fields are transformed to grid space at every timestep via fast Fourier transforms and Gaussian quadrature (a form of numerical integration), and back to spectral space via Legendre transforms and Fourier transforms. The timestepping is thus performed with the waveform representation, while the grid-point physics is incorporated after the transformation into grid space.

The resolution of a spectral model is determined by what is known as the wavenumber of truncation. (McGuffie and Henderson-Sellers, 2005) For example, if a model uses 15 waves to represent each variable in a latitude zone at each vertical level, then it is said to be truncated at wavenumber 15. In fact, 15 was the wavenumber used in early applications of spectral modelling, whereas the norm in current spectral models is to use 42 zonal waves. Higher resolution is also used, and some weather forecasting models for example use 60 levels, which corresponds to a resolution of about 40 km. The resolution of the so-called 'Gaussian grid points', i.e. the points in grid-space, are determined by the truncation level of the model. A resolution too low compared to the truncation level will lead to aliasing of the high frequencies, whereas a resolution too low will not surprisingly lead to excessive computation times. (ibid)

Once again, this description is by no means exhaustive, and there is much more to say about this waveform description. However, just like before, this is not my intention. My intention has been to provide the reader with a basic idea of how a GCM can be implemented on a computer to make a computer simulation of the climate system. It has also been my intention to show how discretizing continuous equations describing a system continuous in time and space is not so straightforward, and can lead to different mathematical alternatives to this process. Hopefully this allows us to readily move on to the next topic: namely how processes that cannot be solved by the fluid equations at a specific resolution are treated.

2.3.2 Parametrizations

The fact that climate models are discretized sets of continuous equations on a grid is probably their most defining feature. The second most important feature then, is how the interactions that happen below the scale of the chosen grid are modelled, or at temporal scales shorter than the time steps used. The climate system consists of many sorts of processes, that happen at many different scales. The formation of clouds for example, can happen at scales down to meters and even millimeters, and from what I have been saying about the computational cost of running climate models, it seems obvious that we cannot resolve models to get this resolution (at least not now, or in the near future).

Despite the fact that we cannot solve for cloud formation, it also seems important to have them in our models. Clouds affect the climate to a large extent, as they both reflect sunlight and absorb LW-radiation from the Earth, not to mention they cause precipitation. Therefore, processes such as this need to be treated with a "sub-grid model", more commonly known as "sub-grid parametrization" or simply "parameterization". This is opposed to the so-called "resolved parameters", such as temperature and pressure, that are solved for by the dynamical equations, and are aptly referred to as the "dynamics" of the model. The sub-grid parametrizations then, are referred to as the "physics". It seems a bit mysterious why the dynamics of the model should be unphysical, especially because parameterizations usually involved various non-physical parameters. The parameters are 'non-physical' because their value do not correspond to anything in nature, instead they are "made-up" to best mimic the effect of the real physical processes, but do not aim to causally describe these effects. Both the need for parameterization and the appropriate values for these parameters are simply artifacts of the computation scheme, and hence non-physical. In addition to small-scale processes, parameterizations can also include processes that are too complex to be physically represented in the discretized model. (McGuffie and Henderson-Sellers, 2005)[ch. 2.5]

Important parameterizations include for example the vertical exchange of water in the oceans and small-scale turbulent mixing. However, by far the most problematic parameterization is that one mentioned above; namely clouds. In earlier climate models, the cloud parameterization simply consisted of calculating the percentage of a grid cell filled with clouds as a mathematical function of a few grid variables such as humidity and convection. However, cloud formation is a complex process involving complicated interactions between large-scale circulations, convection, turbulent mixing, radiation, and microphysical processes like precipitation (which is also dependent on aerosols), melting and freezing, and the nucleation of ice crystals. Modern cloud parameterization schemes therefore involve parameterizing some or all of these processes and then calculating the overall contribution to cloud physics. This is as complicated as it sounds, and in fact much of the differences in model output can be attributed to different approaches to this. (Winsberg, 2018)[ch. 4]

However, although these parameterizations are *non*-physical, that is not to say they are completely *un*physical, because they are usually based on physical considerations. One can differ between three categories of parameterization. (McGuffie and Henderson-Sellers, 2005)[pp. 72-73] The first and simplest one is the null parameterization, where a process, or a group of processes, is simply ignored. Although this perhaps sounds rather unphysical, this choice is still based on careful considerations of what processes actually significantly contribute to the phenomena of interest. If some processes only slightly or maybe not at all contribute to the output, one should not spend unnecessary computation time on them either. (ibid)

The second group of parameterizations have more rooting in reality. They are climatological specifications, and entails using empirical data to fix the parameterization. In early climate modelling for example, it was common to fix the oceanic temperature, including a seasonal variation, based on data of this. This type of parameterization is also utilized in present-day-modelling, but with land-surface characteristics, such as features of the soil and vegetation. It is important to realise that fixing variables like this means that feedback effects are suppressed: the effect of these variables are not allowed to change with changing climate forcings. Another type of parameterization that belongs to this category is that of "tuning". This involves parameterizing processes by relating them to present-date observations, constants or functions describing the relationship between variables are tuned to obtain agreement. (ibid)

The third type of parameterizations are those that have theoretical justifications. (ibid) An example of this is in some two-dimensional, zonally averaged dynamical models, the fluxes of heat and momentum are parameterized by baroclinic theory. (Sasamori and Melgarejo, 1978) This means that although the processes are parameterized, they are determined by theoretical considerations, and therefore can still be said to exhibit some causal structure akin to the real-life phenomenon.

In section 2.1 we saw some of the basic components of the climate and in section 2.2 we saw how we can model aspects of the climate system differently, but that we in principle have a thorough physical understanding of the dynamics of the system that can be described through a set of differential equations. However, in this section, it became apparent that we cannot solve these equations directly, and that simulating them leads to a number of complications. This means that there are inherent simplifications and idealisations in any climate model. Despite of this, climate models are still viewed as one of the most valuable tools we have for studying the climate. In the next chapter, we will therefore look at how these models are used to make hypotheses about the climate in the climate scientific practice.

3 Application of Climate Models: An IPCC Case Study

In chapter 2, the climate system was defined and briefly described, and the some types of models that aim to represent it were explained, as well as how they are numerically solved as computer simulations. From this we can infer that climate models can be used to simulate the behaviour of the climate system, and output certain climatic variables of interest. However, exactly what are these simulations used for in the climate scientific practice? In this chapter we will look at some characteristic examples of this, which will ultimately help us understand why robustness is so important in this discipline.

Perhaps the most defining limitation of a science that aims to describe a system as big and complex as the Earth system is that it is impossible to do traditional experiments to test hypotheses about it.[7] We cannot increase the radiative forcing to see how fast the ice caps will melt, and we cannot take human emissions out of the atmosphere to see how much of an impact they make. We only have one Earth system, the one we are living in; and it only has one history, the one we have observations of. There is no counterfactual information, no appropriate alternatives we can compare our reality with.[8] However, this is exactly what climate model simulations allow us to do. With them, it is possible to model the historical climate without the influence of human emissions, and we can even model what we think the future climate will look like under different conditions. We can also model the past climate in the periods where observations are sparse, to gain a more comprehensive description of it. And in all these processes, the scientists can learn more about the behaviour and mechanisms in the climate system. Evidently, the climate models provide the scientists with an incredibly powerful tool to study the climate.

The number one climate scientific organ is the Intergovernmental Panel on Climate Change (IPCC).[9] The IPCC do not produce their own climate models or do their own climate research, this is left to climate scientific

[7] This is of course not referring to experiments testing the physical theories the models are based on. These are well-established and essential to our understanding of the climate system. This is referring to the testing of the behaviour of the climate system as a whole, or the output of the models.

[8] It is interesting to note that we most certainly can compare components of our climate, in particular the greenhouse effect and climate sensitivity, to that on other planets. (Idso, 1988) However, this still cannot provide us with any knowledge of anthropogenic forcings, climate feedbacks specific to our planet, or possible future scenarios for our climate system.

[9] https://www.ipcc.ch/

research groups of the member countries, but they create extensive reports that are released around every six to seven years, where results from climate science are collected and discussed in one comprehensive report. There are three different workings groups under the IPCC that work on different aspects of climate, and each publish one report every report cycle, the so-called assessment reports (ARs). Working Group I deals with the physical basis of climate science and models representing the climate system. (Chen et al., 2021) Consequently, it is their ARs and the use of models in these that will be the focus in this chapter, as it represents the state-of-the-art of climate science and modelling.

The ARs are based on the Coupled Model Intercomparison Project (CMIP), which for the current AR, AR6, is running in its 6th cycle, CMIP6, whereas AR5 is based on CMIP5, and so on. Climate simulations that are made for the CMIP cycles consist of the outputs of many of the state-of-the-art climate models from the biggest climate modelling centres, and follow a specified protocol and specified experiment (i.e. initial conditions and forcing conditions). Their outputs are then combined to give the overall output of the CMIP cycle, which in turn gives the results and figures in the ARs, that IPCC base their conclusions and hypotheses about the climate on. (O'Neill et al., 2016)

In general, there are three key research uses of climate model simulations, namely detection, attribution and projection. Detection and attribution are closely linked, and involve detecting some change climatic change and then attributing this to specific causes, whilst projections simulate future climatic changes. In other words, detection and attribution deal with identifying changes that have been in the past up until today, whilst projections deal with possible future outcomes. Let us now have a look at each of these in turn.

3.1 Detection and Attribution

Although two separate processes, detection and attribution are really two sides of the same coin. In short, detection and attribution involve quantifying the evidence for a causal link between external drivers of climate change and observed changes in climatic variables. Ergo, it is the process of detection and attribution that allows us to establish whether and how much climatic changes are caused by humans, and other sources of climatic forcing. Since 2010, the IPCC good practice guidance paper on detection and attribution has existed (Hegerls et al., 2009), and its methods are used by the IPCC today. (Bindoff et al., 2013)

It is important to note that the process of detection and attribution is two-fold, and involves first detecting some change, and then attributing it to specific causal factors. It is therefore possible to detect some climatic change without being able to attribute it to specific causal factors. We can have high confidence in the fact that some climatic change is happening, without being so sure about what is causing it. On the other side, to make an attribution, it is necessary to already have established a detection. Let us now look at the characteristics of each of these processes in turn, to understand both their distinctiveness and their interdependence.

3.1.1 Detection

In the IPCC guidance paper on detection and attribution, detection of climate change is defined as

> "the process of demonstrating that climate or a system affected by climate has changed in some defined statistical sense, without providing a reason for that change." (Hegerls et al., 2009) [p. 2]

It becomes clear from this definition what was mentioned before; namely that it is not the aim of detection to give any reason why the detected change is happening. However, what is important is that a detected change should not be attributable to internal variability in the climate system:

> "An identified change is detected in observations if the likelihood of occurrence by chance due to internal variability alone is determined to be small, for example <10%." (ibid) [p. 2]

From this it follows that detecting climate change also requires an estimate of how much a quantity or field of interest might fluctuate as a result of internal variability, a so-called "null-hypothesis". This is not so straightforward to estimate from observational records. Although for some variables, estimates have been made based on paleoclimate data, they are usually obtained from long GCM/ESM simulations in which external conditions are held constant.

