Modelling the earth's climate - an epistemic perspective

Inauguraldissertation

zur Erlangung des Grades eines
Doktors der Philosophie

am Fachbereich für Philosophie und Geisteswissenschaften
der Freien Universität Berlin

vorgelegt von
Friederike Elly Luise Otto

Berlin 2012

Erstgutachter: Prof. Dr. Holm Tetens
Zweitgutachter: Prof. Dr. Anders Levermann

Tag der Disputation: 23. November 2011

Abstract

Climate models and climate modelling are a central part of climate science with particular importance for long term prognoses of future climate development. In the context of global climate change, which is a fact undoubted by climate scientists but sceptically discussed in the public, their importance not only for climate sciences is increasing. However, most findings of climate modelling approaches are highly uncertain and span a very broad range of values for climate variables and impacts of climate change.
Climate models are different from experiments in physics, thus their results must be valued accordingly. Furthermore climate models are epistemically very different from physics theories, which are normally topic of the debate in the philosophy of science. In this thesis climate modelling is analysed according to ascertain the epstemic status of climate models and to discuss its consequences.

Climate modelling is not based on a comprehensive physics theory and is not analogous to experimenting. Moreover, climate models play a double role as an outsourced human brain and a copy of the earth and are thus something in between an experiment and a theory in progress.
Due to this fact several problems of climate modelling result, two of which are fundamental and others are principally to overcome but practically pressing. The fundamental problems of understanding the climate system are the nonlinearity of the system and lack of observational data. The main practical problem of climate modelling is the problem of parameterisation, which is the need to represent processes of the climate system in the modelling approach that are insufficiently understood or on a smaller scale than the resolution of the model.
Parameterisations in nonlinear models make it nearly impossible to detect chains of causes and effects in a climate model. Therefore an intransparent method of fitting the model to data, which is called tuning, results in manipulated physics of the climate model and prevents a meaningful analysis of the modelling results.
As a conclusion of this thesis certain rules are provided that could avoid abuse of climate model tuning. Furthermore basic guidelines are provided to make the climate modelling process more transparent in general and thus to refer to the main uncertainties integral to climate modelling appropriately.

Zusammenfassung

Die Klimawissenschaften erfahren in den letzten Jahren große öffentliche Aufmerksamkeit aufgrund des sich zunehmend bemerkbar machenden anthropogenen Klimawandels. Die Rezeption der Ergebnisse klimawissenschaftlicher Forschung ist dabei keineswegs nur positiv. Im Gegenteil, viele Aussagen, die Wissenschaftler über die mögliche Entwicklung des Klimas, treffen, werden angezweifelt. Und zwar sowohl von Wissenschaftlern anderer Disziplinen als auch von Vertretern politischer und gesellschaftlicher Instanzen. Wenn diese Zweifel begründet werden, beruht diese Begründung häufig auf der Annahme, dass Klimamodelle, mit deren Hilfe Projektionen zukünftiger Klimate erstellt werden, nicht die nötige Qualität aufweisen um ihre Aufgabe zu erfüllen.
In dieser Dissertation wird der wissenschaftstheoretische Status insbesondere komplexer Klimamodelle analysiert und die Frage erörtert, ob Zweifel berechtigt sind.

Der anthropogene Klimawandel ist eine wissenschaftliche Tatsache, die ohne Zuhilfenahme von Klimamodellen zu belegen ist. Für die Folgen der globalen erwärmung spielen diese jedoch eine herausragende Rolle. Klimamodelle sind aus wissenschaftstheoretischer Sicht grundlegend verschieden von wissenschaftlichen Theorien, die im wesentlichen Gegenstand der Diskussion in der Wissenschaftstheorie sind. Einge klassische Fragen dieser Disziplin stellen sich daher anders bzw. bedürfen anderer Ideen zur Beantwortung der Fragen.
Nach einer naturwissenschaftlichen Einführung wird in der Dissertation gezeigt, dass es keine Theorie der Klimawissenschaften gibt, ebenso wie Klimamodelle nicht analog zu klassischen Experimenten verstanden werden können. Klimamodelle, wie andere Computermodelle auch, nehmen stattdessen einen Status zwischen Theorie und Experiment ein und sind eher als Ansatz ein bestimmtes Problem zu bearbeiten zu interpretieren, als dessen tatsächliche Lösung.
Daran anschließend werden die Probleme im Zusammenhang mit Klimamodelierung dargestellt, wobei zwei grundsätzlich nicht lösbare, prinzipielle Probleme einer Reihe von Modellierungsschwierigkeiten gegenüberstehen. Eine der Hauptursachen für letztere ist ein Skalenproblem, da wichtige Prozesse im Klimasystem auf räumlichen Skalen stattfinden, die in den Modellen nicht aufgelöst werden müssen sie parametrisiert werden. Viele Prozesse des Klimasystems sind bisher nicht, oder nur unzureichend verstanden, ein Problem, dass durch die Parametrisierungen verstärkt wird. Ein prinzipielles Problem ist die Nichtlinearität des Klimasystems, die es einerseits nicht möglich macht das System komplett zu verstehen und andererseits nichtlineare Modelle erfodert, in denen es kaum möglich ist Kausalketten zu identifizieren.
Abgesehen von diesen systembedingten Schwierigkeiten erschwert die Tatsache, dass es sich beim anthropogenen Klimawandel um ein singuläres Ereigniss der Klimageschichte handelt, das Testen von Klimamodellen und erhöht damit die Unsicherheit der Klimaprojektionen.
Dass es trotzdem möglich ist Klimamodelle zu validieren und damit zu robusten Ergebnissen der Modellsimulationen zu gelangen ist Gegenstand der weiteren Analysen in der Dissertation. Dabei wird dargelegt, wie mit Hilfe von Klimamodellen insbesondere

durch das modellierern ähnlicher Szenarien in ganzen Ensemblen von Modellen gute Ergebnisse erzielt werden können. Diese Methoden erlauben es jedoch streng genommen nicht, modellierte Klimavariablen so exakt zu prognistizieren, dass Wahrscheinlichkeitsfunktionenen angegeben werden können.
Anhand des Beispiels des Modelltunens werden im lezten Teil der Dissertation Regeln entwickelt, deren Einhaltung einige grundlegende Fehler im Modellierungsprozess verhindern kann. Die konkreten technischen Regeln lassen sich auf drei wesentliche Grundsätze reduzieren: Ein Messdatum nicht zweimal zu verwenden, alle bekannten theoretischen Zusammenhänge im Modellierungsprozess zu berücksichtigen und keine Zusammenhänge, die durch das Tunen des Modells enstehen als Kausalzusammenhänge zu interpretieren. Darüberhinaus werden Vorschläge zum Umgang mit Unsicherheiten im Modell erörtert und anhand von Handlungsanweisungen des Weltklimarats ergänzt. Begrüdet werden diese Regeln damit, dass sie zur Erreichung des Ziels, Modellierungsprozesse und Ergebnisse von Klimasimulationen nachvollziehbarer und transparenter zu machen, beitragen.

Contents

Preface 3

I. Introduction and Background 5

1. Introduction 7
1.1. The debate . . . 9

2. Basic knowledge 15
2.1. Short introduction to climate science . . . 15
2.1.1. Thermodynamics . . . 16
2.1.2. Dynamics . . . 17
2.1.3. The primitive equations . . . 18
2.2. Basic knowledge of climate change . . . 19
2.3. Defining *climate model* . . . 25
2.3.1. Theory of the climate system . . . 25
2.3.2. Climate modelling . . . 26
2.4. Spectrum of climate models . . . 34

II. Epistemic problems of climate modelling 39

3. Perfect climate modelling is impossible 41
3.1. Fundamental uncertainty versus model uncertainty . . . 43
3.2. Fundamental uncertainty . . . 44
3.2.1. Fundamental constraints in understanding the climate system: nonlinearity . . . 44
3.2.2. Fundamental constraints in understanding the climate system: observational constraints . . . 47
3.3. Boon and bane of parameterisation . . . 48
3.4. Using and abusing model tuning . . . 52
3.5. The need for numerical solutions . . . 56
3.6. Theory ladenness of climate models . . . 58
3.6.1. Reanalysis data . . . 61
3.6.2. Paleo data . . . 62
3.7. The impossibility of falsifying climate models . . . 62

3.8. The singularity of anthropogenic climate change . . . 64
3.8.1. Prediction in climate science . . . 66

III. Handling of Problems **73**

4. Validation in natural sciences **75**
4.1. Testing in natural science . . . 76
4.1.1. Climate models as experiments . . . 78
4.1.2. Successful predictions . . . 79

5. Model Testing **83**
5.1. Model testing . . . 83
5.1.1. Testing in general . . . 83
5.1.2. Paleo data . . . 87
5.2. Model comparison . . . 88
5.2.1. Intramodel comparison . . . 89
5.3. Model intercomparison . . . 91

6. Probability in climate modelling **95**
6.1. Reasons for applying probability . . . 95
6.1.1. Deterministic vs. stochastic modelling . . . 96
6.1.2. Uncertainty . . . 99
6.2. Statistical modelling . . . 100
6.3. Ensemble Modelling . . . 103

IV. Conclusion **105**

7. Rules for dealing with climate models **107**
7.1. Common sense rules of model tuning . . . 108
7.1.1. Rules of model tuning . . . 109
7.2. Common sense rules . . . 114
7.2.1. Dealing with uncertainty in climate modelling . . . 119
7.3. Conclusion and outlook . . . 122

Glossary **125**

Preface

This thesis is in the field of the philosophy of science but analyses the rather technical practice of climate modelling in detail. The intention behind the writing is to contribute to the philosophical discussion and at the same to provide something of interest to the climate scientific community. It is thus comprehensible for readers who are experts neither in climate modelling nor in the philosophy of science.
Throughout the argument several words used as philosophical and technical terms appear again and again. Some of them are used in a specific way, which will be defined in the context of the argument, others do not differ from the everyday language understanding. The glossary provides short forms of the definitions and refers to the relevant sections in the text.

Part I.

Introduction and Background

1. Introduction

Climate science is a branch of science, the existence of was denied just a few years ago by several physicists but which is now of increasing importance, the more that people feel the impacts of climate change. Having made the leap from the realm of a small crowd of reputed eccentrics to an influential science in the critical focus of the media, the main tools employed in climate science, climate models, have also come under the scrutinisation of the public.
Looking in the community's most important publication, the latest Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC), released in 2007, the model based projections of future global warming range from a 1.1°C to a 6.4.°C temperature increase whereas more recent publications go even further to assign an increase of more than 8 degrees if the world continues each year emitting more CO_2 than the year before at an accelerating rate. Such an enormous range sounds rather vague and therefore implies an enormous uncertainty in climate modelling results. Additionally, the term model allows people to suspect that a climate model is just a model and not related to the real world. Both assumptions are true, as climate model projections are highly uncertain and climate models do only represent the real world in very specific aspects. Thus, two questions arise: why do we use climate models? And why do we think they tell us something about the world?

The renowned climate scientist Veerhadran Ramanathan answered the first question on his visit to the Potsdam Institute for Climate Impact Research in 2008 as follows: "We do need climate models because it is the core of our thinking, we cannot think nonlinear¡' He is perhaps not entirely correct in this answer but he brought up an important problem of climate science, which is the nonlinearity of the climate system.
We can represent the climate system or aspects of it with the help of mathematical equations, which includes nonlinear equations. Hence nonlinearities can be grasped very well by human minds, but we cannot solve these equations analytically, instead numerical solutions must be found. There are many processes in the climate system which have already been discovered but which are not understood or are too small to consider in climate models. In order to nevertheless represent them in models simulating the climate they are displayed as parameters. Some parameter values can be assigned that have counterparts in the real world but others result only from model fitting. This so-called parameterisation is especially important as it is the only method to begin answering the most pressing questions without fully understanding the system as a whole. The parameterisation and model fitting which is included in the process of climate model tuning is crucial for climate modelling and represents very well what distinguishes climate science from other branches of physics.

This difference between climate science and classical physics is certainly the impossibility of performing controlled experiments, or indeed experiments at all. The only 'experiment' we perform is the modification of the climate of and life on the planet by the current alterations to the atmospheric composition. Thus it is not possible to perform experiments in earth science as the repetition of an experiment, which characterises scientific experiments, is impossible. We have only got one Earth.
Since the role that experiments play in physics is so central also climate physics does not work without it, despite the uniqueness of the earth, by replacing classical laboratory experiments with climate model simulations. This is not a satisfying replacement but a mimicry of experimenting with the climate system, and one that is inadequate for many purposes but quite good for others.

We can think in a nonlinear way in terms of being able to set up nonlinear equations but we cannot understand nonlinear systems if understanding is interpreted as computing or even managing the system. On the other hand nonlinear systems are not intelligible for us as it is impossible to know the causes of the effects we observe in general. Then again there are correlations in the climate system we are perfectly able to understand, which is due to the fact that the climate system is, in a manner of speaking, twofold in thermodynamics and dynamics. The latter is what causes fundamental problems in understanding the climate system, whereas the thermodynamic basis of the climate system is computable and serves as fundamental for every climate model.

It is not only the nonlinearity of the climate system that prevents a thorough understanding of the system but also the impossibility of observing all relevant climate variables in the oceans, the atmosphere, and other spheres. This lack of data is a fundamental problem and the most important one for climate modelling. The problem of little data available to initialise the model, to find parameters, to tune it, and to validate the modelling results, intensifies other problems that cause uncertainty in climate modelling. These are crucial processes in model development and besides the initialisation and the validation they are necessary due to the nonlinearity of the climate system and the sparse observational data. Fundamental problems in understanding the climate system thus result in modelling problems.

Besides these fundamental deficiencies it should be underlined that the basis of climate modelling is physics, thermodynamics and several other important processes, which are physics systems and therefore can be described accordingly in terms of equations of motion and the accompanying equations of state. The impossibility of solving these equations analytically is a characteristic the climate system shares with all other complex mechanical systems. Several aspects of climate physics are also considered in fluid dynamics or chaos theory. What sets climate science apart from such disciplines of physics is the fact that the parts and aspects of the climate system, especially vulnerable to human influence and thus of most concern, can only be considered by merging chaos theory, statistical physics, and fluid dynamics, as well as chemistry and biological research etc., and this would not include anthropogenic influences up until now. Taking

them into consideration would require the additional implementation of socio-economic aspects in the climate models which is additionally difficult due to different modelling strategies in climate and economic modelling.
The climate system in its relevance for life on earth is a system of analytically unintelligible interactions, they can therefore not be analysed separately without great loss of realism. This is an important difference to other physical disciplines, without climate models we would be unable to understand the climate system as a single, integrated whole.

These are some basic arguments why we do use climate models and do think we can learn something, as at least each model run is of value as it presents a 'what if' scenario from which we may learn about the model itself or the earth system (Stainforth et al. (2007)). Nevertheless the value could be higher if some principal problems of climate modelling were addressed more appropriately in terms of uncertainty.
The result of fundamental and current problems of understanding the climate system is uncertainty in modelling the system. In almost every case the uncertainty is of such a degree that the assignment of statistical probability density functions to model output is impossible. It is thus very much a question of whether "non-statistical“ probability functions are really of value or if alternative measures of uncertainty are more appropriate.
All these aspects taking climate modelling apart from classical disciplines of science begs the question of whether these differences in the epistemic foundation lead not only to problems in the practical approach of climate modelling but also reveal different epistemic questions. That this is indeed the case will be seen when trying to define the term climate model and in particular in the concluding part of this thesis, which provides some guidelines to improve climate modelling and also the communication of climate modelling approaches between scientists but also to the public.
The approach of this thesis may be seen as a case study in the philosophy of science on climate modelling.

1.1. The debate

The question of the epistemic status of climate models is interesting not only per se but also in the context of some discussions in the philosophy of science which it touches upon.
There is the debate about the realism of science and its success, which includes such big names as Thomas Kuhn and Karl Popper, Imre Lakatos, and Paul Feyerabend, Larry Laudan, Ian Hacking, Carl Hempel, Nancy Cartwright, and many more. This debate is very much independent of climate science therefore it is not reconstructed at all in this thesis but some arguments and some branches of the main discussions will provide assistance in explaining the failure of climate models as well as their usefulness. The epistemic status of climate science is very seldom addressed in philosophy, even less is that of climate modelling but the probability aspect in particular causes more and more interest for philosophers, for example, Wendy Parker.

Climate models are based on spare data and first principles which come from theories corroborated by a huge amount of data, such as Newton's mechanics and fluid dynamics. Building computer models and in particular climate models is a different undertaking than theory building which is concerned with the main debates in the philosophy of science. Therefore the epistemic questions related to computer modelling are being addressed increasingly by for example, Gabriele Gramelsberger and Paul N. Edwards.
The overall aim of this study is not to bring all this together in a very special case of general circulation climate modelling but to analyse the epistemic status of such models by analysing the the process of modelling. At certain steps in the process of model building some questions arise which are the topics of prominent debates, only in such cases some links are established to selected arguments in the discussion, of which the most important are outlined below.

Philosophy of science is fundamentally concerned with scientific theories, how they come into being, how they are confirmed and refuted, if they have something to do with truth, and which of their properties deserve the label of "scientific" at all. Climate science and especially climate modelling is not built on a comprehensive theory and many aspects of the climate system lack of conceptual descriptions and ideas. Thus some classical questions of the philosophy of science, which are central for widely discussed branches of physics like quantum mechanics, but also for scientific theories in general, do not arise in climate modelling contexts, nor are they central to the epistemics of climate modelling. This lack of a comprehensive theory puts deduction and the deductive results of climate scientific undertakings into marginal cases. However, it is important to highlight that deductive findings are marginal but central for climate modelling since the fact of global warming is to be deduced from very basic physics theories. But apart from that, the debate around inductivism in scientific theories seems thus to be linked naturally to some questions that climate modelling presents to the philosophy of science.
A very common view of how inductivism works (e.g. Chalmers (1999)) is that first the principle of induction holds which demands a large number of experiments under a wide variety of conditions and no exceptions to the observation for an inference to be drawn. Secondly, the clue of inductivism is that if laws and theories are inductively derived from experience, the predictions derived from these laws are deduced. Of course inductivism is also used to confirm theories especially nowadays when observations cannot be done straightforwardly. Traditionally, theories that were built as inductively correct were regarded as automatically justified. But there has never been an argument as to what 'inductively correct' means.
It was this view that led David Hume (Hume (1993)) to come to the conclusion that there is an unsolvable problem of justifying our theoretical knowledge. He was the first one to characterise the problem of induction realising that inductivism does, from a logical point of view, not provide any correct knowledge. The problem of induction is not that inductively confirmed theories are not *certain* but that we have no reason to believe that any of them deserves *any confidence at all*, as Hume famously stated. All these interpretations of inductivism argue from the point of view of inductive theory building or confirmation, which is not what climate modelling is about.

In contrast to some other modern sciences some basic content of climate modelling can be confirmed inductively by measuring for example, temperature or rainfall, whereas a quark, in contrast, cannot be observed in such a way. This points to one of the most important problems for climate modelling, because inductive confirmation of some findings would be possible, if reliable data was available, which is in principle a problem. The problem is not that the data underdetermines the findings, which is the case in every inductive argument according to Lipton (1998), but that there is no data with respect to some crucial processes and variables in climate modelling.
Russell (1999) describes the problem of inferring from regularities to generalities with the example of a chicken expecting to be fed by the farmer as it is every day which one day finds the farmer breaking its neck. This example shows where arguments with respect to inductivism may become of importance for climate modelling because the fact of anthropogenic climate change is a singularity in the history of climate. How to infer inductively and at all on such a soft basis is thus a crucial question.
The literature of the philosophy of science provides no comprehensive answer to the problem of inductivism as is the case for most basic questions in the philosophy of science. This thesis will certainly not change this fact but may show that in the context of climate modelling the crux of some questions has shifted. An example is a topic discussed in terms of inductivism under the phrasing "new data" and "novel prediction". A very central guideline which prevents basic modelling errors is: never use a datum twice. There is a controversy in climate science about whether or not data to confirm a theory must be new, where new means not known by the scientists who are providing the hypothesis to be confirmed. Opposing views are given by explanationists claiming theories to be valid if they explain known phenomena and predictionists assigning this potential only to newly predicted phenomena. There are several views between these two extreme positions which are given as a paradox by Carl Hempel and contrasted by Musgrave (1974) as logical versus historical confirmation. In climate modelling it is very plausible to explain that the use of the same datum within the modelling setup and the validation leads to circular reasoning, thus data for model evaluation must be new. However, the crucial question now is what it means for the datum to be the same.
Nevertheless, predictions are one of the most important reasons to build climate models. A very central question is thus: how to identify a successful prediction? Is there a demarcation line that can be drawn when the predicted climate and its confirmation resemble each other enough to deserve the term confirmation. This line does not seem to be easy to draw and the even more compelling question is that of how a prediction could be identified as successful before the reality check is possible.
Climate models are built to predict global warming, and as this is a fact known independently of models, the simple fact that they predict a warmer future is for practical purposes a necessary condition but not very illuminative. However, the conditions for predictions and successful predictions in the philosophy of science are basically discussed in the context of the novel prediction debate. Within this context it is furthermore to question of how hindcasting, which is the simulation of past climates with climate models, fits into the corroboration of models and their predictions. In the importance of hindcasting it can be seen that prediction is not necessarily time dependent as the his-

torical view, Alan Musgrave (1974) discusses, implies. Menke (2009) concluded for the methodological value of predictions that newness is vague in every case and needs further refinement. For him predictions do have a methodological value because they refer to other values especially to simplicity. However, he realises that simplicity is an equally unsatisfactory, vague, ambivalent, and opaque term. All attempts to compile a methodological value of predictions show that it always goes with something owning its own methodological value. Nevertheless there is some result, again from Menke, that at least logical and historical values do not contradict each other. Climate modelling approaches use old evidence, in the sense of knowing it, and build their models to represent it. Only to use the exact data once either in building the model or in its validation transfers the debate from newness to exactness or sameness.