In its periodic assessments, the IPCC has reached increasingly strong conclusions about the detection of climate change in observations. In AR5 it was concluded that it is "virtually certain" (which means a probability greater than 99%) that the increase in global mean surface temperature seen since 1950 is not due to internal variability alone. (Bindoff et al., 2013)[p. 885] That is, the probability that the warming is due to internal variability alone was assessed by the scientists, on the basis of available evidence and

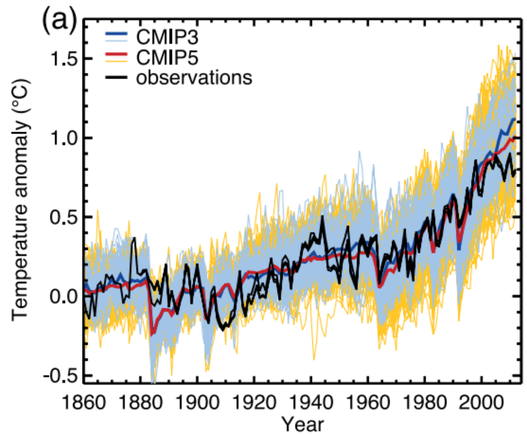

Figure 2: The black lines represent three different observational estimates of the Global Mean Surface Temperature (GMST), whilst the dark blue and red lines represent the median of the Coupled Model Intercomparison Project 3rd cycle (CMIP3) and Couped Model Intercomparison Project 5th cycle (CMIP5) outputs respectively. The light blue and the yellow lines correspond to the individual model outputs within these model ensembles. Note that temperature is represented as anomalies relative to 1880–1919 and not absolute temperatures. The figure has been taken from the AR5 (Bindoff et al., 2013) [p. 879, figure 10.1a]

expert judgment, to be less than 1%. Indeed, it was noted that, even if internal variability were three times larger than estimates from simulations, a change would still be detected. (ibid)[p. 881]

So let us look at an example of detection of global warming. Perhaps the most notable and familiar detection of warming is that of rising global mean surface temperatures (GMST). In figure 2 we can see a detection of rising GMST.

In figure 2 it seems apparent that the GMST is rising. However, remember that to classify as a detection of a climatic change, this change also has to be very unlikely to be caused by internal variability in the system. An apparent change could always very well be caused by natural variability, so we should be careful when "seeing" such trends with the naked eye. However, it has been shown that this observed trend in GMST since the 1950s is very large compared to model estimates of internal variability. When the observed trends in GMST were compared with a combination of simulated internal variability and the response to natural forcings, it was found that the observed trend would still be detected for trends over this period even if the magnitude of the simulated natural variability were tripled. (Knutson et al., 2013)

The detection of rising GMSTs therefore seems evident according to the definition of detection given above. Furthermore, if there is more energy added to the climate system, we would also expect to see a rise in ocean heat content (OHC). A warmer ocean also means that the volume of the ocean should increase, i.e. we should also see an increase in sea level. Some of the energy should also be absorbed by the Earth's cryosphere, and we would expect see the ice masses melting at an increased rate. Therefore, the IPCC also support the detection of rising temperatures with the detection of related variables. This reduces the possibility that the detection of any of the variables is faulty, or that the rise in GMST is due to some other causal process. The observation of changes in such variables can be seen in figure 3.

When a detection is made, and if it the change is unlikely to be because of natural variability, what remains is then to understand why it is happening. Put differently, what remains is to attribute the change to some causal factor.

3.1.2 Attribution

Apart from the difficulty in establishing a baseline for natural variability, and uncertainties in observational data-sets, detection is relatively unproblematic. Attribution on the other side, is a bit more complex. In the IPCC guidance paper on detection and attribution, attribution is defined as

> "the process of evaluating the relative contributions of multiple causal factors to a change or event with an assignment of statistical confidence. The process of attribution requires the detection of a change in the observed variable or closely associated variables". (Hegerls et al., 2009) [p. 2]

Ergo, unlike detection, attribution also requires knowledge of physical causal factors, as well as statistical analysis. Remember that we in section 2.1 reviewed some climate forcings, i.e. processes that have the potential of disturbing the energy balance of the climate system, thus changing it. Attribution then, aims to ascribe a detected climate change in terms of one or more of these forcings.

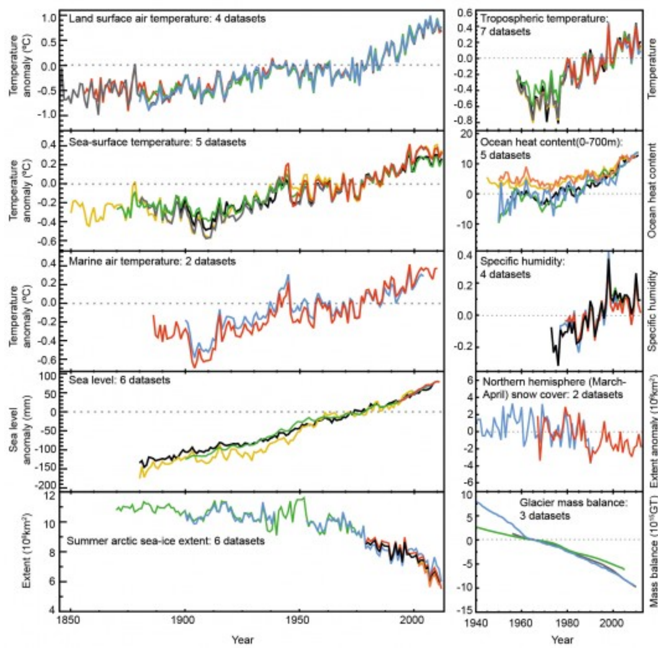

Figure 3: Detected changes in various climate variables. In the left section of the figure, we can see the detected changes in land surface air temperature, sea-surface air temperature, marine air temperature, sea level, and summer arctic sea-ice extent. In the right section of the figure we can see the detected changes in troposheric temperature, ocean heat content, specific humidity, Northern hemisphere snow cover, and glacier mass balance. The different coloured lines represent different data-sets. Note that most variables are given in anomalies relative to a baseline rather than absolute values. The figure is taken from (Hartmann et al., 2013) [p. 199]

From the definition above we can also understand that an essential feature of attribution is that an observed change does not only have to be shown to be consistent with a certain causal factor, it has to be shown to be inconsistent with an alternate explanation that lacks this causal factor. In practice, this usually means that the observations should be shown to be consistent with results from a process-based model that includes the specific causal factor, and inconsistent with another, otherwise identical model excluding this factor. This means that when climate scientists say that the temperature rise can be attributed to anthropogenic forcings, they are not only saying that the observations are consistent with model results implying this, but that the observations are completely inconsistent with a model that excludes anthropogenic factors.[10]

Let us return to an example, and see what such an attribution process can look like. In figure 4 we can see the attribution of the rise in GMST to anthropogenic forcings. As we can see in part (a) of this figure, there is seemingly a detection of a rising GMST, as also discussed in the previous section. However, to attribute this change to specific causal factors requires providing a counterfactual image, an estimate of the climate system without the causal factor of interest; namely the anthropogenic forcing. This is what we see represented by the blue line in figure 4 (a): a simulation of the warming of the GMST in a climate system without the anthropogenic (greenhouse gases and aerosols) forcing, i.e. with natural (solar and volcanic) forcings only. The orange line then, is the how this same evolution would look like in a climate system with anthropogenic forcings only. These estimates are obtained from the mean of the CMIP3 and CMIP5 ensembles modelling these counterfactual scenarios.

Ergo, the detected temperature change from our example is demonstrated to be completely inconsistent with natural forcings only, and thus it is attributed to anthropogenic forcings, finalizing one instance of a detection and attribution cycle. (Bindoff et al., 2013)

3.2 Projections

So detection and attribution deal with identifying climatic changes and their causes. However, perhaps the most important and pressing task for climate

[10]It might be of interest, although a total digression, to note that this is consistent with how science and technology theorists such as Bruno Latour describes objectivity in modern science, namely that the property of objectivity is about being able to defend a hypothesis against any objection, and furthermore, that this is what establishes it as a fact. (Latour, 2017)

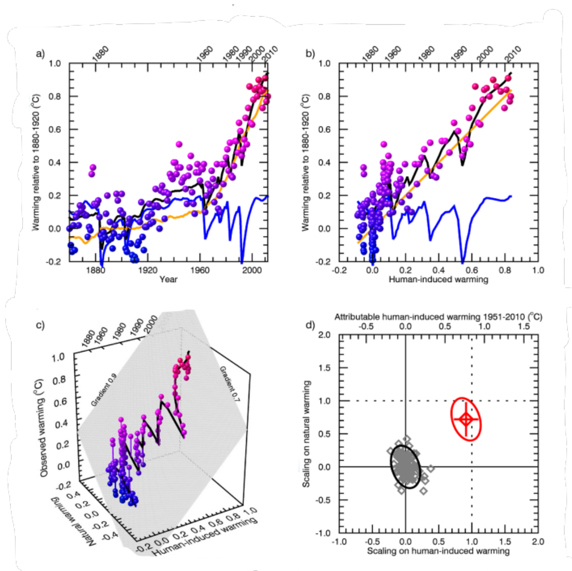

Figure 4: The simplified steps in a detection and attribution study. The coloured dots in part (a) are the observed global mean temperatures relative to 1880-1920, the black line is the best-fit linear combination of the observations, the orange line is the CMIP3/CMIP5 modelled mean temperatures with anthropogenic forcings, and the blue with natural forcings only. Part (b) depicts the same, but with all data plotted against model-simulated anthropogenic warming in stead of time. Part (c) shows the observed temperatures versus model-simulated anthropogenic and natural temperature changes, with best-fit plane shown by coloured mesh. Part (d) depicts the gradient of the best-fit plane in (c), or scaling on model-simulated responses required to fit observations (red diamond). Obviously, there is no overlap between the observations and model runs without anthropogenic forcings! Figures are taken from box 10.1 in (Bindoff et al., 2013) [p. 876])

models is to predict future climatic changes. Specifically, how the climate will respond to certain emission scenarios, i.e. different radiative forcings, is interesting in order to inform decisions about mitigation and adaption. This is what "projections" aim to do.

The IPCC defines a climate projection as

> "a climate simulation that extends into the future based on a scenario of future external forcing." (Kirtman et al., 2013)[p. 960]

I will address and explain what is meant by "scenarios" and "future forcings" shortly. However, firstly it is important to note that a "projection" is a technical term, and should not be confused with a "prediction". A weather prediction for example, begins with some specific initial conditions, namely the present weather conditions, and the simulation is run from these and a few days into the future. After that the weather prediction will start to be very uncertain because of the chaotic nature of the atmosphere. Furthermore, there are things which are not that important for weather prediction simulations, such as the conservation of mass and energy. This might sound counter-intuitive, but it is because whenever a new weather simulation is run, its initial conditions are manually updated to match observations of the current state of the atmosphere. This means that the continuation between a previous weather prediction and the current one is unimportant, making it irrelevant whether such quantities are conserved over time. Thus, the weather predictions are fully determined by the initial conditions we force them with.

So whilst weather predictions are based on initial conditions, climate projections try to eliminate any dependency on such initial conditions. This is partly because we do not want our climate models to be dependent on the specific details of the constituents of the climate system at the time the simulation happen to start. There is no guarantee that the specific time that is chosen is representative of the year or even period it is from. However, there is also a deeper reason for why it is undesirable to have a projection dependent on initial conditions. As mentioned before, it is impossible for a weather prediction to be accurate beyond a certain amount of days. This is because it is a chaotic system, and is consequently highly dependent on initial conditions. A small change in the initial conditions could result in a dramatically different prediction. (Dragutin, 2020)

Instead, the simulation is usually equilibrated for a certain amount of time, and then the starting point is set to be after the equilibration, and not from the point of the initial condition. Another method used to eliminate the dependence on the initial conditions is to manually or probabilistically vary the initial conditions within a model ensemble, so that the model ensemble as a whole is free from specific initial conditions.

So a projection aims to be independent of any particular initial conditions. Instead, it is dependent on certain boundary conditions, specifically on a particular "scenario of future external forcing". These scenarios of future external forcings entail specific scenarios of the amount of radiatively active species (i.e. greenhouse gases or aerosols) that is emitted into the atmosphere. Firstly, a possible emission scenario is estimated, and then biogeochemical models are used to calculate the corresponding concentrations of constituents in the atmosphere. Then, using various radiation schemes and parametrizations, these concentrations are converted into radiative forcing. The response of the different climate system components is then calculated in a comprehensive climate model, such as an GCM or ESM, resulting in the climate projection. (Meehl et al., 2007) From this, the possible future values for specific variables in a certain emission scenario can be extracted.

The scenarios operated with by the IPCC have changed over the years. Perhaps the most well-known ones are the so-called "Representative Concentration Pathway" (RCP) scenarios that were developed for and extensively used in AR5. (Kirtman et al., 2013; M. Collins et al., 2013) These consist of four different 21st century pathways of greenhouse gas emissions, air pollutant emissions and land use. They are then named after the overall radiative forcing they would result in at the end of the century. So for the scenario RCP2.6, this would entail a radiative forcing of 2.6 W m^2, and similarly for RCP4.5. RCP6.0 and RCP8.5. Although these scenarios are still used in AR6, there are now used together with RCP1.9 (which corresponds to the aspirational goal of the Paris Agreement), RCP3.4 and RCP7, as well as the Shared Socioeconomic Pathways (SSPs), that also try to take various political and social factors into the development of the scenario. (Chen et al., 2021) Although the details of each of these are interesting, what is important to note is that all of them somehow represent a specific future emission scenario.

So instead of delving further into these details, let us turn to an example of a projection. In figure 5 we can see the projected global relative temperature change up until year 2100. In this figure we can see many of the components mentioned so far. SSP scenarios and their corresponding radiative forcings are shown, and the different projected temperature changes correspond to different scenarios. The number of models that have gone into making each projection is shown next to the scenario name. The reason why some scenarios consists of a lot more models is mainly because of

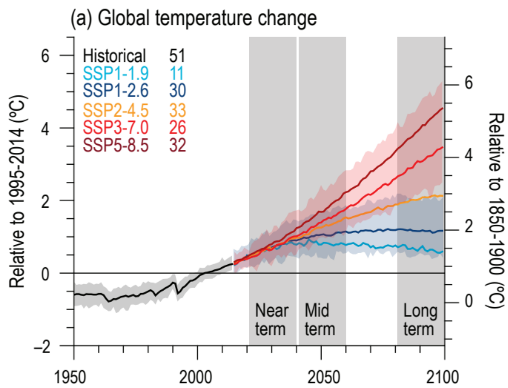

Figure 5: Temperature as an indicator of global climate change up until year 2100. The different scenarios are marked by the different colours that can be read in the top left corner. The two different vertical axes represents the temperature change relative to two different baselines, namely relative to 1995-2014 and 1850-1900 respectively. It is assumes that by 1995 some warming due to anthropogenic emissions had already occurred, and the relative change is therefore smaller. Figure is taken from figure 4.2 in (Forster et al., 2021)[p. 571]

which scenarios have traditionally been prioritised and developed, whereas for example SSP1-1.9 is new for AR6. The shaded area around the overall projections represent the range of the single model outputs, and as mentioned before, this is taken to be an estimate of the uncertainty of the projection.

Figure 5 also highlight an essential feature of how the IPCC use models, namely that the output is not based simply on one model, but the overall output of a so-called "multi-model ensemble"(MME). It was briefly mentioned before that the ARs are tied together with a corresponding cycle of CMIPs, and we could also see the individual model outputs of the GMST in figure 2. However, this aspect become even more apparent looking at the ranges of the projections in figure 4, as well as seeing the number of models going into the different projections. And indeed, it is such a defining feature of IPPC's practice that it is worth elaborating on. Empirically, it has been shown the result of such an MME overall performs better than individual models, even if the confidence in a certain model is high. (M. Collins et al., 2013) This can be explained by the fact that no model is completely "correct", and it is not so straightforward to tell just exactly which one performs better in a given area and why. The average of the models therefore seems to balance out the model biases of the different models, giving an overall more plausible result. The range of the different model projections then provides a basis for quantifying uncertainty in the projections. This agreement or "robustness" between different climate models is therefore taken as a confirmatory virtue, and exactly this will be the topic of chapter 4 and 5.

However, before examining this further, let us look at the apparent discrepancy between the IPCC's very confident statements about the climate system and its behaviour on one hand, and the many uncertainties and lacks of the models used to support these hypotheses on the other hand.[11] This will naturally lead us to the topic of model agreement and its epistemic power in chapter 4.

3.3 Uncertainties and Confidence Levels

> "It is *virtually certain* that, in the long term, global precipitation will increase with increased global mean surface temperature." (M. Collins et al., 2013)[p. 1032]

> "There is *very high confidence* that globally averaged changes over land will exceed changes over the ocean at the end of the 21st century by a factor that is likely in the range 1.4 to 1.7. (M. Collins et al., 2013)[p. 1031]

> "It is *very likely* that the Atlantic Meridional Overturning Circulation (AMOC) will weaken over the 21st century but it is *very unlikely* that the AMOC will undergo an abrupt transition or collapse in the 21st century." (M. Collins et al., 2013)[p. 1033]

These are statements that can be read in the IPCC's AR5, based on the scientists' confidence in projected climate changes. As we can see, they are assertive statements expressing strong claims about the likelihood of future climate changes. However, on the other side, it is possible to read in AR5 that

[11] It is important to highlight that this does not mean that I think it is problematic that the IPCC bases hypotheses by the help of these models. I simply want to point to the perhaps counter-intuitive idea that the climate scientists are able to put forward hypotheses of high confidence in spite of these complications. One such mechanism is the agreement between different models in MMEs, and as we will see in the chapter 5, this is an epistemologically sound method.

> "Projections of climate change are uncertain, first because they are dependent primarily on scenarios of future anthropogenic and natural forcings that are uncertain, second because of incomplete understanding and imprecise models of the climate system and finally because of the existence of internal climate variability." (M. Collins et al., 2013) [p. 1034]

The fact that the projections are uncertain because they are dependent on scenarios of future forcings is perhaps the most obvious uncertainty. We simply do not know exactly what anthropogenic emissions will look like in the future, and that is why the IPCC looks at different possible emissions scenarios, to account for different possible futures. In addition to this comes the uncertainty regarding natural forcings in the future. We do not know when there could be a possible volcanic eruption or how much aerosols it will release into the atmosphere. And although we have extensive knowledge of past solar cycles and sun spots, we cannot know the precise future activity from the sun either. So in summary, although no matter how many factors that are accounted for, it is simply impossible to know the exact future radiative forcing, both because it depends on decisions that are not yet made, and because it involves potential unpredictable natural forcings.

The second factor mentioned in the quote above, namely the "incomplete understanding and imprecise models" is a complex category involving many different aspects, and applies to detection and attribution studies as well as projections. Firstly, from chapter 2 we know that there are many ways to construct a climate model. For the case of GCMs, we saw that although they are based on the same fluid dynamical equations, there are numerous ways to realise the simulations. Firstly, how the Earth is to be spatially discretized needs to be addressed; we saw two main methods for doing this, namely finite-grid methods, which entails even further choices about the shape of the cell, and the spectral methods. It is also necessary to define how data is stored within a grid-cell, the size of the cell, as well as to define an appropriate time-step. This means that there is no "best" method, no completely accurate way to make the simulations of the models, leading to inaccuracies in the projections as well.[12]

[12]Note that there is a difference between saying the models outputs are uncertain and that the models themselves are uncertain. I am refraining from labelling the models as "uncertain", although this is often how their impreciseness is defined. Strictly speaking, the models are not "uncertain". Given the same conditions, it is perfectly possible to recreate and predict what the model output will be (unless of course, there are stochastic variables involved. Still, with the same randomly generated numbers, the model output

Another major factor contributing to the impreciseness of the climate models is parameterizations. Recall from section 2.3.2 that some processes cannot be resolved at the grid-level, and therefore need to be parameterized. Processes such as these can either simply be removed, be empirically estimated, or be based on physical equations. The need for parameterizations can be a result of either the resolution of the model, or of a lack of understanding about certain processes in the climate system. The parameterizations vary greatly between models, and are therefore a major source of uncertainty and differences between models.

A lack of understanding of the climate system also include limited knowledge about what the IPCC calls the "response of the climate system". The fact that the climate system is nonlinear, and that we do not exactly understand how all the feedback mechanisms work, means that there will be uncertainties regarding how the climate system will behave in the future. Even though there are data and reconstructed data-sets of the climate system in the past, under different forcing conditions, the concentrations of greenhouse emissions we are headed towards simply are unprecedented, and it is therefore not possible to straightforwardly use information of past climate to simulate the future. Note that this is related to the uncertainties regarding parameterizations, because parameterizations are often based on past data and therefore will not change accordingly with unknown radiative forcings. There is also of course the possibility that there are elements that are currently not accounted for in the models at all, but that will be of importance in the future.

The last category of uncertainties listed above is those that arise with respect to internal variability of the climate system. The internal variability is a natural consequence of the chaotic nature of the system, and is difficult to completely get rid of. The internal variability can be sampled and estimated explicitly by running ensembles of simulations with slightly different initial conditions, or can be estimated on the basis of long control runs where external forcings are held constant. Nevertheless, even in these cases internal variability is always an estimate based on model simulations, and is therefore subject to the same uncertainties mentioned above.

However, despite these uncertainties, the IPCC makes strong claims about future conditions like the ones we saw above. The IPCC operates with a framework for assessing uncertainties, which appeals to two mea-

will also be the same). Hence, there are no uncertainties involved in the model and its simulation. However, the models are inaccurate and imprecise representations of the climate system, meaning that projections of future climate change, and attribution and detection studies, have some uncertainties in them.

sures for communicating uncertainty. The first is a qualitative "confidence" scale, which depends on both the type of evidence and the degree of agreement. The second measure is a quantitative scale for representing statistical likelihoods for relevant climatic variables. It is an interesting question in itself how exactly this evaluation framework works, and how a qualitative measure of "confidence" is translated into a quantitative measure of "likelihood". There are some interesting discussions inherent in the construction and application of this framework, and a critical analysis of it can be found in for example (Wüthrich, 2017). However, this is not the question of this thesis, so we will forget this framework and instead focus on the general question of how confidence in climate hypotheses can be achieved.

What I have tried to show so far is that climate models are complex and incomplete and their simulations inherently imprecise in many ways. Despite their many problems, the output of MMEs are used to make strong claims about the behaviour of the climate system and future climatic changes. So what exactly is it that gives the scientists the confidence in hypotheses supported by these modelling outputs? And in hypotheses about the climate in general? In the next chapter I want to focus on one important virtue that does this, namely robustness.

4 Robustness Analysis

For very complex systems, such as the climate system, it is not possible to confirm theories and models merely by traditional methods of validation and verification. It is certainly possible to test a model's performance compared to observational data, but even this is not always so straightforward either, since the models often are tuned with the very same data they are compared to. Not to mention that data often is lacking over long periods and sparse in many areas. Furthermore, for future scenarios data are obviously not existent at all, and we are already seeing unprecedented levels of greenhouse gases, that likely will continue into the future. The behaviour of the climate system might therefore not be so easily comparable to past data, and therefore hard to predict given our lack of complete understanding of many processes in the climate system. Robustness analysis provides such an alternative method.

We have already seen in chapter 3 how the output of multi-model ensembles is used in detection and attribution and projection studies. In general, it seems like there is more confidence in a result if more models agree on it. Pirtle, Meyer, Hamilton found 118 articles in climate science where the authors refer to agreement between a variety of climate models to encourage confidence in the results of the models. (Pirtle et al., 2010) This idea is also formalised by the IPCC, which state that confidence depends on

> "the type, amount, quality and consistency of evidence (e.g. mechanistic understanding, theory, data, models, expert judgment) and the degree of agreement" (Mastrandrea et al., 2010, p.1)

Some pressing questions that follow from this practice of establishing confidence based on the agreement between models, namely whether this robustness by itself really can carry the epistemological weight ascribed to it, and whether model agreement by itself really can serve to increase the confidence in a hypothesis?

Although it might seem like an intuitive idea, why exactly do we have more confidence in a result when there is such agreement? And what exactly counts as more evidence? These might sound like trivial questions, but let us consider an example to shed some light on the questions that arise from this.

Let us imagine you're a bird enthusiast. You read about the sighting of a particularly rare bird in your area in the newspaper. However, with your knowledge of birds, you know it is very unlikely to be seen in your area.