Another very central wish in the discussions in the philosophy of science is to find criteria and arguments to distinguish science and pseudoscience. There is no need to show that climate modelling is pseudoscience since it is a branch of physics. However, some important arguments to determine the epistemic status of climate models stem from this discussion most prominently from Karl Poppers undertakings to reject positivism. Popper's intention was to be able to distinguish between science and pseudoscience (Popper (1998)) by one demarcation criterion, with the aim of being able to label Marxism or psychoanalysis as pseudoscience. We may find that he was a bit too rigorous in excluding disciplines because there is not much left but the pure natural sciences, such as physics and perhaps chemistry, after applying his criteria. In my opinion and in the debate following Popper's ideas this strict classification of academic work into scientific and not scientific by means of one criterion- the demarcation criterion as Popper (1959) calls it - is not meaningfully possible. The possibility of logically falsifying scientific findings, which is Popper's demarcation criterion, is not given in climate modelling, but this may also be the case for most theories. This is even if it may be arguable, if Nancy Cartwright (1983) is right in claiming that the laws of nature are not even approximately true but known to be false, that scientific work is successfully possible without logically falsifiable theories.
However, Popper identified a very important aspect which indeed is a criterion of scientific work: the critical and rational discussion. To be able to hold such discussion, transparency of methods is necessary. It will be seen that climate modelling approaches could do with some improvement in that direction.
Thomas Kuhn (1996) provided four criteria for a good scientific theory: accuracy, consistency, scope simplicity, and fruitfulness. There is no comprehensive theory of the climate system and the criteria are open for discussion but Kuhn himself also does not regard them as demarcation criteria but rather in terms of norms or values, as they influence theory choice but do not govern it.

It is also a controversial discussion in the philosophy of science whether or not normative rules and methodological rules should govern scientific development. Paul Feyerabend (2010) is the most prominent advocate of an "anything goes" point of view against methodological rules for science.

Nevertheless the conclusion of this thesis will include a set of guidelines for good climate modelling practice. These are in particular rules that must be kept in order to avoid typical mistakes. Rules are generally justified if they help to realise certain aims. The aims of climate science are, among other things, to provide sound climate modelling results. Therefore the models must be built accordingly. The rules that will be provided to achieve this are very vague but nevertheless they help to concretise what is meant by transparency. Working with these rules may furthermore enable a refinement. In demonstrating that some rules have been adhered to, accusations of manipulation could be rejected or justified.

Throughout this work the modelling processes are analysed and questions in need of answers are outlined. However, as is quite common in philosophy there are no decisive answers to be given, if an answer can be attempted at all. At a few points when not only questions arise but answers are attempted they are of more of a technical nature and, in the words of Alan Musgrave, when trying to find a position that is neither realist nor antirealist, devoid of philosophical content. However, the questions are important for the philosophy of science and may inspire new answers in the untypical context of climate modelling.

2. Basic knowledge

To examine the possibilities and limitations of modern climate models it is not only necessary to know the basic facts about the climate system but also facts about its representations- climate models. The following chapter gives the necessary background knowledge for an understanding of the terms *climate system* and *climate model* and for an examination of the question of whether there can be said to be a theory of the climate system. A comparison between the latter and the former will allow an elucidation of the epistemic status of climate models.
The following chapter is therefore dedicated to four main topics, the climate system, climate change, climate modelling and its theory, giving the background knowledge necessary to follow the arguments in this discussion of climate modelling.

2.1. Short introduction to climate science

To state the epistemic status of climate modelling it is essential to see the difficulties in climate modelling and in understanding the climate system. The latter in particular is more easily understood if a description of the climate system is divided into separate descriptions of thermodynamics and dynamics, which is done in the following introduction to the climate system. Furthermore, as all climate modelling is to some extent a mathematical operation the most basic set of climate system equations is briefly discussed. The implications of the facts stated below are discussed in chapter II.

Traditionally climate is defined as the average weather over a period of time, ranging from several years to millions of years, the usual period being thirty years. This definition is not exact, neither is that of weather. According to LeTreut et al. (2007), weather is the fluctuating state of the atmosphere around us. It is characterised by weather elements like temperature, precipitation, wind and clouds. It is the result of rapid development and decay of weather systems such as pressure systems and frontal zones. Climate is the average weather in terms of the mean and its variability over certain time-spans and areas. Climate varies strongly from place to place and over time but statistically relevant variations of the mean state of the climate system or its variability occur only once in several decades or centuries. Using this definition the climate sciences are confronted with a somewhat contradictory situation. On the one hand weather is a chaotic phenomenon which we are able to predict for only one week into the future. On the other hand, it is possible to predict global temperature development for the next decades. As such prognoses are possible even if they cannot be called true predictions (see section 3.8.1)the long-term behaviour of weather does not seem to be chaotic. Thus

climate "may be taken as the distribution of states on some *attractor* of weather" (Stainforth et al. (2007)). The definition of climate as average weather is most common and easy to grasp as weather belongs to our every day experience. The attractor definition is not different from that but shows that the climate system belongs to the realm of nonlinear dynamics. In dynamic systems an attractor is a time-invariant entity of the phase space that attracts other points of the phase space so they tend to be part of that entity too. Within this definition of climate, the fact that the climate system is an inertial system is underlined. In this approach, as in every day experience, climate means "average weather".
The reason for our ability to predict a temperature development but not future weather lies in the fact that temperature basically depends on thermodynamics, whereas weather is a dynamic phenomenon. The thermodynamics of the climate system are quite well understood, but the most important aspects of dynamics are not (see below).
The earth's climate is the climate of a complex, nonlinear system consisting of several complex and nonlinear subsystems. Thus, to enlarge the meaning of *average weather*, *climate* refers to the state of the climate system as a whole, including a statistic description of its variations. Even if the definition by means of weather is most common it is usual in climate science to approach the climate system as a system of subsystems. This is therefore the approach that will be taken in the following discussion with an emphasis on the openness of the climate system. To truly describe this system countless subsystems should be taken into account. Not only the standard set atmosphere, oceans, cryosphere but also biosphere, lithosphere and all other spheres and systems on micro and macro scales form the climate system. There are no fixed boundaries of the system. This fact is especially important when dealing with the modelling of the system, which will be discussed further in section 3.3.
But the most important components of this system are the atmosphere, the hydrosphere, the biosphere, the lithosphere and the kryosphere. Every part of this system is in itself a complex system with internal feedbacks, with distinct characteristic time scales ranging from days to millenia. Each subsystem is subject to forcing from other earth system components and to external forcing. The basic external forcing for the climate system is the incoming solar radiation.

2.1.1. Thermodynamics

The external forcing is normally given as an amount of energy per square meter per second averaged over the entire earth. But due to the different distances between the sun and points on the earth's surface, the forcing is not at all homogeneous around the globe. Approximately 30% of this energy is reflected back to space by clouds, aerosols and the surface of the earth. Due to the reflecting properties of aerosols natural events like volcanic eruptions or aerosol-containing industrial emissions may induce considerable changes to the earth's energy balance. The remaining two thirds of this energy is absorbed by the earth and, to achieve a balance, emitted back as longwave radiation. If this longwave radiation was directly emitted to space the average surface temperature

of the earth would be −19°C which is 34°C less than it actually is. However, only ten per cent of the outgoing surface radiation goes directly into space. This is due to greenhouse gases and some types of clouds that absorb a small part of the surface radiation and reflect more than 80% back to earth. It is the same effect that the glass roof of a greenhouse produces. The most important greenhouse gases are water vapour, existing only for days, and carbon dioxide, which can survive for millenia.
These principles of thermodynamics, which form the basis of a climate system, are well understood.

2.1.2. Dynamics

The most important origin of atmospheric circulation is the difference in incoming solar radiation from equator to pole, leading to energy differences on the earth's surface. This difference caused by the spherical shape of the earth generates different temperatures and therefore varying atmospheric pressure, the compensation for which induces wind systems. Naturally the same is true for the oceans and their circulation systems.
Other important determinants of atmospheric circulation are the rotation of the earth, the orography, and the location of continents and oceans and the latent heat. Approximately one fourth of the incoming solar energy is used to evaporate water from the earth's surface. This energy is saved as latent heat which is released again when the vapour condenses.
The earth's rotation alters the circulation from the solar radiation-induced north-south pattern to a more east-west directed circulation. The planet's topography causes geographically stationary wave systems of low and high pressure on a planetary scale. Thus changes in the structure of the surface cause changes in the dynamics of the climate system. These changes are especially substantial if they involve positive feedback mechanisms which amplify the effects of changes. A well-known example of positive feedback is connected to the melting of ice sheets. A warming of the polar regions induces their melting. With this the colour of the surface in this region changes from white to the colour of dark soil or water, which absorbs a larger fraction of the incoming sunlight and therefore exacerbates the warming of the region, which in turn increases the speed of the melting.
Besides this so-called ice-albedo effect, many other feedback mechanisms belong to the climate system, positive as well as negative. Two positive mechanisms which are especially important in the context of global warming (see section 2.2) are firstly the increasing inability of the oceans to assimilate carbon dioxide due to increasing ocean warming and secondly the increasing ability of a warming lower troposphere to contain water vapour which functions as a greenhouse gas. As the oceans are *the* natural CO_2 sink and as water vapour is the greenhouse gas of the highest concentration in the atmosphere, these examples of feedback become more and more important with increasing climate change.
Strong negative feedback mechanisms are not as easy to identify. One potential negative mechanism is an increase in cloud covering due to increasing temperatures. The increased

cloud thickness or extent could reduce incoming solar radiation and limit warming. But this is only true if the emerging clouds actually have this ability. There are also clouds that work in a similar way to greenhouse gases. Thus cloud mechanisms are an example of an unsolved mystery of climate physics too, apart from poorly understood feedback mechanisms.
So far the climate system is described as the more or less balanced interaction of the earth system's spheres. The balance depends on the steady continuity of external and internal forcings, where the latter are especially positive and negative feedback mechanisms.
It is important to state that thermodynamics and dynamics are not equally fundamental. A climate system without any dynamics is still a climate system, whereas the dynamics only emerge on a thermodynamic basis. As will be discussed in chapter II this classification is mirrored in the different degrees to which thermodynamics and dynamics are understood.