Therefore, you want to increase your belief in the sighting, but what exactly then would count as "more" evidence? Just counting the pieces of evidence is not enough, consider simply if you had 13 copies of the same newspaper writing about it, this will obviously not increase your belief accordingly. However, if you read about it in a second newspaper, this might increase your belief that it is true, and that it was not just a glitch caused by the first newspaper. You then go on to read about it in more newspapers, but after reading about it in the fourth, fifth or even ninth newspaper this again might cease to count as "more evidence" of the sighting. However, if your local ornithological community writes about it in their newsletter, this might increase your confidence as they are less likely to have confused the specific bird in writing with another, more common species. However, you still might have doubts, but hearing your like-minded bird enthusiast friend seeing it for herself, hearing a zoologist describing it, or perhaps seeing one of its characteristic feathers at the location of the sighting, will certainly count as more pieces of evidence that will increase the credibility of the rare bird actually being in your area.

These might seem like mundane observations of what counts as more evidence, and it seems obvious that reading about the bird sighting in the ornithologist newsletter, or hearing a friend observing it, gives us more confidence in the sighting than what 13 copies of the same newspaper would. It is not merely the quantity of the pieces of evidence, but, perhaps more so, the quality. So what if an ensemble of climate models tells us that we should expect the ice sheets of Greenland to melt at a certain rate? Is this more like reading 13 copies of the same newspaper, or is it more like spotting the feather? The challenge is that climate models are often very similar, sharing assumptions, calibration methods, and even code, and therefore not independent the same way as your friend's observation and the feather might be said to be. So can they then count as individual pieces of evidence? And perhaps more fundamentally, do we really need the pieces of evidence to be completely independent for them to increase our confidence in the result? If not, what condition will?

Robustness analysis then, is the attempt giving a systematic account to answer questions like these, and exactly pin down how this consilience of evidence should work to increase our confidence. When exactly, and under what conditions, do more pieces of evidence contribute to give us more confidence in a hypothesis? And what is the relation between the pieces of evidence that contribute to such a robustness? So in order to understand *how* and more importantly, *if* multiple models in an ensemble can contribute to give us confidence in climate hypotheses, let us look at some different accounts of robustness analysis (RA).

4.1 Robustness Analysis

So robustness analysis is an area of the philosophy of science examining and assessing if and how the robustness across various means of detection can be confirmatory. Although robustness and model agreement is widely referred to in climate science, and consequently is also a topic debated by philosophers focusing on this science,[13], robustness analysis is not a concept exclusive to climate science and modelling. It has also been extensively applied to for example biological models (Kim et al., 2006), cosmology (Gueguen, 2020) and economics (Kuorikoski et al., 2010). A shared feature among many notions of robustness analysis has been the idea of *independence* between the models or other means of detection as a necessary condition to explicate confirmation for a hypothesis. In this section we will, closely following Jonah Shupbach's criticisms (2018), see how such an independence account is unfeasible. We will then review Shupbach's alternative framework, namely Explanatory Robustness Analysis, and preliminary discuss its potential application in climate modelling and Winsberg's interpretation of this.

4.1.1 Probabilistic independence accounts of RA

Biologist Richard Levins (1966) was the first to introduce the concept of robustness analysis into the philosophical debate. Biological models can be very complex, and because of this many biological models rest on heavy simplifications and idealisations, just like climate models. Levins was concerned with when we could know whether detected results derived from such models depended on essential causal features of the model, or on the details of these assumptions. He argued that it is possible to figure this out if

> "...we attempt to treat the same problem with several alternative models each with different simplifications but with a common biological assumption. Then, if these models, despite their different assumptions, lead to similar results, we have what we call a robust theorem which is relatively free of the details of

[13]See for example a variety of articles put forward by Elizabeth Lloyd and Wendy Parker, who have been particularly active in this discussion and in some ways can be said to represent opposite views, Lloyd arguing that robustness analysis can be confirmatory in the context of climate science (Lloyd, 2015) and Parker arguing that it cannot. (Parker, 2011)

> the model. Hence our truth is the intersection of independent lies."(Levins, 1966) [p. 423]

So Levins argues that by considering various models of a common causal core but of varying assumptions, a common prediction among these models would qualify as robust, and we could have confidence in the commonly detected result.

It is worth to note that although Levins' description was aimed specifically at models, other means of detection can also easily be included, like Shupbach did in his paper "Robustness Analysis as Explanatory Reasoning" (2016). Here he develops an account of robustness analysis applicable to both experimental results and model results, using the examples of the experimental detection of Brownian motion, and the Volterra principle as detected by multiple predator-prey models. Furthermore, he claims that his notion of RA can be further extended to any type of detected results, including "observations, measurements, predictions, theorems, and so on". (ibid)[p.2] Shupbach's line of argument will be followed closely in this section, and although a digression from Levins, the examples found in Shupbach will be explained now as they effectively illustrate the concept of robustness analysis, and also will be invaluable to see why Levins' account falls short.

The experimental detection of Brownian motion refers to a set of experiments conducted to observe Brownian motion, firstly carried out by botanist Robert Brown, and later other scientists, before finally arriving at Einstein's hypothesis that Brownian motion was caused by molecular interactions. (Perrin, 1913) When Brown first observed the motion, he hypothesized that it was characteristic to the specific type of pollen, which had a particular shape. However, when detecting the motion in other pollen particles as well, this proved not to be the case. He then hypothesized that it was caused by the vital forces in pollen, however, when detecting the motion in inorganic material as well, this could similarly no longer be a likely hypothesis. The motion was later observed in different environmental conditions, in different types of containers and with different equipment. The Brownian motion was robust across the various changes to the experimental set-up, including the size of the particle, the medium, the container, etc., whilst it was sensitive to others, including the size of the particle and the temperature. Performing a robustness analysis of this therefore allowed the scientists to infer that it was molecular and thermodynamic properties that could explain the phenomenon. (Shupbach, 2018)[pp.1-2]

The Volterra principle is an example of model-based RA previously also cited in for example (Weisberg and Reisman, 2008). In line with Shupbach's conviction, Weisberg argues that this principle, arising from the Lotka-Volterra predator-prey model, is a prime example of robustness analysis. This model is a mathematical biology model representing the population of predators and preys by a set of differential equations. The Volterra principle then, states that if a constant proportion of both the predator and the prey populations are continuously removed, for by biocides, the average number of predators will decrease relative to the average number of prey. This principle emerges in a range of predator-prey models with varying assumptions and idealisations. This principle is therefore detected by various means of detection (the means being the models themselves) and is therefore a robust result. As Weisberg argues then, this shows that the Volterra principle is robust and should give us confidence that it describes a real ecological phenomenon. (ibid)

In addition to this, let us throughout this section keep a climate hypothesis described by Winsberg (2018) that will be of great importance in section 4.2 as well as in chapter 5, namely that the Equilibrium Climate Sensitivity (ECS) lies between 2.1°C and 4.7°C. If hypotheses based on a commonly detected results between such climate models can be given confidence based on robustness analysis, we would want our conception of robustness analysis to be able to account for this example too.

However, back to Levins, we can see that the intuition underlying his description of RA is the same intuition we arrived at above: that we can get confirmation through a diverse set of means of detection agreeing on some detection. This is also something we can infer from Shupbach's examples of experimental and model-based RA from the history of science. However, the question still remains; what should the relation be between such means of detection making a robust detection? Although Levins did not explicitly define this relation himself, a key word in his account of RA is the word "independence". Perhaps because of this, the discussion that followed Levins focused on probabilistically independent means of detection as a criterion for how different the models in a robust collection have to be. After Shupbach (2018) can call this measure "RA-diversity", describing the relation between models/other means of detection required in order to be able to provide confirmation for a detection.

Firstly, let us take Levins' notion of "independence" literally, like Orzack and Sober (1993) did in their critique of his account. In this account, unconditional probabilistic independence is required as a criterion for RA-diversity. Let R be the detection that has been robustly detected by various means. Then let the proposition that this result is detected using the k'th means of detection as R_k. Then, two means of detection are RA-diverse if the fact that R is detected by means i should have no influence on the probability that R will be detected using means j, i.e.

$$Pr(R_i R_j) = Pr(R_i) \times Pr(R_j) \tag{17}$$

assuming that the probabilities $Pr(R_i)$ and $Pr(R_j)$ are both greater than zero. Furthermore, this implies that

$$Pr(R_i) = Pr(R_i \mid R_j) \tag{18}$$

and vice versa that

$$Pr(R_j) = Pr(R_j \mid R_i) \tag{19}$$

Orzack and Sober then, argued against such an unconditionally probabilistic account. They argue that requiring the models to share a "common biological assumption", like Levins did, in effect guarantees that the models are not independent. More specifically, when one model in an RA ensemble implies a particular result, it is natural to assume, again like Levins did, that the result is driven by the causal common core of the model. However, detecting such a result will then effectively raise the probability of detecting the same result with another model, also sharing this causal core. (Orzack and Sober, 1993) Adapting this to our previous terminology, the detection of R by a model i will increase the probability of detecting R by model j sharing the same causal core, i.e.

$$Pr(R_i) < Pr(R_i \mid R_j) \tag{20}$$

Orzack and Sober argue that this can be true in cases where models are considered RA-diverse, also exemplified by Shupbach's examples. To understand this, let us take a closer look at Shupbach's example of Brownian motion. In this example, it is simply not true that the diverse means of detection are probabilistically independent. If we take any two of the experiments detecting Brownian motion, such as the experiment suspending dust particles in water and that of suspending them in ethanol. Shupbach then argues that detecting Brownian motion in one of these experiments will also affect the probability of detecting it in the other experiment. This is a result of the fact that there are many different factors that can contribute to any of these detections, such as the type and size of particle, the suspension of the particle, and other environmental conditions. Shupbach argues that this is true in any case where the experiments are allowed to potentially be influenced by factors other than the one in which the experiments differ.

In light of this, it is easy to see that unconditional probabilistic independence cannot be a necessary condition for RA-diversity, at least not if we want it to describe these paradigmatic examples of RA in science.

Revisiting our climate science example, it is also easy to understand that any hypothesis supported by multi-model ensembles, such as that about the ECS, cannot live up to this condition of independence. Seeing the many similarities shared by the models, it is trivial that a detection made by one model will influence a second, similar model's detection of this result too. This means that such an independence account cannot help us explain why model agreement is given such epistemic power either.

Abandoning the idea that the means of detection must be completely probabilistically independent of each other, Wimsatt (1994) proposes that instead "the *probability of failure* of the different means of access should be independent". By abandoning the condition that the results of various means of detection must be completely probabilistically irrelevant of each other, this account escapes the specific problems mentioned above. Instead, the probability of the various means of detection to detect a wrong result must be independent. This means that if two given models in an RA ensemble detect the wrong result, they should do so for two different reasons. And moreover, if we learn that one model has given us the wrong result, this should not affect the probability that another also will do so. The *reliability* of each model is independent, and this account can therefore be called the *reliability independence* account.

As Shupbach notes, this account elegantly emphasizes the epistemic appeal of RA-diversity: "whereas a linear chain of justification can be no stronger than its weakest link, a web of independent lines of justification is no weaker than its strongest member". (Shupbach, 2018)[p. 6] This epistemic advantage can be shown by considering an ensemble of n means that detect a common result, and these means are reliability independent. For simplicity we assume that the probability of each means of detection detecting the wrong result to be p_0. If we have a result that the means have all commonly detected, and this result turns out to be wrong, that means that all of the means have failed independently. Therefore, the probability of this happening is $p_p = p_0^n$, and if we assume that $p_0 < 1$, we have for $n > 1$ that $p_p < p_0$. So the probability of a single means of detection detecting the wrong result is always greater than the probability of a collection of robust means detecting the wrong result, and the more means in the ensemble the smaller the probability becomes (given reliability independence).

Despite of this, as Shupbach argues, reliability independence cannot serve as a necessary condition for RA-diversity either. It is perfectly possible to imagine RA-diverse means of detection that are not reliability independent. Returning to the example of experimentally detected Brownian motion, there are many features common to all the means of detection that could lead us astray: the specific experimental set-up, the environmental conditions surrounding the apparatus, the medium used in the experiment. This means that learning that one means of detection is detecting the wrong result could affect the probability of another being wrong too; their reliability is in effect not probabilistically independent. This means that if we, like above, still want to stick to the case of Brownian motion as a paradigmatic example of RA, reliability independence cannot be necessary to explicate RA-diversity either.