This conceptional differentiation has to be dealt with in modelling the climate system. On the one hand the aim is to understand and simulate an equilibrium climate, on the other hand it is to model dynamical climate development. The latter is not to be confused with the dynamics of the climate system, but instead aims at understanding transient climate phenomena such as anthropogenic climate change and processes of natural climate variability. A prominent example of a transient climate phenomenon is the so-called el niño event, which causes abnormal warming of the south American Pacific coast.
The simulation of an equilibrium climate is not concerned with such phenomena and instead aims to provide an average climatology. Such a static picture of the world does not only serve an understanding of the climate system in principle but is also important in order to identify variabilities and to tell transient climate phenomena apart from average climatological behaviour.
Both concepts of climate simulation require different ways of thinking which have resulted in diverse climate modelling approaches.

2.1.3. The primitive equations

Every subsystem of the earth's climate system as well as the system as a whole is a physical system that can in principle be described by physics equations, which principally allow for a reasonable computation of all temporal and spatial spheres of the climate system. Unfortunately, an analytic solution of the equations is impossible. So it can be said that they are basic to climate modelling but not to climate models.
There is nevertheless a set of equations, the so-called primitive equations, that is a basis for climate models. Within these equations the division between thermodynamics and dynamics is represented. These equations are the basis for all sophisticated climate models dealing with energy balance as well as equations of motion. Thus they are the starting point for most dynamical studies, from day-to-day weather forecasts to complex paleo climate reconstructions.

Principally there are three governing laws framing the climate system: the continuity equation, the Navier-Stokes equation, and the thermal energy equation. The continuity equation grants the conservation of mass and constrains the mass field. The conservation of momentum is granted via Newton's second law of motion. This law leads to the equations of motion, normally given as the Navier-Stokes equation, with respect to all occurring forces such as stress and friction. The second equation of state, next to the continuity equation, which governs the internal energy balance evolving from the first law of thermodynamics is the thermal equation which accounts for all system internal fluxes of heat.
The Navier-Stokes equations represent the dynamics of the climate system, i.e. the nonlinear part of the climate system. While the thermodynamics of the climate system are in principle understood and can be given to a known degree of accuracy, i.e. it is at least known what has been omitted, the latter cannot be said of dynamics. It is possible to write the principally correct form of the Navier-Stokes equation but it is impossible to solve it analytically or numerically. This is due to the fact that the dynamics of the climate system are not known on small scales and also poorly understood on large scales. Therefore even the most basic set of equations needs approximations. Differences in climate models are essentially differences in representations of dynamics and approximations, which depend on the aim of the modelling approach as well as on computational resources and numerical skills. The shorter the period one wants to model, the more details can be represented, and the same is true for the regionalisation of models. If only thermodynamics are represented the outcome will be a comprehensive climate model. In contrast to that, the dynamics of the model need thermodynamical triggering, without thermodynamics there can be no dynamics.
Thus, even if they are called primitive the basic equations are highly complex. They are the basis of complex climate models but need several simplifications.

2.2. Basic knowledge of climate change

Besides the principal difficulties, which will be discussed in detail in the following chapter II, the potential imbalance of the climate system places additional obstacles in the way of computing the climate system. When understanding climate as average weather bound to an attractor it is easy to explain why an increasing amount of carbon dioxide in the atmosphere is of much greater consequence than the term global warming implies. Due to the increasing amount of greenhouse gases in the atmosphere the distribution of the climate system has moved from its initial attractor and is now in a transient state instead of a stable one as before. What this means can be seen in the spatial description of the climate system.
The increasing amount of carbon dioxide emitted into the atmosphere primarily from fossil fuel combustion considerably modifies the chemical structure of the atmosphere, with serious consequences for the climate system. Most obviously the fraction of reflected long wave radiation which warms the surface increases. This is especially dramatic as

the energy of the climate system is now balanced. Thus every alteration of factors of this system can unbalance it and force it towards a new balance which does not necessarily include comfortable conditions for human life. Throughout the history of the earth many such uncomfortable climates existed.
Therefore the man-made climate change we are confronted with reveals additional difficulties in climate modelling in addition to illustrating principles inherent to the system that prohibit a true understanding of the climate system. The next section briefly depicts the facts of climate change.

The facts of climate change are based on elementary physics and are thus nothing truly new for the scientific community. What is indeed comparably new is the possibility of using modelled experimenting to learn more about potential influences of climatic changes caused by increasing atmospheric CO_2.
The fact of anthropogenic climate change is undoubtable when the scientific evidence is considered. The predicted and measured consequences are highly dependent on the magnitude of climate sensitivity. This is the equilibrium change of the global mean temperature under a doubling of atmospheric CO_2[1], which is most probably $3 \pm 1°C$. This magnitude of climate sensitivity is the basis for all calculations of emission reduction and its influence on the climate system. It is only if science had erred dramatically in this assessment of climate sensitivity and found it much smaller, that climate change would be a less worrying fact. As mankind has already increased the amount of CO_2 in the atmosphere from preindustrial 280 ppmv to the current 387 ppmv and has observed an increase in global temperature which accords with the projected magnitude of climate sensitivity, such findings are highly unlikely. The opposite possibility, i.e. a higher climate sensitivity than estimated, is not very unlikely and would imply a greater process of global warming and more dangerous consequences. Our knowledge about climate sensitivity is independent of climate modelling, only the assumptions about future changes in climatic behaviour are based on model simulations.
In the public debate about global warming and climate change, the subliminal assumption that all we know about climate change is due to calculations and simulations of climate models is discernable. In fact our knowledge about the connection between CO_2 and temperature results from observational and paleoclimatic data and knowledge of properties of greenhouse gases. In 1895 Svante Arrhenius had already suggested that the increase or decrease of CO_2 triggers glacial and interglacial periods (LeTreut et al. (2007)). Even if we know today that historical climatic changes preceded the change in CO_2, the fact that CO_2 is a greenhouse gas has been known since that time.
Thus climate models are not necessarily needed to state important facts of the climate system and possible changes.
Climate sensitivity as a basis for our knowledge of the climate system and especially of global warming is certainly the most important model-independent factor for public discussion. But apart from that the observation and interpretation of climatologically

[1]As CO_2 is not the only but the most important greenhouse gas except for water vapor, the influence of all other gases influencing global warming are translated into equivalents of CO_2.

important variables during the last century was and is *the* grounding for our knowledge and of course for the development of models.
During the 20th century we observed a temperature increase of approximately $0.6°C$. From paleoclimatic data we know that the global temperature had a variability of several degrees during earth's history. But we can also observe that none of the known factors that caused climate warming in the past is causing warming now. Thus only simple observation gives reason to link the increased concentration of greenhouse gases in the atmosphere to its warming, as researchers like Arrhenius did more then a century ago.
But observation alone was not the only key to learning about the climate system in premodelling time. Of course, classical methods of physics and chemistry provided and continue to provide an important part of our knowledge. As the pressing question today is about the effect of CO_2 and other anthropogenic greenhouse gases on the climate, climate sensitivity is an example of the combined merits of physics and observation. As a modified composition of greenhouse gases influences the amount of infrared radiation warming the earth, the emerging question concerning these changes is twofold. Firstly, how the radiation changes due to a higher CO_2 concentration in the atmosphere? Secondly, what is the temperature increase caused by an increased radiation? Climate sensitivity can be described as $°C$ per unit of radiation or, more common, in terms of a $°C$ temperature increase due to a doubling of CO_2 concentration. This is equivalent to a heightened radiation of $4\frac{W}{m^2}$.
According to Rahmstorf and Schellnhuber (2006), three methods are used to calculate the magnitude of climate sensitivity. The first method uses only physics. With the help of experiments a temperature increase of $1.2°C$ is measured caused by a doubling of CO_2. As such experiments do not take into account any feedback effects, thus this value is the lower limit of climate sensitivity. The only possibility of reaching a lower climate sensitivity would be by a very powerful negative feedback mechanism, but as the most potent known feedback effects, namely the ice-albedo-feedback and increased water vapour, are positive, such a finding is highly unlikely.
The second method is based on observational data of past climate variabilities. These data must be of high quality as they are used to separate the effects of single processes and factors. The results are best if the concentration of CO_2 in the data varies highly but most other important factors, such as solar constants and land distribution, are equal to today's values. The best source for these data are ice cores from Greenland or Antarctica. With this method a magnitude of climate sensitivity between $3°C$ and $4°C$ is found.
Before being able to build climate models scientists must accumulate the necessary know-how, mostly on the basis of the physics of fluid nonlinear dynamical systems. However, next to this relatively young branch of physics lies some very basic physics research, important for climate science. That is the theory of gases, starting with simple concepts such as temperature and pressure, linking this to radiation absorption and emission characteristics of different gases and fluids. Historically, after the general behaviour of gases the first knowledge of turbulence was gained, which led to modern theories of chaos and complexity. Apart from the latest findings in the latter field, all of them are perfectly suited to computer modelling free research. The theoretical insights of these sciences are

also fundamental to climate science. Only upon their findings and mathematical theories of numerics do we have the possibility to start thinking about modelling. Before ever thinking about model building we knew that the Navier-Stokes-Equation provides the equations of motion for the climate. The development of physics and especially fluid dynamics was necessary in order to be able to deal with subjects like the atmosphere, and at the same time it underwent important further development with regard to the study of weather and climate phenomena. As early as the 19th century Henri Poincare was praised for his ansatz to find an answer to the question of whether the solar system is stable. Similar stability analyses can be undertaken with regard to climate and weather phenomena. Even if these equations were not solvable at the time, the possibility of writing them down required considerable insight into the internal dependencies of weather and climate phenomena.

Apart from such theoretical understanding practical knowledge about the interrelations of the climate system, especially the atmosphere, was also gained. In 1859 John Tyndall identified the absorption of thermal radiation by complex molecules via laboratory experiments. He drew the conclusion that changes in the amount of such atmospheric components as H_2O and CO_2 may have caused historical and prehistorical climate changes. The insight that thermal radiation does not pass through transparent material as easily as solar radiation was already gained by Horace Benedict de Saussure in 1760.

Even if it has only been established recently that in prehistoric times the temperature rise preceded the increasing concentration of CO_2 in the atmosphere, the correlation was posited almost 200 years before the very first climate model was invented. Before this time attempts were also made to combine theoretical knowledge and the findings of experiments with greenhouse gas absorption. Callendar (1938) found that a doubling of the CO_2 concentration will lead to a global warming of about $2°C$ with the warming being highest at the poles. He came to this conclusion by solving a set of simplified numeric equations. His findings are still up to date as they compute a climate sensitivity, without referring to it by this term, of $2°C$, which lies perfectly in the range of today's magnitude of $3{\pm}1°C$.