The shortcomings of reliability independence is just as easily demonstrable with examples from modelling. As Shupbach points out when it comes to the Volterra principle, some models share similar assumption, e.g. the unrealistic assumption that prey cannot take cover or learn. But then discovering that one of the models is unreliable should often greatly increase our confidence that the other is too. In general, fully RA-diverse means of detection can nonetheless be susceptible to many of the same potential defects. In such cases, learning that one of our means of detection is unreliable will often greatly increase the likeliness that more of the means of detection are similarly unreliable.

Furthermore, revisiting our ECS hypothesis supported by the multi-model ensemble, it is also easy to understand that the models that support this hypothesis are not reliability independent either. It suffices to say that many of the models of the models share common assumptions, such as how to parameterize certain physical processes for example, and that learning that one of these models is unreliable can increase the confidence that another model sharing similar assumptions is unreliable as well.

4.1.2 Explanatory Robustness Analysis

It seems like the accounts above cannot give a satisfactory notion of RA-diversity. So what can? As Shupbach notes, "it seems [...] that philosophers working on RA have been lured away from the concept of RA-diversity by probabilistic independence". (Shupbach, 2018)[p. 10] It is worth to mention that Shupbach does not deny that the notions of RA mentioned above will in fact provide the conditions required to gain confidence in a detected result. In fact, he notes that such independence-accounts "imply interesting senses in which diverse bodies of evidence may be specifically confirmatory." [p. 5] However, he does not want to accept the restrictions constrained by them as necessary to achieve the required diversity, as they fail to describe essential examples of RA in science. I share Shupbach's conviction that probabilistic independence is too strong and perhaps the wrong thing to pursue in the first place when looking for confirmation of common results detected by various means. Let us therefore take a closer look at the account Shupbach offers instead; namely Explanatory Robustness Analysis (ERA).

Instead of the top-down approach taken by many of his predecessors, where the starting point is a notion of RA-diversity which is then applied to examples of RA, Shupbach starts in the opposite end. By looking at his two examples of RA he attempts to identify what defines and unifies them, and from this he builds his conception of RA-diversity. Shupbach then, claims that what really differentiates one means of detection from the next is that it is capable of ruling out another class of competing potential explanations. When Brownian motion was detected in various types of pollen for example, the possibility that it was a phenomenon caused by the pollen itself was ruled out, and similarly, when detecting the result in inorganic materials, the possibility that it was caused by the vital force of organic materials was also ruled out. Each detection therefore had the potential of ruling out some explanation, until the only explanation that seemed likely was the final hypothesis that Brownian motion is due to internal, invisible movements in the medium. Shupbach argues that the same tendency is true for the Volterra principle: the different models rule out competing explanations of the principle being detected as the result of specific assumptions in the models. Shupbach sees these notions of *explanation* and *elimination* as the essential components of RA.

Shupbach then defines RA-diversity in the following way:

> "Means of detecting R are *RA-diverse* with respect to potential explanation (target hypothesis) H and its competitors to the extent that their detections (R1, R2, ..., Rn) can be put into a sequence for which any member is explanatorily discriminating between H and some competing explanation(s) not yet ruled out by the prior members of that sequence." (Shupbach, 2018)[p. 11]

Ergo, RA-diverse means of detection are defined by being able to explanatorily discriminate between pieces of evidence, and are able to successively eliminate alternative hypotheses. The definition above sheds some light on what Shupbach's account consists of, but is still informal and lacking in terms of describing its central terms, such as what it means to be "explanatorily discriminating", or what "competing explanation(s)" really are.

However, closely following the above definition, Shupbach begins defining the "logic of robustness analysis" and provides the following formal account of a successful increment of RA:

1. **Past detections**: We have a result R that we have detected using $n-1$ different means. We let $E = R_1 \& R_2 \& ... \& R_{n-1}$.

2. **Success**: The target hypothesis H explains the detections in the conjunction E, but so does an alternative hypothesis H'. Formally we have that $\epsilon(E, H)$, $\epsilon(E, H')$, where ϵ is defined as the explanatory power that a particular explanans (the explanatory account) (h) has over explanandum (the fact to be explained) (e).

3. **Competition**: H and H' epistemically compete with each other with respect to E. Formally, we have that i) $Pr(H \& H') = 0$ or ii) $\epsilon(E, H \mid H') \leq 0$ (i.e. the probability of both H and H' being true is zero, or less strongly, given H', H no longer holds any explanatory power over E).

4. **Discrimination**: There is an additional nth means of detecting R for which H would strongly explain the detection of R by the nth means, R_n, and H' would strongly explain not detecting R by this means, $\neg R_n$. Formally, $\epsilon(R_n, H \mid E) \approx 1$, $\epsilon(\neg R_n, H' \mid E) \approx 1$.

5. **Success**: We learn that R_n, i.e. the nth means also detects R. (Shupbach, 2018)[pp. 12-15]

Note that Shupbach defines the explanatory power ϵ as

$$\epsilon(e, h) = \frac{Pr(h \mid e) - Pr(h \mid \neg e)}{Pr(h \mid e) + Pr(h \mid \neg e)} \tag{21}$$

This means that the explanatory power is a function with range $[-1, 1]$, where the closer to 1 it takes, the more powerful potential explanation of e is offered by h.[14] Conversely, the closer to -1 the function takes, the less powerful is the potential explanation e of h. And if $\epsilon(e, h) = 0$, then h is explanatorily irrelevant to e. Details of this measure of explanatory power can be found in (Shupbach and Sprenger, 2011). In their paper, Shupbach

[14]Depending on the reader's background, it may or may not be useful to point out here that these equations are not necessarily meant to be calculated as equations, giving a numerical value. Rather, they are formal logical relationships used to prove that Shupbach's account satisfies Bayesian epistemology.

and Sprenger develop this functional relationship from a few adequacy conditions for explanatory power. However, these details are not important for our purpose. What is important to note, is that the measure of explanatory power has to do with a hypothesis' ability to decrease the degree to which we find the fact that is to be explained surprising.

In addition to defining some of the steps described above in much more detail, Shupbach goes on to show how the successful completion of these steps also necessarily leads to incremental confirmation, according to Bayesian principles. I will not spend any time outlining the steps of this proof, but simply say that I think Shupbach's argument holds, and point any interested reader to see the details in (Shupbach, 2018)[pp. 13-18].

In my opinion, Shupbach successfully identifies the way in which means of detection must be *relevantly* different. And because he does so, it is no longer necessary to have such a strong criterion of probabilistic independence, a criterion that perhaps does entail RA diversity, but in an excessively strong way. In this process he manages to perfectly describe the normative power of ERA, by showing that whenever various means of detections are able to rule out competitor hypotheses, we get incremental confirmation of a hypothesis. This means that if we are able to align the models of MMEs with Shupbach's ERA, we have justified exactly *why* the agreement of multiple models has epistemic power, thus finally confirming the intuition we have had all along, but have not been able to formally account for.

It is important to note that there are a few fundamental differences between Shupbach's account and the ones we have looked at above. First of all, the accounts above have focused on the criteria necessary for a detected result to be *accepted*, pace Levins' notion of "*truth* at the intersection of independent lies". Shupbach is shifting this focus to accumulating *incremental confirmation*, and what conditions are necessary to *increase our confidence* in a hypothesis. Furthermore, Shupbach's account is altogether hypothesis-based; rather than increasing the confirmation for a *result*, it is the detected result which increases the confirmation for a *hypothesis*. Secondly, it is not so relevant anymore whether means of detection are strongly diverse or independent in some absolute sense. What matters for RA-diversity is that the means (which may actually be quite similar in most respects) are different in exactly the sense required to rule out the competitors of the target hypothesis. These two differences make Shupbach's account very promising for multi-model ensembles.

4.2 Winsberg's Application of ERA to Climate Science

We have now seen some of the arguments provided by Shupbach for why probabilistic independent accounts of RA are insufficient. We have also seen how models in MMEs are not probabilistically independent, and therefore such RA accounts cannot explain why the agreement of models in MMEs should lead to increased confidence. However, we have also seen how Shupbach's ERA provides a promising way to relate different means of detection so that incremental confirmation follows. The exciting question is then whether it is possible to extend Shupbach's notion of ERA to climate science.

Shupbach's ERA has been the source of much discussion after the publication of the article in 2016, and this discussion has also reached its way into the debate of model robustness in climate science. One particular philosopher that has endorsed the notion of ERA and tried to establish its potential in confirmation of climate scientific hypotheses is Eric Winsberg. Let us therefore review Winsberg's application of Shupbach's framework to climate science.

One of the interesting features of Winsberg's application, is the fact that he takes Shupbach's claim that ERA can be applied to both experimental and model-based results. In fact, Winsberg argues that this lack of separation between model- and experimental evidence means that both results detected by models and experiments can be taken into the same robustness analysis. He states that "[r]ather than trying to show how RA is *part of* a complex epistemic landscape in which other sources of evidence also play a role, our aim will be to show how RA can offer a comprehensive picture of that complex landscape - of how all these sources of evidence work together." (2018, p.184) Winsberg thus endorses Shupbach's account, and further promotes it as a holistic approach to describe the overall confirmation process in climate science - not restricting it to climate models.

However, before applying this unified approach, Winsberg does attempt to look at ERA strictly in the context of climate models. (Winsberg, 2018)[section 12.3]. He considers specifically the climate hypothesis we have considered previously in this chapter, namely that of Equilibrium Climate Sensitivity (ECS) lying in the range between 2.1°C and 4.7°C, as found in IPCC's AR5, supported by the outputs of the CMIP5 multi-model ensemble. (M. Collins et al., 2013) Winsberg then argues that there are two alternative explanations for why the models agree on this. Firstly, that the ECS actually falls within the detected range. Secondly, that the ECS does not fall within this range, and that the detected range is a result of a common failure among the models, with an especially likely candidate of this failure being how the models treat cloud feedbacks.

In Shupbach's terminology then, Winsberg is putting these two hypotheses forward as epistemic competitors. However, he further argues that even if our models agree with past observations, it is difficult to conclude whether the ECS does fall in the projected range, or whether the models are getting the right results for the wrong reasons. Because the models do not isolate specific causal processes, the outputs agreeing with the data could be because of compensating errors, rather than their ability to correctly predict the behaviour of the climate system. These compensating errors could be cancelling each other out in this specific case, but will not continue to do so when we get to regimes for which we do not yet have data. Consequently, neither the target hypothesis nor the alternative hypothesis seems like a more plausible explanation. We can say that in Winsberg's view, it is not possible to use the ouput of our current models to discriminate between the two competitor hypotheses, and therefore it is also impossible to successfully perform ERA based purely on model results in this context.

Because of this, Winsberg quickly moves on to considering the cases where model agreement could be enough to explicate confirmation of a hypothesis. He convincingly argues that this is a complicated process, and points to the process of finding "Emergent Constraints" in climate science as a particularly powerful method of RA that has the potential of completely ruling out competitor hypotheses based purely on results detected by models. As mentioned above, he also points to the inclusion of experimental results can help in cases, such as with ECS, where this elimination of competitor hypothesis is problematic solely based on model agreement.

However, although I do not necessarily disagree with Winsberg's points, I think he too eagerly gives up the model-based ERA for the MME in his example. I think he is right to argue that ERA cannot be used to completely confirm our target hypothesis in the example of ECS. However, remember that on of the most significant consequences of ERA is exactly that it does not rely on complete confirmation, but can serve to increase our confidence in any case where it can be applied. By phrasing his competitor hypotheses in terms of correct vs. wrong, Winsberg effectively retracts to the question of full acceptance, because to eliminate the competitor hypothesis in this case, it would essentially be necessary to prove that the model output is correct. This, in turn, greatly limits the potential of ERA. Ironically, I will use Winsberg's own insight to reformulate the approach of ERA applied to MMEs, namely when he says that it is possible to make such an ensemble of models RA-diverse "without altering the ensemble by adjusting your hypothesis." (Winsberg, 2018)[p. 202] In the next chapter we will see exactly what difference such a reformulation can make.

5 Towards a Full Unification of ERA and Climate Science

In the last chapter, we saw that whilst independence-based accounts of robustness analysis fail to describe the multi-model ensembles in climate science, Shupbach's framework of Explanatory Robustness Analysis seemingly escapes some of the problems associated with this. We also saw how Winsberg argues how this framework can be applied, but in the case of the ECS hypothesis only by also considering experimental evidence. I do not disagree with Winsberg in thinking that the inclusion of experimental evidence can make the robustness even better. As Shupbach demonstrated, the inclusion of more means of detection with the capacity of eliminating competitor hypotheses will increase our confidence in a hypothesis correspondingly. However, I argue that this is in fact not necessary. I will demonstrate that ERA can be applied to show how model agreement by itself can serve to increase our confidence in the very same climate scientific hypothesis put forward by Winsberg. I will also discuss how ERA can potentially be used as a guiding principle to contruct better MMEs, and furthermore, how it can be applied to means of detection beyond models.

5.1 Revising Winsberg's application

In order to illustrate exactly where and why I part ways with Winsberg, I will step-by-step go through Shupbach's formal account of the application of ERA. This will also lay the groundwork for later placing specific models into it, ultimately showing exactly how it can be applied to real-world models (section 5.2) Let us therefore revisit Winsberg's hypotheses in light of the formal steps provided by Shupbach to initiate this discussion.

Remember, the first step of a successful increment of ERA was

1. **Past detections**: We have a result R that we have detected using $n-1$ different means. We let $E = R_1 \& R_2 \& ... \& R_{n-1}$.

It is easy to understand how this step works with an ensemble of models detecting a result supporting the hypothesis of a certain ECS. An ensemble starts with one model, and adds more and more models detecting a result within the range supporting the hypothesis. The result R is thus the projected range, and the conjunction E is the combined output of all models. Following this, we have the step of

2. **Success**: The target hypothesis H explains the detections in the conjunction E, but so does an alternative hypothesis H'.

It is also easy to see how Winsberg is defining the hypothesis of ECS falling within the projected range as the target hypothesis H, and the hypothesis that ECS does not fall within that range but that the detected result is an artifact of a systematic failure in cloud feedback as the alternative hypothesis H'. We then get to the step of

3. **Competition**: H and H' epistemically compete with each other with respect to E. Formally, we have that i) $Pr(H\&H') = 0$ or ii) $\epsilon(E, H \mid H') \leq 0$ (i.e. the probability of both H and H' being true is zero, or less strongly, given H', H no longer holds any explanatory power over E).

This is where I will begin to part ways with Winsberg. Winsberg has formulated his hypotheses in such a way that the condition $P(H\&H') = 0$ is true. Either the detected range of ECS is correct, or it is not correct and a consequence of systematic failure of the models. The next step should then be that of

4. **Discrimination**: There is an additional nth means of detecting R for which H would strongly explain the detection of R by the nth means, R_n, and H' would strongly explain not detecting R by this means, $\neg R_n$. Formally, $\epsilon(R_n, H \mid E) \approx 1$, $\epsilon(\neg R_n, H' \mid E) \approx 1$.

which leads to problems for Winsberg's hypotheses. It requires that a new means of detection somehow can better be explained by one hypothesis over the other. However, as Winsberg notes, it is not possible with our current models to decide whether the ECS is actually correct or if it is wrong and there is a systematic failure of the models' ability to model feedback mechanisms. Adding another model to our ensemble that detect the result R can therefore not help us discriminate between the two alternative hypotheses. This is where Winsberg turns away from pure model-based ERA to experimental and observational evidence to help this process. However, this is where I instead will reformulate the hypotheses.

Winsberg formulated his hypotheses about ECS as follows:

> 1. ECS actually falls in that range.
> 2. ECS does not fall in that range and the value predicted by the ensemble is an artifact of the systematic failure of all the models to accurately capture all of the feedbacks - with cloud feedbacks being an especially likely candidate.

where the first one is the target hypothesis H, and the second the competitor hypothesis H'. Because we cannot effectively eliminate H' (at least given our current (lack of) knowledge of feedback mechanisms), then we also cannot use ERA to confirm H. However, I propose to reformulate these hypotheses into

1. Models detect range of ECS because it is correct.
2. Models detect range of ECS as a result of the specific feedback parameters.

where the first one is my target hypothesis H, and the second the competitor hypothesis H'. My target hypothesis H is very similar to Winsberg's, but is just reformulated for ease of comparison with my competitor hypothesis H'. My competitor hypothesis on the other hand, significantly differs from Winsberg's competitor hypothesis, and is weakened in two distinct ways. Firstly, it no longer says anything about the detected ECS being wrong. Secondly, it is now referring to "specific feedback parameters" instead of "systematic failure of all the models to accurately capture all of the feedbacks".

With these changes, it is easy to see that the two hypotheses are not necessarily inconsistent in the strong sense of $Pr(H\&H') = 0$. It could be true that the detected range of ECS is correct *and* that it is detecting the range because of the specific feedback parameters. In other words, the two hypotheses are not ruling each other out by default. Instead, the second condition for epistemic competition given by Shupbach holds, namely that two hypotheses epistemically compete when only one of them is needed to explain the detected result.

It is worth to further elaborate exactly what it means for hypotheses to epistemically compete in this way, to understand how it works in our case. As Shupbach writes, "[p]otential explanations of some explanandum E often compete, despite being consistent, when any one of these suffices to do the explanatory work of the others. Once we have accepted one, the explanatory work in accounting for E is done and hence there is no remaining explanatory reason from E to accept the others." (Shupbach, 2018)[p. 14] He further elaborates using the example of Brownian motion. In this case, the potential explanations for Brownian motion competed with each other like this. If we are inclined to accept the molecular explanation of Brownian motion, then it seems futile to aditionally accept the vital force hypothesis as another explanation; it no longer holds any explanatory power over E. And similarly, upon accepting our hypothesis H that the models are detecting the range because it is correct, this means that it is no longer necessary to consider hypothesis H'. It is important to note that this does not oblige us to accept the target hypothesis in any increment of ERA, only that this is what it can look like for two alternative hypotheses to be epistemically competing.

So by reformulating the hypotheses about detected ECS into two competing hypotheses that are not mutually exclusive, I believe I have provided a more feasible framework for performing model-based ERA with climate models. Let us now proceed to the next steps of Shupbach's account to see if these can now be completed.

After identifying two suitable epistemically competing hypotheses, the next step is discrimination, which I argued was not possible in the case of Winsberg's hypotheses. The detection of R by another model cannot tell us whether the ECS is right, or if it is wrong and an artefact of the systematic failure of the models to capture the feedback mechanisms. However, adding another model detecting R *can* help us discriminate between my revised hypotheses. If a new model with different feedback parameters detects R, the fact that it is a result of the specific feedback parameters of the models in the ensemble so far can no longer explain the detection. However, if the new model does not detect R, this would be explained by H'. The step of discrimination is therefore fulfilled, which then brings us to the last step of

5. **Success**: We learn that R_n, i.e. the nth means also detects R.

This means that if the new model is shown to detect R, this has successfully increased our confidence in H.

I have therefore shown that unlike what Winsberg seems to argue, model-based ERA *can* increase our confidence in a target hypothesis in cases where its epistemic competitor is not necessarily completely inconsistent with it. The weakening of this condition allows us, like I have done, to phrase the competitor hypothesis in such a way that it can be explained or not explained by adding an additional means of detection, i.e. model, to the collection of models. This turned out to be seemingly fruitful for a successful application of ERA to MMEs.

The discussion so far has worked as a proof-of-concept of the successful application of ERA to a potential MME detecting ECS, with realistic competitor hypotheses and realistic ways the models can differ. However, albeit realistic, we still have not applied this to real models. Let us therefore now take our discussion out of this theoretical vacuum and bring it back to the ESMs actually used by the IPCC.

5.2 Re-visiting model-detected ECS

In order to apply ERA to real models used by the IPPC, let us turn to the very same example used by Winsberg, namely that of the detection of ECS in IPCC's AR5. As mentioned above, the output of the CMIP5 models detected the ECS to lie between 2.1 degreees °C and 4.7°C, and the hypothesis is that ECS actually lies within this range with high confidence. (Flato et al., 2013) The ensemble used for this detection consists of 30 models, and therefore it might not be surprising that each one of them will not be considered. Since we are looking at the hypothesis that the models could detect the wrong result because of how the feedback mechanisms are modelled, I will only focus on the models that have explicitly defined feedback parameters, and that have quite different treatment of these. A complete overview of the models and their feedback parameters can be found in table 9.5 (Flato et al., 2013)[p. 818].

For the reasons just stated, the models BNU-ESM, CCSM4 and MIROC5 were selected. In table 1 we can see an overview of the value of ECS detected by these models, and the values of the different feedback parameters fed into the model. Let us now attempt to apply Shupbach's formalism based on this information.

To prove the application of ERA to the ensemble of models detecting ECS it is possible to start from any conjunction of models in the ensemble. However, to work more elegantly with the description of Shupbach's steps given above, let us assume our starting point is the CMIP5 models listen in table 9.5 (Flato et al., 2013)[p. 818], minus the three models from table 1. That means that the ECS is already detected to be within the range of 2.1 and and 4.7 °C. This then, fulfills the first step of ERA, namely that we have a result R detected by $n-1$ different means.

Model	Equilibrium Climate Sensitivity	Planck Feedback	Water Vapour Feedback	Lapse Rate Feedback	Surface Albedo Feedback	Cloud Feedback
BNU-ESM	4.1	-3.1	1.4	-0.2	0.4	0.1
MIROC5	2.7	-3.2	1.7	-0.6	0.3	0.1
CCSM4	2.9	-3.2	1.5	-0.4	0.4	-0.4

Table 1: The equilibrium climate sensitivity (ECS) given in °C for the CMIP5 models BNU-ESM, MIROC5 and CCSM4. The corresponding values for the feedback parameters are also listed, all given in W m^{-2}°C^{-1}. The values are all taken from table 9.5 (Flato et al., 2013)[p. 818].

We can then move on to the second and third step of Shupbach's framework, which concern the target hypothesis and an alternative hypothesis that both potentially explain the detections, and that these hypotheses epistemically compete. Like before, we will define our target hypothesis H as the hypothesis that models detect the result because ECS actually lies within this range, and similarly we will again define the alternative hypothesis as the models detecting this result because of the specific treatment of feedback mechanisms. The hypotheses epistemically compete not in the sense of $Pr(H\&H') = 0$, but in the sense that given H, H' no longer holds any explanatory power over R. If the models detect R because R is correct, then H is doing the explanatory work of H', upon accepting H it is no longer necessary to look for alternative hypotheses.

So we have a $n-1$ past detections detecting the result R that ECS within the range 2.1 and and 4.7 °C. The conjunction E denotes $R_1\&R_2\&...\&R_{n-1}$, which we will define to be the results detected by all models in the ensemble not stated here plus the first model in table 1, the BNU-ESM model. We have two hypotheses that can both potentially explain E, and these hypotheses epistemically compete. The next step in Shupbach's framework is discrimination, where an n'th means potentially detecting R would be strongly explained by H but not H'. Let us say that this n'th means of detection is the MIROC5 model. We can easily see that this model differs from the previous means of detection, the BNU-ESM model, in its treatment of lapse rate and water vapour feedback. Similarly, it differs from the other models in the ensemble, as can be seen in table 9.5 (Flato et al., 2013)[p. 818]. Detecting R by this n'th means of detection would then strongly suggest that E is not a result of the specific feedback parameters, and that instead, the target hypothesis is correct.

As we can see in the table 1 the MIROC5 model detected an ECS within the range of the conjunction E, meaning that we can add the successful detection R_n to the conjunction, and that this then, completes an increment of ERA, meaning our confidence in H is increased.