These historic findings underline the fact that the most important mechanisms triggering the climate system can be examined without the help of climate simulations. This means that the principles of climate change can be studied on the basis of physics and observation, thus providing qualitatively valid results. Quantitatively the picture that can be drawn is much more vague. Of course, the magnitude of climate sensitivity is computable without simulations but it is only one aspect of describing climate change, and the most basic one. And when seeking an understanding of the climate system as a whole, climate sensitivity alone is even less useful.

Considering that climate change is the most pressing question of public interest, it is not about the magnitude of the global temperature rise resulting from a doubling of CO_2 concentration in the atmosphere. What interests people is whether there will be enough rain to irrigate their fields, whether there will be more hurricanes, more floods or whether the sea level will rise and endanger their home. The question is also of course not only if this will happen but of when and where and if there will be areas where the climate will change to be more suitable for habitation by humans. In short, what

interests people outside of science are the impacts of climate change. In order to assess these, the possible consequences of global warming, modelling is the most important tool we have. Predictions are meaningful, if at all, only in modelling contexts. The only prediction that can be made on the basis of climate sensitivity alone, without modelling, is the fact that the global mean temperature is rising and will rise further.
Some aspects of these impacts can be assessed independently of modelling while comparing today's observations with prehistoric data on climate changes. But as there never has been a climate change that equalled the one we are witnessing today, this comparison can only give hints but never be sufficient. Especially if the impacts must be known on regional scales the only possibility of learning about them before they occur is the use of highly developed model simulation techniques.
But before being able to compute the possible effects of climate change, models are necessary which allow us to be certain of the anthropogenic accountability of global warming. This is because "without a model of what would happen *without* anthropogenic atmospheric change, scientists cannot separate out the effects of rising greenhouse gas concentrations from natural climate variability" (Edwards (2001)).The most pressing questions are not those of complex climatic feedback mechanisms and the chemical composition of the stratosphere but of temperature and water supply. The first goal of numerical climate modelling is to find answers to the question of whether this planet can continue to sustain life.
A comparatively difficult task is to link droughts or storms to man-made climate change. As these phenomena emerge due to the high and chaotic internal variability of the climate system their analysis calls for a coupling of statistical analysis and model simulations. Therefore cleverly designed complex models are needed.
With regard to this topic another great challenge, apart from that of the impacts of climate change evolves. This is the internal climate variability. Changes in solar constant, land use or other external and internal forcing factors can be detected when analysing statistical records. But it is hardly possible to learn about variabilities generated by internal dynamical feedbacks without designing, or modelling, a system with comparable internal feedbacks.

Besides this classical comprehension of the climate system and climate change as described above, it is possible and perhaps more instructive for our purposes to approach the topic from a system theoretical point of view.
According to a typical climate system description the earth is an complex sphere with its many subspheres, hence the physical climate system. However, since the Romans already began to alter the environment by deforesting huge areas to build their naval fleets, the influence of human activities on the ecosystem is not to be neglected. At least given today's knowledge of anthropogenic climate change, the interference of civilisation with the ecosphere is significant. Thus the development of the climate system is dependent on the evolution of the ecosphere and correspondingly on that of the anthroposphere as strongly coupled factors.
The scientific discipline evolved from this important insight is Earth System Analysis, which investigates how this interrelated complex of ecosphere and anthroposphere re-

acts to certain disturbances and how their consequences can be dangerous to nature and humankind. While the management aspect of this branch of climatology is essential, it will not be considered in detail in this paper as it would bring up as many more complex problems as we are already faced with. Nevertheless, the idea of managing the climate system requires an even deeper level of understanding. It is not enough to simulate past and present climate states correctly, rather a feasible prediction of future climate states and their changes due to intentional and unintended human interference is also needed. In this context of 'understanding', only the ability to change the system according to our conception is a true understanding, not the construction of model systems alone. This means that the model system must be as reliable, in order to enable us to check management strategies in the model system, the results of which will be equal to those seen in the real world.

However, this is of course not the only way of defining understanding and it is certainly not the one that can realistically be achieved in climate science. The next less deep form of understanding would refer to the ability to compute the system, but even this cannot be achieved realistically, as will be explained in chapter II.

On more superficial levels, it is hard to still use the word understanding, which seems to imply some deep knowledge of the system. In my opinion the deepest understanding that can be achieved in climate science is not to be surprised by the system's actions and reactions on larger scales. The reasons why this is not a modest aim and the possibilities of reaching it within today's research frameworks will emerge in the context of the following chapters and will be the subject of the discussion of the concluding chapters 6 and IV.

2.3. Defining *climate model*

As soon as we include the word model in the term climate model we are in the middle of a much disputed field in the philosophy of science: the definition of the terms model and theory and their demarcation from each other. Even if a detailed discussion of this issue is of no further interest for the purpose of this paper the term climate model needs to be delineated. This section attempts to provide a philosophical definition of these terms which will be complemented by an overview of examples of actual climate modelling in section 2.4.

2.3.1. Theory of the climate system

If the climate system was a simple physical system it could be described with a classical physics theory, consisting, in the words of Ludwig (1974), of three parts: a *mathematical theory*, a *reality domain* and an *instruction for use*. While the latter is the set of axioms linking mathematics and reality, the *mathematical theory* is the straightforward part of the physics theory. In principle the *mathematical theory* consists of axioms that allow the production of proofs leading to true statements. The *mathematical theory* has to be consistent and completely independent of the physics theory it is designed to constitute. In Ludwig's interpretation of theory building, the *reality domain* entails a part of reality which we regard as given. It is the fraction of reality we perceive independently of the physics theory. That does not mean that it is independent of any physics theory but of the one in question. This *basic domain* of the *reality domain* can be explained using an example from paleo-biology. Bones and fossils can be found according to which we believe in the existence of dinosaurs. The bones are part of the *basic domain* of the theory while the dinosaurs belong to the *reality domain*, and not to the basic one. The prehistoric existence of dinosaurs is a fact only in the context of the theory, while the bones can be found independently of it. In physics theories the *basic domain* of reality is the domain of experiments.
To summarise Ludwig's interpretation of physics theory building, physics theories consist of mathematical theories, a carefully chosen section of reality and a set of axioms to link both spheres, where the axioms as the essential part of the *instruction for use* are the core and characteristic of the physics theory.
If we thus had an optimal climate system we would have these three parts and with them a theory of the climate system. That is, we would have:

1. An empirically appropriate description of those rules and laws showing the causality of how the states of the climate system evolve out of each other.

2. A description of the system that is dense. That is, if D and $D*$ are different data we always find $D_1....D_n$ that $D*$ can be deduced with the help of laws and rules from D.

3. The rules and laws can be checked empirically.

4. There are no observations that cannot be explained via the known rules and laws.

5. Data can be gained in arbitrary good resolution.

6. The rules and laws and data lead to good predictions.

For mechanical systems, for example, these conditions are met. As Ludwig (1974) gives his definition of a physics theory in the introduction to his book on the fundamentals of theoretical physics he had systems like that in mind. As has already become clear in section 2.1, we do not have an optimal climate system. The main shortcomings are:

1. It is not possible to calculate states of the climate system using known laws of physics as we do not know the causal mechanisms.

2. This is partially due to the fact that we do not have sufficient data whether in spatial nor temporal resolution.

3. Experiments are impossible.

4. The interactions of the subsystems and subscales of the climate system are insufficiently understood.

5. Predictions of quality are impossible.

Thus we do not have a theory of the climate system. In this thesis the shortcomings of the system preventing meaningful theory building will be discussed in detail. Nevertheless we have a huge amount of climate models, most of which are epistemically meaningful.

2.3.2. Climate modelling

Although the uses of climate models have only been touched upon so far, it has become apparent that it is not a trivial matter to ask what is meant by the term climate model. There are several possibilities of conceptualising climate models. The term is obviously not synonymous to computer model or numerical model, although both of these notions seem to be entailed in the concept of a climate model. But there is more to be subsumed under the complex idea of a climate model. When reading the term in literature or using it, climate model also refers to a conceptual representation of the climate system that is not only a set of equations but also contains a story behind it. Thus it must be taken into account that the second part of the concept is the term model which is one of the terms in scientific and popular history representing a particular broad set of ideas. A model is a beautiful person that gets money for being beautiful. It is a miniature airplane or the picture of it and my new kitchen on a computer screen. A model is an idealised context in which simple laws of physics are true without further assumptions. It is the relation between different quantities derived from observations. A model is a set of equations implemented in machine code to solve them. And a model is a draft of a theory that needs some finishing thoughts. All of these connotations are associated with the term model and so they also accompany the term climate model.

In analogy to the term theory as described above, a useful interpretation of the term model can be given. The task of this section is to find a realistic definition of the term climate model which makes the implicit meanings explicit without adding superfluous content. For the main aim of stating the epistemic status of such models and ascertaining crucial rules in the construction of climate models it is important to find out what is really meant by the term in its every-day usage and not what would be convenient to be meant by it. Thus the philosophical handling of the term is helpful for the finding of implicit interpretations of the term but the governing discipline for interpretation must be climate physics. Therefore, from paragraph 2.3.2 onwards the physical process is analysed, whereas the previous paragraphs are dedicated to sharpening the abstract term climate model.

In the words of Tetens (2003), the part of reality known as W serves as a model for that part $W*$ if and only if a representation or presentation of W is useful to discover something about $W*$, because W is in parts analogical to $W*$. The presentation can be direct observation, formal description, mathematical or in prose, or in short every method commonly attached to the term model.
Important in this definition of model is the need for partial structural analogy. In this way not every idea connected with the climate system that comes to mind becomes a model. The definition covers two important aspects of modelling: representation and interpretation. The former purpose of a model is given if the analogy is in the results, i.e. the modelling results look similar to the reality part it represents. A good example of representative parts of climate models are the parameterisations. The mechanisms are not analogue to those in reality but the represented climate variables are comparable.

This is an important feature of climate models and is different from purely explanatory and interpretative models as for example, billiard balls explaining molecule behaviour. The representative character becomes especially evident in the common analysis of climate model data via graphical maps which look exactly the same as maps constructed from observed data. This is not a random choice of representation or only for the purpose of better comparison, but it is used because a climate model tries to simulate the climate system, even if under deliberate or inevitable simplifications. Seen from this point of view a climate model is an experimental setup.
The interpretative aspects of climate modelling are those involving physics theories. For even if there is no theory of the climate system as a whole, parts of it are very well covered by physics theories. These parts are not subsystems but physics theories of for example nonlinear dynamics and thermodynamics. The primitive equations as shown above (section 2.1.3) are based on these branches of physics.