Now let us repeat the process for the CCSM4 model. For consistency, let us stick to the same alternative hypotheses as in the last increment, although if we look to table 1, we can see that it could be possible to formulate a more specific alternative hypothesis with respect to cloud feedback, as this is the main difference between the models. Also for consistency, let us call the CCSM4 model the n'th+1 means of detection, that can be used to discriminate between our two hypotheses. Just like before, detecting R by the CCSM4 model would strongly suggest that E is not the result of the specific feedback parameters, and strongly explain our target hypothesis

H. Once again, from table 1 we can see that CCSM4 detects an ECS within our hypothesis range, meaning that another increment of ERA is successfully completed, and that our confidence in H should be increased correspondingly.

I will argue that my choice of models for this proof-of-concept was arbitrary, and that it is not difficult to imagine that the pattern I have just described can be extended to the other models in this particular ensemble. I will therefore not repeat this process for more models, but simply claim that with a suitable formulation of alternative hypotheses, adding new models to any MME can increase our confidence in a hypothesis based on detections from previous models.

Lastly, it is worth to once again point to the importance I have placed on rephrasing the hypotheses to make it possible to discriminate between them using the models of the MMEs. Like Winsberg points out, because of our current lack of knowledge about the feedback systems, it is impossible to completely rule out the hypothesis that the models are wrongly representing these. However, as I have shown, by weakening Winsberg's hypothesis, it is possible to apply Shupbach's framework to step-by-step, model-by-model, show how specific conditions in the models cannot explain the detected results, consequently increasing our confidence in the target hypothesis. This can be likened to Shupbach's example of detecting Brownian motion. Here the scientists systematically changed the experimental set-up to find that the detected result did not in fact depend on specific conditions. This is exactly what the climate scientists are doing by using MMEs. They cannot prove that their causal processes are modelled "correctly", however, they can systematically increase our confidence in the target hypothesis by varying their "experimental" set-up. In the example of Brownian motion the scientists could not initially pin down a single explanation, but they could systematically rule out variables the phenomenon was dependent upon, each time an alternative hypothesis was ruled out. This is analogous to the way climate scientists use MMEs to rule out similar hypotheses about the results detected from the models.

From this we finally have an answer to the questions we started out with: Why should model agreement give us more confidence in a hypothesis about the climate? And under what conditions can it do so? Model agreement is an epistemologically sound method to increase our confidence in a hypothesis because it eliminates alternative hypotheses, which as Shupbach shows should incrementally increase our confidence in the target hypothesis. Furthermore, it will be so in any case where two epistemically competing hypotheses can be differed between by another means of detection, in our case climate models.

5.3 A Contemplation On the Potential of ERA Applied to MMEs

I want to stop and contemplate on the application of ERA to MMEs, and once again try to make us feel the full force of what I believe is its potential. We started out with some vague idea that more pieces of evidence leads to increased confidence, and a very complex scientific discipline where models are added to big ensembles, and scientists trying to justify this process as a way of increasing confidence. But why should it do exactly this? The models are often similar, meaning they do not add independent bits of information, they are inconsistent in their assumptions, the MMEs consists of a somewhat random collection of models. It is therefore not evident, why this model agreement should have such a heavy epistemological weight in the climate scientific practice. Well, it should because the MMEs are able to, model-by-model, rule out alternative explanations of the hypothesis supported by their commonly detected result. In this chapter, I have showed that this is true formally. Ergo, ERA is able to explain and justify the use of model agreement in MMEs to support hypotheses about the climate.

This means that the process of “model agreement” in order to increase confirmation for hypotheses about the climate is epistemologically sound, and can be applied in any case where the models can be defined as epistemic competitors, no matter how similar or how different the models are. I think this is a massively significant result, one that I think Winsberg touched upon, but did not elaborate enough. However, it still leaves us with a major challenge. When does this incremental confirmation lead to full confirmation of a hypothesis? By shifting the focus from obtaining full confirmation to incrementally increasing confidence, it seems like we were able to delay this question for a while, all whilst building a normative epistemological framework for model agreement. However, the question still remains, can model-based ERA ever lead to full confirmation of a hypothesis? And, perhaps more interestingly, can our current MMEs provide such confirmation?

As Winsberg pointed out with the ECS hypothesis, the current multi-model ensembles supporting it are not able to completely rule out competing explanations. And if it is not possible to rule out competing explanations, one might think that we are better equipped at supporting our hypotheses than we were in the beginning (despite of establishing the epistemic power of robustness). At first glance then, it might seem that we have done a lot of work just to arrive at similar problems to the ones we out started with.

However, we are now in a completely different situation to answering them. Realising exactly what can increase our confidence in a hypothesis, and exactly how the models can work together to differ between hypotheses, we are also in a situation where we can use this knowledge to make our MMEs as effective as possible.

Constructing more efficient MMEs is a matter Reto Knutti, a climate scientist with a keen interest in the philosophy of science, and that has co-authored several of the IPCC ARs cited in this thesis, has advocated for. (Knutti, 2010; Knutti, 2018) He argues that the selection of models that go into an MME is neither random nor systematic. Instead, the modelling groups contribute with their "best" model, which in many cases can mean the models that give a similar output to those of other modelling groups. As Knutti notes, "it is easier to be in the middle of the crowd than far outside" (Knutti, 2010)[p. 397], meaning that no modelling group will try to push their model into extreme behaviour. Effectively, one can say that the modelling groups focus solely on the performance of their own model, rather than how it will work together with the others in the ensemble. Furthermore, Knutti argues that because of this in many respects similar behaviour of the models, the current MMEs effectively contain too many models without much added value. This is perhaps not a problem by itself, but in light of the limited amount of resources we have for running these immensely complex models, it certainly is. The question is then how many and which models to include in an MME?

As Knutti also points out, it is not necessarily so easy to understand what exactly is the "best" model. As emphasized elsewhere in this thesis, it is problematic to compare models directly with data, as data is lacking both for the future climate, and to some extent also the past. Moreover, it is difficult to establish appropriate performance metrics, because metrics applicable to one specific purpose might not be so relevant in others. Not to mention that a good fit with existent data not necessarily equates to a high-quality model. (Knutti, 2018) Rather, it could suggest an over-dependence on tuning, i.e. calibrating the model to fit with certain data, or, as discussed before, compensating errors in the model giving output consistent with data by "accident".

Perhaps then, ERA can act as a guiding principle for the selection of models that go into an MME. If adding a model to an ensemble provides the scientists with a way to eliminate an alternative hypothesis, then it should be included in the ensemble. However, if on the other side, a model cannot provide this potential elimination, or if it can only eliminate a hypothesis already discarded or thought to be unlikely by the scientists, then one might think twice about including it. This adds value to constructing a model that might not produce such similar outputs to the ones of the other modelling groups, as pointed out as a systematic weakness by Knutti above. Effectively it shifts the focus from producing the "best" model, which is problematic, to producing the best model ensemble.

It is important to note the specification that a model which cannot be used to eliminate a competitor hypothesis, *or* "can only eliminate a hypothesis discarded or though to be unlikely by the scientists" above. This specification is necessary because my application of ERA to multi-model ensembles potentially opens up for the possibility that *any* model that can potentially eliminate *any* competitor hypothesis will increase confidence in our target hypothesis. However, it is of course unfeasible and inefficient to make an ensemble that eliminates *all* potential competitor hypothesis, even if we had the computational power we might not want to do this. As Lloyd points out, parameterizations and simplifications of the models in an MME are based on different lines of evidence, including experimental and observational evidence. (Lloyd, 2015) And they should be. Of course not every competitor hypothesis is likely, and should be selected based on physically realistic explanations based on the scientific knowledge we already have.

It is therefore contended that a selection of models for MMEs based on their potential to eliminate possible explanations constrained by the physically viable options, could be an appropriate and effective way to construct MMEs. However, exploring this possibility and the details of it further is beyond the current scope.

5.4 Going Beyond Model-Based ERA in Climate Science: Revisiting Detection and Attribution Studies

So far, I have focused only on model-based ERA. This has allowed us to pin down exactly why, and in what situations, model agreement should increase our confidence in a hypothesis about the climate. However, as Winsberg argues, Shupbach's breaking down of the boundaries between model-based and experimental-based robustness analysis allows us to include both model results and experimental evidence in the same robustness analysis. In this section I will point to some examples of this in climate science, and argue that this is a particularly characteristic type of reasoning, which, once again, Shupbach's framework can help us understand.

Firstly, let us revisit the example of detection and attribution studies that were discussed in section 3.1. This is perhaps a particularly obvious example of ERA. Recall that detection and attribution studies consists of two separate stages, the first one detecting a climatic change, and the second one attributing this to certain causal factors. As I will argue, there are clear examples of ERA happening on many levels in a process of detection and attribution.

Firstly, as discussed in section 3.1, the process of detection involves detecting a change, and making sure it is not only a result of internal variability. This is because a possible explanation for any detected change is that it is a result of the variability of the climate system. It is easy to understand how this can be phrased in terms of Shupbach's framework, by defining the hypothesis that the change is a result of factors external to the variability of the system, and its epistemic competitor hypothesis as the hypothesis that it is a result of the natural variability. As mentioned before, usually model simulations with external conditions held constant are used to simulate natural variability, and rule out the possibility that the climatic change is a result of this until the "likelihood of occurrence by chance due to internal variability alone is determined to be small". So once again we see an example of ERA with models being used to distinguish between the alternative hypotheses.

However, detecting a change in a climatic variable and using models to rule out the possibility of it being a result of internal variability does not exhaust the process of detection. As illustrated in figure 3 various means of detection are used to establish the temperature rise of the past centuries. Because even if the detected temperature rise of one observable is established as significant, there is still the possibility that the data set of the most immediate measure of temperature, i.e. surface temperature, is faulty or biased in some way. In order to eliminate this possibility, the scientists compare with observations of another climatic variable, for example the ocean heat content, which measures the heat contained in the oceans. If the temperature really is rising, then we should expect this to rise as well. As we can see in figure 3, this is in fact the case, which increases our confidence in the target hypothesis. However, scientists are skeptical creatures, and in the figure we can see how more and more means of detection are added to eliminate the possibility of alternative explanations such as systematic failures, biased data sets, and missing causal connections. Every separate data set detecting rising temperature increases the confidence in the detected change being actual. Again, this well described by Shupbach's framework, meaning Shupbach's framework can be used to explain why this type of reasoning is epistemologically sound.

Attribution starts where the process of detection ends, and aims to map the causes of the climatic change. For the detected temperature change over the last centuries, climate scientists put forward two possible explanations, the first one that the temperature rise is a result of natural forcings; the second one that is a result of anthropogenic forcings. These can be said to be yet another school-book example of Shupbach's epistemically competing hypotheses. Either the detected temperature rise can be explained mostly by natural factors, or by anthropogenic influence. As mentioned in section 3.1 the observed detection has to be shown to be inconsistent with the alternative explanation, meaning that the competitor hypothesis has to be fully eliminated. Attribution studies then consists of different ways to try to eliminate one of these competitor hypotheses.

The main way in which these competitor hypotheses are distinguished between is to model the climate for the past centuries including anthropogenic forcings, and including natural forcings only, and then comparing them to the observations of the detected changes.

Ergo, attribution and detection studies then are a prime example of how climate scientists use a process of distinction and elimination of alternative hypotheses to increase our confidence in hypotheses about the climate system, and consequently fit into Shupbach's framework nicely. They also illustrate how diverse ERA in climate science can be, varying between models and experimental evidence to perform this elimination. Furthermore, we can see how this process appear in a rather nested fashion, where multiple layers of ERA can appear in the process of increasing confidence in a single hypothesis. We can then understand better what Winsberg meant when he said that "[r]ather than trying to show how RA is part of a complex epistemic landscape in which other sources of evidence also play a role, our aim will be to show how RA can offer a comprehensive picture of that complex landscape - of how all these sources of evidence work together." The detection and attribution studies in this section shows us exactly what this complex epistemic landscape can look like.