Climate models as models in general serve representative and interpretative purposes. Thus they can provide a better understanding of the climate system in terms of a better knowledge of the interconnection of climatic processes and an explication of climatic phenomena caused by the implementation of new processes. Apart from these two characteristics of a climate model they can also have a predictive purpose. If models succeed

in fulfilling this purpose it corresponds with another level of understanding the climate system. Being able to predict the development of a system is arguably the deepest understanding that can be achieved for a system.
The climate system confronts us with a complexity of processes that are not open to simple deterministic predictions. A climate model is an interpretation of this complexity. This interpretation can be said to be a description of phenomena for practical purposes, but not necessarily, for the purpose of predicting future climates. An aim of a climate model is quite often to extract one or several influencing factors to connect them to patterns in the variability of the climate. This aim is very practical but rather preparatory for predictions.
While all models are representative as well as interpretative, only a certain class of models have this predictive character. It is only inherent in models simulating the transient climate.
Coming back to the notion of a model as an experimental setup, it should be noted that another specific characteristic of all climate models is that they are tools. Unlike in other scientific disciplines, it is not possible to take parts of the climate system into the laboratory and conduct experiments under controlled conditions. However, with the help of climate models the climate system can at least be simulated. As experimental setups climate models must be understood as having similar imperfections as other breadboard constructions. Thus they are not fixed structures but setups that allow a huge range of experiments to take place. Given the same initial conditions these experiments can be repeated and will produce the same output. It has to be emphasised that a climate model is an experimental setup or, even more basic, can be seen as a laboratory, but it is not an experiment. Only a single simulation with a climate model under specific parameter settings and initial conditions can be compared to an experiment. Both terms, experimental setup and laboratory, do not fit exactly the experimental function of a climate model. It is more than an experimental setup because, depending on the complexity of the model, the parameter settings allow for a whole range of simulations, whereas the notion of laboratory is not specific enough as a climate model limits possible simulations through the considered processes, the way they are represented, the dimension of the model, and the tuning specifications. A climate model, for example ECHO-G at the Max-Planck Institut für Meteorologie in Hamburg, is an experimental setup with a large tool box allowing simulations on very different scales and with different parameters.
However, besides the millions of potential worlds it offers, it is also common to say that ECHO-G shows a much warmer average sea surface temperature for the next hundred years than CM2.0 from the Geophysical Fluid Dynamics Laboratory at Princeton University (Randall et al. (2007)). Thus the term climate model refers not only to the setup but also to the result of a special experiment. In both cases the average is taken from a whole ensemble of model simulations, therefore the term climate model is used in analogy to an experimental standard, and not to the single instance of an experiment, with the standard here being the average resulting from several simulations run under slightly varying conditions.

Double role of climate models

Climate models are not only a special kind of model but something new to our conventional way of regarding things. With the increasing use of computers not only our scientific way of working, but also our conception of reality and its categorisations changes and requires revisions and perhaps new categories.
Above I state that climate models are not physical in the same way as experimental setups in classical physics are. But in another way they are very much physical. Everything in the process of climate modelling is information and information is physical. Every piece of information is a state either of atoms in the outside world, or bits and bytes in a computer, or synapses in our brain, providing thus three spheres wherein information is processed.[2]
Looked at in this way, climate models can be seen as playing two roles, or better one double role. A climate model is a copy of the earth even if it is a very bad one, which allows us to experiment with the earth and its climate as if the climate model were a physical model in a laboratory. A climate model is furthermore an ideal world. Since we are unable to solve the equations analytically or numerically, results cannot be given in reality. But the equations that are the basis for concrete climate modelling can be solved because, within the simplified context of a climate model, the nonlinearities can be handled.
The climate model thus exists in all three spheres of processing information: physically it is a computer model, by implementation of theory it is related to our minds, the sphere of thoughts and neuronal information processing, and by modelling, in terms of parameterisation and tuning, it is linked to the world of matter. Parameterisations are that part of a climate model representing processes not resolved within the model. Tuning a model is then done to adjust the parameters within the parameterisations, with the help of observational data to get the model to simulate a realistic climate. Both methods are critically discussed in chapter II, in sections 3.3 and 3.4 respectively. Figure 2.1 displays this double role a climate model has within the three spheres of information.

Besides the aspects described above the diagram shows additional relevant links between the parts of the three spheres.
Basic physics denotes the basic equations used as model input, which are mainly the primitive equations. The central Navier-Stokes equation describes the motion of particles or in other words micro physics, which is known to us due to classical physics research.
As it is impossible to solve the basic equations, simplifications are made. However, with the implementation of incomplete and simplified theoretical assumptions the world cannot at all to be simulated. To be able to play its role as a model of the earth the climate model must be completed with parameterisations and eventually tuned on the basis of observed data to show a climate resembling more or less the real earth's climate. Such a complete climate model is now used as a copy of the world's climate system. In simple to intermediate climate models it is not only one copy but huge ensembles of possible

[2]Even though all of these three spheres are made of atoms the perception of information is threefold, thus the distinction is meaningful.

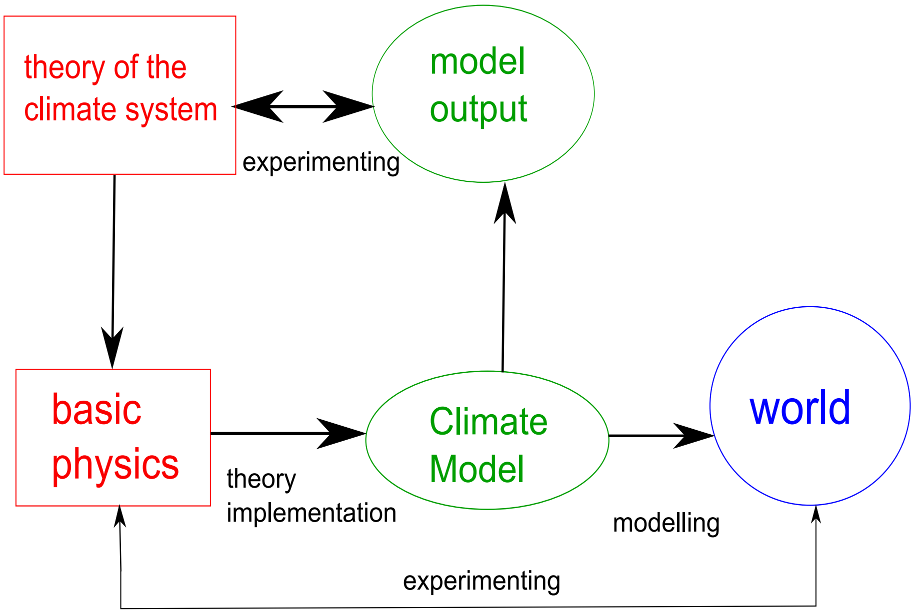

Figure 2.1.: Schema of the double role of a climate model as an outsourced brain and a model Earth. The latter holds for the modelling of reality as well as possible, whereas the model acts as an own brain in computing nonlinearities incomprehensible for human brains. The colours denote the three spheres of processing information. Human minds (red), computers (green) and the real world (blue.)

worlds, whereas for a general circulation model, developed to represent the earth's climate as true as possible, not more than ten different, complete model simulations and thus worlds are created. The created world becomes manifest in the model output.
The model is only determined on the basis of the defined state variables, parameters and model equations. When initial and boundary conditions are added the model will be restricted to one model world. Thus the model output is not part of the formulation of the climate model that can represent many different model worlds but represents one of all possible model worlds. Thus the model output is part of the model only insofar as it is part of one realisation of the modelling approach.
It is important to highlight this aspects of the output as the fact that it belongs to a model world shows that it is a very big step to compare it with the real world we live in. The model output can be analysed as if it were the result of a single experiment. Due to the imperfections of the copies and the, in all probability, stochastic nature of the climate system special care must be given to statistical analysis of model output (see chapter 6).

The analysis of climate models as interpretative and representative models given at the beginning of this chapter applies generally without limitations concerning the type of model. Thus climate models as "interpretative models establish a link between abstract theory and model, whereas representative models establish a link between model and

world" (Cartwright (1999)). In this respect climate models can be taken as mediators between reality and theory. Mediator is thus another term to clarify the double role climate models play.
The distinction between representative and interpretative models is represented in the scheme depicted in figure 2.1. The arrow between model and basic physics represents the interpretational role a model plays whereas the linking of model and world gives the representational role.
For Cartwright these two roles of models relate to different modelling approaches but climate models represent both roles at the same time. Whether the interpretative or representative character of a model dominates crucially depends on the type of model. To fit these purposes, and additionally the role of prediction, a whole range of climate models have been developed. This range is called the spectrum of climate models. An overview of the spectrum is given in section 2.4 highlighting their advantages and shortcomings.

Story telling

Explaining the double role of climate models is epistemically fruitful for understanding the different aspects of climate science subsumed under the term climate model, but insufficient for understanding the process of climate modelling. Climate models are not physical models in the way that laboratory experiments are but they are also not only gedankenexperiments or theoretical models that exist inside our heads and on paper only. On the one hand a climate model does not belong to one sphere or the other, on the other hand it belongs to both, the sphere of laboratory experiments and that of the mind.

A climate model is a mathematical model consisting of equations. The results of simulation runs are huge fields of numbers. In order to start a simulation the model of course does need some input variables, the initial and boundary conditions as well as specific parameters to represent physical processes (see section 3.3). The input consists not only of measured values of observables and parameters but also of immeasurable values for parameters. Those values are assumed but normally on the basis of solid scientific calculations and advice. A climate model is furthermore nothing that one can touch like a laser, but rather the computer model has a code which translates the whole modelling setup into zeros and ones.
Throughout the modelling process we are confronted with many assumptions on very different levels of uncertainty. There are the basic theories of physics as named above, by means of which concepts of mechanisms in the climate system are constructed, which in turn are corroborated by observational data and model simulation. The assumptions made about climatic interrelations are considered part of the model, not a stand-alone theory. That this approach is reasonable is best shown by considering the actual practice of modelling in an example.
The every-day work of a scientist using any kind of climate model consists in improving the model and thus our physical understanding of some small climatological process. Let

us consider, for example, a scientist working in a group modelling atmospheric chemistry. They have implemented a rudimentary scheme of chemistry in an atmospheric model. Having made some simulation runs and compared them to observational data they discovered that the chemistry scheme behaved rather poorly in terms of the stratospheric transport of CO_2. Therefore the scientists look back at the reaction equation and chemical transport laws they implemented in their scheme. They compare their approach with works of other modelers and reread theoretical papers on CO_2 transport in general, then make adjustments in the transport equation. After technical implementation of the changes the simulation study can start again. Then, after some model runs the output needs to be interpreted within the context of the previous general assumptions concerning this atmospheric transport. Perhaps a consistent interpretation along the lines of the previous assumptions is not possible, which means new assumptions must be made. These new assumptions will be based at least partially on the output of the climate model.