Recall that the IPCC relies on "type, amount, quality and consistency of evidence (e.g. mechanistic understanding, theory, data, models, expert judgment) (Mastrandrea et al., 2010) [p.1], and the degree of agreement" to support confidence in a hypothesis about the climate. I have so far discussed the significance of model agreement, and also touched upon how data is also used to support hypotheses. It would be interesting to try to understand further how this epistemic landscape works, and how these other types of evidence work together. From the work done in this thesis is seems plausible that this too can be placed within Shupbach's framework as a way to understand its epistemic significance. However, this is again beyond the scope of this thesis, but could be an interesting angle to pursue in the future.

6 Conclusions

In this thesis, I have looked at the construction, use, and epistemological role of climate models in the climate scientific practice. This was done by looking at climate models, their application in the climate scientific practice, and how the practice of model agreement can be justified by using the framework of ERA.

Chapter 2 was dedicated to developing an understanding of the components of the climate system, and the methods for modelling these. It became evident that the climate system is a complex system, and that consequently, the models representing this system are also complex. Due to our inability to perfectly represent this system, idealisations and simplifications are necessary, and there are therefore uncertainties related to how well these models can represent the climate system.

The focus of chapter 3 was how the main scientific body for climate science, namely the IPCC, uses models in establishing hypotheses about the climate. I pointed to two main uses of climate models, namely detection and attribution studies and projections about the future climate. From this it became apparent that models are heavily used to form hypotheses and to build confidence in these hypotheses. However, some obvious epistemological questions also arose from this. If there are so many inaccuracies related to the models, exactly how do they give us such high confidence in their outputs? The IPPC's use of multi-model ensembles also became apparent, with hypotheses relying on the common output of multiple models. This led us to two questions: How and if this model agreement should be given such a significant epistemological weight; and exactly how they work together to increase our confidence in hypotheses about the climate.

In chapter 4, robustness analysis was highlighted as a way to answer questions such as these. Robustness analysis aims to describe how pieces of evidence can work together to give certain knowledge in cases where the individual pieces might have intrinsic shortcomings. I argued that the more traditional notion of robustness analysis that is based on independence between these pieces of evidence is not feasible. Instead, I looked at Shupbach's explanatory robustness analysis as an alternative to this, where increased confidence is achieved in any case where the pieces of evidence, be it models or experimental evidence, work together in a way that allows us to eliminate alternative explanations to the detected phenomenon.

The final chapter was dedicated to revising Winsberg's application by rephrasing the hypotheses about Equilibrium Climate Sensitivity. It was then shown that, unlike what Winsberg argues, ERA is able to show how model agreement in this case is confirmatory. Shupbach's framework was also applied, step-by-step, to three example models from the CMIP5 that are used to support the ECS hypothesis. This establishes the practice of robustness among models in an MME as epistemologically sound. After focusing solely on ERA in terms of model agreement, detection and attribution studies were once again revisited. From this we could spot the nested way in which multiple layers of elimination and discrimination can work with means of detection that consist of both models and observations. This shows that Shupbach's type of reasoning can be said to be highly characteristic for the climate scientific practice, and also how it is able to account for the various types of evidence involved in making hypotheses about the climate.

I believe my application of Shupbach's framework to the climate scientific practice is valuable for two main reasons. Firstly, because it allows us to delegate a normative epistemological power to the so-far ambiguous practice of "model agreement". For many reasons, confirming hypotheses about the climate system is a complicated process, a highly contextual process, and a process that might be very difficult to understand for anyone outside the discipline. However, I believe that it is still possible to highlight and describe the mechanisms at play in this process, and therefore to understand why such confidence should be given to the knowledge produced. I further believe that this is not only interesting, but indispensable for a science such as climate science, which is so highly politicised and its knowledge so defining for the future of us all.

Secondly, I believe this description and application of ERA is valuable because by exactly defining the factors at play, it is also possible to start optimizing the process. Like some climate scientists have pointed out, multi-model ensembles are sometimes constructed somewhat randomly, meaning that their potential is not fulfilled. I believe that the framework of ERA can be one such method that could help maximising the use of MMEs, and this possibility was briefly discussed. Because in the end, a more directed use of the incredible climate models we have at our disposal could only serve to further improve our knowledge and understanding of the climate system.

References

Arguez, A., & Vose, R. (2011). The Definition of the Standard WMO Climate Normal. *Bulletin of the American Meterological Society*, *92*(6), 699–704.

Baum, C. M., Low, S., & Sovacool, B. K. (2022). Between the sun and us: Expert perceptions on the innovation, policy, and deep uncertainties of space-based solar geoengineering. *Renewable and Sustainable Energy Reviews*, *158*.

Bender, F., Hegerls, G. C., Flynn, C., & et al. (2022). The value of values in climate science. *Nature Climate Change*, *12*(4-6).

Bindoff, N., Stott, P., AchutaRao, K., Allen, M., Gillett, N., Gutzler, D., K., H., Hegerl, G., Hu, Y., Jain, S., Mokhov, J., I.I. Overland, Perlwitz, J., Sebbari, R., & Zhang, X. (2013). Detection and Attribution of Climate Change: From global to regional. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Bradley, R., & Steele, K. (2015). Making Climate Decisions. *Philosophy Compass*, *10*(11), 799–810.

Campsiano, C. J. (2012). Milankovitch Cycles, Paleoclimatic Change, and Hominin Evolution. *Nature Education Knowledge*, *5*(4(3)).

Chen, D., Rojas, M., Samset, B., Cobb, K., Diongue Niang, A., Edwards, P., Emori, S., Faria, S., Hawkins, E., Hope, P., Huybrechts, P., Meinshausen, M., Mustafa, S., Plattner, G.-K., & Tréguier, A.-M. (2021). Framing, Context and Methods. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

Collins, M., Knutti, R., Arblaster, J., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, R., Krinner, G., Shonqwe, M., Tebaldi, C., Weaver, A., & Wehner, M. (2013). Long-term Climate Change: Projections, Commitments and Irreversibility. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Collins, S. N., James, R. S., Ray, P., Chen, K., Lassman, A., & Brownlee, J. (2013). Grids in numerical weather and climate models, 111–128.

Copernicus. (2023). Copernicus: 2022 was a year of climate extremes, with record high temperatures and rising concentrations of greenhouse gases [Accessed: 02.04.2023].

Cubasch, U., Wuebbles, D., CHen, D., Facchini, M. C., Frame, D., Mahowald, N., & Winther, J.-G. (2013). Introduction. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Dragutin, T. M. (2020). Chaos in modelling the global climate system: Reminiscences and some current meditations. *Curr Adv Geo Geosci*, *1*, 1–10.

El amrani, M., & Seaid, M. (2008). Eulerian-lagrangian time-stepping methods for convection-dominated problems. *International Journal of Computer Mathematics*, *85*, 421–439.

Flato, G., J., M., 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., & Rummukainen, M. (2013). Evaluation of Climate Models. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., D.J., L., Mauritsen, T., Palmer, M., Watanabe, M., Wild, M., & H., Z. (2021). The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

Gueguen, M. (2020). On Robustness in Cosmological Simulations. *Philosophy of Science - Proceedings of the 2018 biennial meeting of the Philosophy of Science Association*, *87*(5), 1197–1208.

Hartmann, D., Klein Tank, A., Rusticucci, M., Alexander, L., Brönnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., & Zhai, P. (2013). Observations: Atmosphere and Surface. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Hegerls, G. C., Hoegh-Guldberg, O., Cassasa, G., Hoerling, M., Kovars, S., Parmesan, C., Pierce, D., & Stott, P. (2009). Good Practice Guidance Paper on Detection and Attribution Related to Anthropogenic Climate Change. *Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Detection and Attribution of Anthropogenic Climate Change*.

Idso, S. B. (1988). The CO2 greenhouse effect on Mars, Earth, and Venus. *Science of the Total Environment*, *77*(2-3), 291–294.

Kim, J., Bates, D., Postlethwaite, I., Ma, L., & Iglesias, P. (2006). Robustness analysis of biochemical network models. *IEEE Proceedings - Systems Biology*, *153*(3), 96–104.

Kirtman, B., Power, S., Adedoyin, J., Boer, G., Bojariu, R., Camilloni, I., Doblas-Reyes, F., Fiore, A., Kimoto, M., Meehl, G., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G., Vecchi, G., & Wang, H. (2013). Near-term Climate Change: Projections and Predictability. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Knutson, T., Zeng, F., & Wittenberg, A. (2013). Multimodel Assessment of Regional Surface Temperature Trends: CMIP3 and CMIP5 Twentieth-Century Simulations. *Journal of Climate*, *26*, 8709–8743.

Knutti, R. (2010). The end of model democracy? *Climatic Change*, *102*, 395–404.

Knutti, R. (2018). Climate model confirmation: From philosophy to predicting climate in the real world, 325–359.

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic Modelling as Robustness Analysis. *British Journal of Philosophy of Science*, *61*(3), 541–567.

Latour, B. (2017). *Facing Gaia: Eight lectures on the new climatic regime.* Translated by Catherine Porter. UK: Polity Press.

Le Treut, H., Somerville, R., Cubasch, U., Ding, Y., Mauritzen, C., Mokssit, A., Peterson, T., & Prather, M. (2007). Historical Overview of Climate Change. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*.

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., Engelbrecht, F., Fischer, E., Fyfe, J., Jones, C., Maycock, A., Mutemi, J., Ndiaye, O., Panickal, S., & Zhou, T. (2021). Future Global Climate: Scenario-based Projections and Near-term Information. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, *54*(4).

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.

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K., Frame, D., Held, H., Kriegler, E., Mach, K., Matschoss, P., Plattner, G.-K., Yohe, G., & Zwiers, F. (2010). Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties.

McGuffie, K., & Henderson-Sellers, A. (2005). *A climate modelling primer* (3rd ed.). Chichester, West Sussex ; Hoboken, N.J. : J. Wiley.

Meehl, G., Stocker, T., Collins, W., Friedlingstein, P., Gaye, A., Gregory, J., Kitoh, A., Knutti, R., Murphy, J., Noda, A., Raper, S., Watterson, I., Weaver, A., & Zhao, Z.-C. (2007). Global Climate Projections. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.*

North, G. R., & Stevens, M. J. (2006). Energy-balance climate models, 52–72.

O'Neill, B., Tebaldi, C., van Vuuren, D., Eyring, V., Friedlinstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G., Moss, R., Riahi, K., & Sanderson, B. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*, 3461–3482.

Orzack, S., & Sober, E. (1993). A critical assessment of levins's "the strategy of model building in population biology. *The Quarterly Review of Biology*, *68*(4), 533–546.

Parker, W. S. (2011). When Climate Models Agree: The Significance of Robust Model Predictions. *Philosophy of Science*, *78*(4), 579–600.

Peiro, J., & Sherwin, S. J. (2005). Finite difference, finite element, and finite volume method. *Handbook of Material Modeling*, *1*, 421–439.

Perrin, J. (1913). *Les atomes*. Ox Bow Press. Translated by D. L. Hammick (1990).

Pirtle, Z., Meyer, R., & Hamilton, A. (2010). What Does It Mean When Climate Models Agree? a Case for Assessing Independence among General Circulation Models. *Environmental Science and Policy*, *13*, 351–361.

Sasamori, T., & Melgarejo, J. (1978). A Parameterization of large-scale heat transport in mid-latitudes. Part I. Transient eddies. *Tellus*, *30*, 289–299.

Shupbach, J. (2018). Robustness Analysis as Explanatory Reasoning. *British Journal of Philosophy of Science*, *69*(1), 275–300.

Shupbach, J., & Sprenger, J. (2011). The Logic of Explanatory Power. *Philosophy of Science*, *78*(1), 105–127.

Solanki, S., & Krivova, A. (2003). Can solar variability explain global warming since 1970? *Journal of Geophysical Research*, *108*(A5).

Weisberg, M., & Reisman, K. (2008). The Robust Volterra Principle. *Philosophy of Science*, *75*(1), 106–131.

Winsberg, E. (2018). *Philosophy and climate change*. Cambridge University Press.

Wüthrich, N. (2017). Conceptualizing Uncertainty: An Assessment of the Uncertainty Framework of the Intergovernmental Panel on Climate Change. *EPSA 2015 Selected Papers: The 5th conference of the European Philosophy of Science Association in Düsseldorf*.