Gramelsberger (2006) explains this modelling process schematically by paraphrasing climate modelling as “story telling with code”, which has an underlying process of derivation from theory to mathematics, from mathematics to code and finally from code to the ’story’ about the climate system. Even if the accompanying text suggests an evolutionary process from theory to story Gramelsberger concludes her statement as a system of equations:

$$Theory = Mathematics = Code(f90) = Story.^{3} \tag{2.1}$$

If this process was not evolutionary there would be no difference between the theory and the story. Because if the theory *equals* mathematics, which equals the story because the code is also equal to the mathematics and to the story, then the theory as well as the story is just the prose accompanying the mathematical equations. Only if the transcription of the theory to mathematics and code added something new to it, would the resulting story be unequal to the input theory and would thus deliver new insights. But this is of course what happens, as modelling would be useless if it did not happen. Thus the equals sign is false or at least misleading. The theoretical assumptions made about the climate system comprise mathematics and the story around it. The formal theory, the mathematics, is transcribed into numerically resolvable equations and afterwards encoded to be implemented in a model. The single pieces of code forming the model are programmed to exchange information. Thus they interact with each other in an unforeseeable way because,although the existence and placement of nonlinearities are known, their effect is not. Subsequently the interpretation of the modelling results forms a new story.

$$Story \Rightarrow Mathematics \Rightarrow Numerics \Rightarrow Code \Rightarrow Story \tag{2.2}$$

[3] f90 stands for the programming language FORTRAN in it’s 1990 version. FORTRAN was the first programming language used for weather forecasting models and early climate models, thus it is still used for complex modelling. Nowadays different programming languages are increasingly used.

This diagram represents the processes of climate modelling along these lines. Even if the climate model is only used to test if some theoretical assumptions are correct, the input story would not totally equal the output story because the assumption has at least been linked to a greater system. Normally the output will not be exactly as expected.
In light of the discussion of the double role of climate models another refinement of the diagram seems advisable. The relationship between the left and right item of the diagram is not a straightforward one-way relation from left to right, instead the right item influences the left one as well. The need to solve the equation constrains the mathematics used to formalise the initial idea. The translation into numerically solvable equations, which can be implemented into code, is constrained by the computational power at hand. The latter also influences the number of interaction steps within the model code. Thus the relationship has to be seen as a two-way relationship as the equals sign also suggests, but instead of equality the double-arrow symbolises that the modelling parts influence and constrain each other. Furthermore it is necessary to mention that the described process is repeatable and normally will be repeated several times. Thus a diagram more closely representing the modelling process is:

$$Story \Leftrightarrow Mathematics \Leftrightarrow Numerics \Leftrightarrow Code \Leftrightarrow Story \tag{2.3}$$

In these approaches theoretical assumptions exist and they contain mathematics as well as implicit translation rules but they are part of an incomplete "theory-in-progress" of the climate system. It is not only incomplete but varies, either slightly or more extensively, from scientist to scientist. Thus the term story and its mathematical representation is a more appropriate description of the loose beginning of modelling. The included mathematics are led by ideas and not so much by computational needs that a further revision to that end is needed. That is what the item numerics stand for. But in the described modelling process the theoretical assumptions are additionally influenced, maybe sometimes even unconsciously, by computational needs. An equally appropriate term for the beginning of a modelling process would be theoretical assumption but the term story underlines the fact that the end of a modelling approach is very often the beginning of a new one.
This paragraph shows the process of modelling whereas the paragraph on the double role of such models gives the epistemic status of the result of such a process. To subsume the discussion above the term *climate model* is to be interpreted as a climate modelling approach.

2.4. Spectrum of climate models

At the beginning of climate research it was unthinkable for one single model to simulate the highly complex climate system with its several subsystems and many timescales, so the concept of a hierarchy of climate models was invented (Schneider and Dickinson (1974)). This concept is much more than a relic of the beginning of computer use. As computational power is still limited and different research questions require different levels of complexity, a spectrum of climate models is essential for climate research. Nowadays the term spectrum is commonly used instead of hierarchy as there is no hierarchical order. More complex models are not generally superior to simple ones and vice versa, complexity is thus not a criterion of quality for climate models. Moreover there is no general criterion as there is no best climate model but the best climate model for a certain research question. Nevertheless there are bad climate models for every purpose, although their fault does not lie in conceptual discrepancies but in the modelling practice. These problems are the topic of the following chapters, especially chapter II.

The simplest possible approach to constructing a model of the climate system is to balance, or very nearly balance, the incoming energy in the form of short wave electromagnetic radiation to earth with outgoing energy in form of infrared electromagnetic radiation from earth. In all climate models this balance is considered, but even this energy balance alone already offers a so-called-zero-dimensional climate model, where the dimension represents the number of independent variables of space. Any imbalance in the energy balance results in a change in the average temperature of the earth, which consists of the effective blackbody radiation temperature of the earth plus the temperature resulting from the greenhouse effect.
Simple energy-balance models without any spatial resolution are only thermodynamical models. If supplemented by convective energy transport in the atmosphere (vertical dimension) or by single grid boxes entailing the equations for different latitudes and longitudes (horizontal dimension) they are already dynamical models where the dynamic is represented within the continuity equation. Models composed of two or more boxes can also be used to represent oceanic flows. Models can be built for every component of the climate system, with varying resolution of climatic processes, which in turn can be coupled so as to represent the whole climate system and the interaction of the components. The most complex models available are general circulation models which discretise and numerically solve the full equations for mass and energy transfer and radiant exchange. The more complex a model is the more unknown data and processes must be dealt with, i.e. the more parameterisations are needed. Coupling of spheres is done via calculating energy and mass exchanges between them, with the output of one sphere serving as the input for the other at every time step in the process.
All of these model types have advantages and disadvantages concerning the representation of processes of the climate system and the need for computational power, thus the whole spectrum of climate models is in use in modern climate science. The most important types, in research and philosophical debate respectively, are briefly explained below. It is important to underline that the hierarchy of models is not a historical de-

velopment. Most kind of models started their development at the beginning of climate research and all of them are constantly being improved.

The simplest models are zero-dimensional models. The following is an example from (Ghil (2001)) of a zero-dimensional atmospheric model. The atmosphere consists of only one equation linking the surface-air temperature to changes in the global radiative balance.

$$c\frac{d\overline{T}}{dt} = R_i - R_o \tag{2.4}$$

$$R_i = \mu Q_o\{1 - \alpha(\overline{T})\}, \qquad\qquad R_o = \sigma(\overline{T})\overline{T}^4.$$

R_i refers to the incoming solar radiation while R_o represents the outgoing terrestrial radiation. The heat capacity c is that of the atmosphere combined with that of the upper layers of the ocean depending on the relevant timescale. Q_0 is the incoming radiation at the top of the atmosphere, σ the Stefan-Boltzmann constant, and μ is a factor for the amount of incoming radiation which is 1 for present day conditions. α and m represent the planetary albedo and a greyness factor in dependence on $\overline{T}$, which is the global mean temperature. Such a model reproduces a future evolution of global mean temperature. It reproduces approximately the same global mean temperature a general circulation model can do, quite well if the constants are adjusted accordingly. Models of this complexity are easily solved analytically, as they do not represent the nonlinearities of the climate system. As they are of such low complexity they can be used to test more complex models. Those models can be expected to reproduce similar mean temperatures, because the zero-dimensional model depicts the basic thermodynamics that is to be represented by higher models too. In section 4 it will be discussed whether such testing of complex models is one of the very few methods to gain confidence in a climate model. To identify the thermodynamic development of the climate system it is important to simulate it alone, because otherwise the influences of thermodynamics cannot be isolated. As only thermodynamics are understood sufficiently this is very important.
The more complex type of an Energy-Balance-Model (EBM) is of one dimension because it additionally represents the horizontal heat fluxes. The simplest possible way to develop such a model is to add a term referring to those heat fluxes to equation 2.4. Alternatively, the height dependence of the radiation can be simulated using 1-d models including vertical convection.
The next more complex type of model is a combination of a radiative-convective (RC) model and a one-dimensional EBM. Another class of 2-d models is achieved by the extension of EBMs to zonal and meridional heat transport. Both kinds of two-dimensional models are developed with reference to the earth's dynamics or thermodynamics respectively. Those models concentrating on statistical dynamics (SD) simulate especially the interactions of stationary atmospheric waves and travelling weather systems.
The top of the hierarchy pyramid (see Claussen et al. (2002)) is occupied by general circulation models (GCMs), which aim at representing all relevant phenomena, including all spatial dimensions. Atmospheric GCMs typically consist of a grid resolution from 3° to 5° and up to forty vertical layers becoming broader at increasing height where the

atmosphere becomes thinner and less turbulent. That means they fragment the earth into boxes of 150 to 500 km length with height ranging from some meters above the surface to several kilometers in higher layers, with the model equations being computed at each node. For each grid box the equations of state and motion are computed for radiation and convection and the dynamics of the system. Modern AGCMs also represent aspects of land surface and oceans to compute their influence on dynamics. A GCM can be made more complex by interactively coupling GCMs of different climatological subsystems, typically atmospheric and oceanic GCMs.

Besides this straight-forward hierarchy a second branch of complex climate models has been established during the last decades. Especially in the non-scientific community GCMs are regarded as the most sophisticated and trustworthy models available, as they aim at representing climatic processes as precisely as possible. This is true, but besides the fact that such models require enormous computational power problems arise on the conceptual level. As it is possible to truly model only climatic processes that are measured and more or less understood, the level of complexity of a GCM varies according to our level of research. This leads to the somewhat paradoxical situation that some small-scale processes are simulated but some highly complex ones on bigger scales are not. They are only reproduced as parameters. In the view of many climate scientists this lack of consistency may lead to at least as many errors as GCMs prevent due to their high resolution.
To handle this lack of consistency some research centres concentrate on developing Earth System Models of Intermediate Complexity (EMICs). These models can but need not represent all spatial dimensions, but they use coarser grids and fewer vertical layers as GCMs. Therefore they require less computational power and can be used for simulations on longer timescales. Their conceptual advantage is that they aim at consistency, which in this case means preventing errors due to wrong emphases on small-scale climate processes.
A different approach to handling the conceptional problems regarding GCMs is the attempt to interpret their output in terms of the understanding gained from low-dimensional models or, most importantly, independent of models.

In addition to the hierarchy or spectrum of models, as this term also includes EMICs, a classification distinguishing three model classes is given. Tutorial models are used as 'geo-cyberspace toys' (Schellnhuber and Kropp (1998)), which are invented to study phenomena and mechanisms of various systems without claiming to mimic any system of reality. A very famous example is the daisyworld model by Watson and Lovelock (1983). To underline the gaia hypothesis that the earth could be understood in analogy to a huge creature, they modelled a planet full of black and white daisies only. Black daisies absorb sunlight whereas white ones reflect most of it, so - depending on start temperatures - a temperature balance and thus a constant population of black and white daisies establishes. It was the aim of this model to show the self-regulating mechanisms of a planet.
In contrast, conceptual models as a second model class try to simulate the essence of

aspects of the climate system but without representing the dimensions of the real world processes. Apart from GCMs and perhaps some EMICs, all climate models fall into this category. Conceptual models vary from zero-dimensional box models to relatively complex more-dimensional models. The reduction of essential features often allows the reproduction of important mechanisms like ocean circulations or atmospheric patterns on large scales. Only analogical models, the third model category, try to simulate the climate system's dynamics as accurate as possible. Their purpose is the simulation of past or future climates in order to make quantitative statements. They are not useful for analysing basic mechanisms as they need enormous computational resources and produce a huge amount of data. Yet, models of this type are the only ones that enable us to see effects and mechanisms not genuine to a special subsystem but instead evolving from the interaction of different climatic subspheres.

All types of models within the spectrum vary very much within their category. The spectrum only sketches the most important differences. The greatest variations in conceptualisation are due to fundamentally different aims, either simulating an equilibrium climate to understand principles or model transient climatological behaviour to understand climate change.

One of the more advanced problems of all earth system modelling is the influences of human activity on the climate system in the future, which includes not only carbon dioxide use but more complex economic developments. Attempts to cope with them are done via integrated assessment modelling, which is in principle conceptual but could also be considered as a fourth type of modelling. The characteristic trait of this branch of climate science is its interdisciplinary approach, including modelling concepts from economic and the social sciences in climate models. Only such models will be able to provide the broad basis of the scientific understanding of the 'world system' that is needed to make meaningful normative climate politics. A very deep understanding of the interactions of the climate system, economy and other globally relevant factors is the precondition for effective political interventions. Such integrated assessment models do not include highly developed climate system dynamics as GCMs do, but are rather energy-balance models or are based on aggregated GCM output. This is necessary as their purpose is mainly to analyse possible impacts of different climate scenarios on various socio-economic future settings. Although the term assessment implies their aim is the comparison of several scenarios to discover future trends, their outputs do not lead to statistically significant predictions. Due to this fact their basic theories are much less sophisticated than the basis of pure climate models. All deficiencies of climate modelling are also found in economic climate models and only become worse as the problems of economic modelling are added. As the aim of integrated assessment models is not to understand the climate system but to analyse the impacts of climatic changes with an emphasis on socio-economic dependencies, they are normally not considered in the hierarchy of climate models. This is acceptable in the sense that they do not provide a new level of complexity, however, due to their increasing importance, especially in public discourse and politics, they should be included in the spectrum of models.

A big advantage of simple modelling approaches is, as also sketched above, the knowledge of the primitivity of such models that makes it easier to validate the models and use their output. A zero-dimensional model that is able to simulate today's and historic energy balances can be used to gain knowledge about the earth's energy budget under changing conditions. Everything else the model might provide as well will not be taken to have any scientific value. But of course, the statements gained in that process always have to be understood in light of the underlying assumption: 'if nothing else on smaller scales happens'. These simple models do not seem to reveal very large epistemic problems as their limitations are so overwhelming that they are not used for prediction but instead to understand general processes in the climate system. Conceptual models and models of limited complexity are normally not considered to represent reality but to isolate aspects of climate processes.

An exception to this understanding - that they are not suitable for predictions - are the EMICs, which are designed taking into account the demands of politics, which explicitly requires predictions of climate development in view of specific socio-economic development. Particularly concerning the atmosphere EMICs have only a very weak dynamical basis as they mainly use parameterisations of zonally averaged fluxes in terms of average temperatures and winds (see Claussen et al. (2002), McGuffie and Henderson-Sellers (2001)). Therefore they rely heavily on tuning parameters which may lead to unreliable predictions of the earth's climate on the whole.

Part II.

Epistemic problems of climate modelling

3. Perfect climate modelling is impossible

In chapter 2 it was shown how important knowledge of the climate system can be gained without even having invented a single climate model. Therefore, even if climate modelling was a completely impossible undertaking, there would be evidence of climate change and knowledge of the climate system. However, it is difficult to draw a dividing line between those modern scientific findings dependent on modelling the climate and those independent of it. The global data sets which are available today are so enormous that it is impossible to view and categorise them without the use of computers. In fact, their analysis is not meaningfully possible without the help of reanalysis and interpolation models. According to Edwards (2001), even their collection requires interpolation modelling and model-based reanalysis of data sets. Thus climate science strongly depends on climate modelling. Although reanalysis models are not intentionally the same as climate models, many of the uncertainties to be found in climate modelling are also problematic for reanalysis models. As Edwards (2001) defined the problem, "without global data sets, modelers could neither validate nor parameterise their models. Without computers and satellites, uniformly gridded global data sets could not even be created, much less manipulated. Without numerical weather prediction models and GCMs, these data could not be understood." But the purpose, the object of a reanalysis model, is different from that of a model to simulate future climate development, to understand processes of the climate system or to predict future climate development.

The references above show the problem as it evolves in real life, but on a conceptional level the difference in purpose is quite evident. If the intention behind the analysis is the prediction of future climate development, based on knowledge of past and present behaviour of the climate using empirical statistical methods only, we will not talk about modelling but of extrapolation with the help of regression equations. If instead the purpose is to make projections based on the physical, chemical and biological processes we believe to govern the climate system, the result is a climate model. Thus the intention behind climate modelling is the representation of the climate system based on the "first principles" of the climate system. Another way of expressing the different approaches to studying and predicting climate system development is to talk about practical approaches when using statistical extrapolation and about a theoretical approach if the basis are governing equations. Purely empirical or statistical models based on present-day observations cannot be expected to perform well for climates apart from those existing today.

The criterion for considering a computer-based approach in a climate model must remain unclear to some extent because of the mixing of both methods in most climate models,

especially where parameterisations are concerned and on higher levels of the hierarchy of models.
This is insofar important as observational data are extremely important for climate modelling. They are used to gain knowledge of the climate system, to initialise models and to validate the modelling approaches. Thus the interdependency of observational data and model data is crucial to an understanding of climate models. This closeness reveals many aspects that make climate modelling a tricky undertaking, far from representing the world as it is.
Before I present the different arguments against perfect climate modelling in the following sections, some more terms must be defined.

Apart from comparing modelling results with observational data the biggest part of model evaluation is done via model intercomparison. To understand its meaning it is important to distinguish between simulation and experiment. To draw this line is not a simple task because a definition is never given explicitly but always assumed to be implicitly clear. While it is commonly agreed that a simulation is a single model run with fixed parameter values, boundaries and initial conditions, there is no such agreement concerning experiments. In the context of the IPCC a model setup with fixed initial and boundary conditions but variable parameters seems to be the best description of an experiment. At least "experiment" is often used in this way. But strictly speaking it is not possible to conduct experiments using models, at least if defining experiment in the common context of the natural sciences. An experiment is the creation of a controlled environment[1] where an exactly defined cause is placed in a certain isolated environment, via interconnection with the system, changes in system properties, application of fields etcetera. The observables of this system can now be analysed as their changes can only result from the prescribed cause, causality assumed. Postulated coherences can thus be tested and falsified. In addition, probabilistic statements can be gained via the reproduceability of the experiment.
In contrast to that, a simulation is an imitation of an already known causality. Within a simulation all implications of this causality law, including the inapparent ones, are made explicit. Its effects can be seen in all observable and even non-observable variables, which makes it such a good tool.
The comparison of simulations and measurements of real world data can also be interpreted as an experiment of some sort but with the difference that the observed systems are very complex and separability and repeatability are not guaranteed. Thus simulations are used to make assumptions in a scientific frame where no consistent theory exists, whereas classical experiments as described above are used to test theses evolving from theories or theoretic models. In climate science theory building and derivations of theses or prognoses are sourced out to computers as an experiment of some sort but with the difference that the observed systems are very complex and separability and repeatability are not guaranteed. Thus simulations are used to make assumptions in a scientific frame where no consistent theory exists, whereas classical experiments as

[1] under the precondition of separability of physics via scales and exclusion of interconnections

described above are used to test theses evolving from theories or theoretic models. In climate science theory building and derivations of theses or prognoses are sourced out to computers. Nevertheless "experiment" in the context of climate modelling is often used synonymously with "parameter study", which has fixed initial and boundary conditions but varying parameters.
Of course, all measurements in physics are systematically inaccurate due to the instrument calibrations, data reduction and addition of artificial distortions owing to laboratory conditions. Enhanced computer modelling of data can provide a reduced and interpretable data set as good as data from traditionally controlled experiments, or even superior to them, but its setup and interpretation requires different skills and is accompanied by different systematic errors. Systematic errors in a complex, nonlinear system are normally harder to detect and eliminate then those in classical systems.

3.1. Fundamental uncertainty versus model uncertainty

In the following sections in this chapter different sources of uncertainties are discussed and analysed with respect to their consequences for comprehensive modelling approaches to predict future climate development.

These uncertainties are of two different categories on an epistemic level: fundamental uncertainties and model uncertainties. While the latter uncertainties are to be overcome by perfect climate models the former are not. In fact, both categories of models constrain the quality of modelling results to a certain extent, as perfect climate models are practically impossible. In principle, however, model uncertainties could be avoided by unlimited computational power, resolution of the Navier-Stokes equation and the availability of several different and comprehensive reanalysis data sets. Fundamental uncertainties result basically from fundamental limitations of the human brain or the world and are thus unavoidable.
While model uncertainties are manifold I could only detect two fundamental uncertainties constraining climate modelling (section 3.2.1 and section 3.2.2). The causes for model uncertainties are to be found at all stages of the modelling approach and are thus for the modelling practice equally important and even bigger constraints in actual modelling approaches.
In the context of this thesis many problems concerning climate models are discussed. Some of them are really important and influence the quality of climate predictions in a strongly negative way. Nevertheless they do not prevent meaningful climate predictions in principle, which is also true for fundamental uncertainties. This is due to the fact that basic aspects of the climate system are understood and understandable, namely the thermodynamic basis of the climate system which is part of every modelling approach and not to be questioned, taking model or fundamental uncertainties into account.
But some uncertainties do indeed prohibit the meaningful prediction of certain climate variables, at least for the time being.

The implications of theses categorical deficiencies in climate modelling are further discussed in the context of probabilistic modelling in section 6.

3.2. Fundamental uncertainty

There are two fundamental uncertainties in understanding the climate system which result in fundamental uncertainties in climate modelling. These are the nonlinearities of the climate system and the fundamental incompleteness in climate system observations which will be discussed in the following two subsections.

3.2.1. Fundamental constraints in understanding the climate system: nonlinearity

Being able to model something normally implies an understanding the model system as well as the modelled system. For the climate system and its modelling approaches both are impossible. There are two fundamental and principal constraints that prohibit the exact computation of the climate system, its sound application to a model system, and the comprehension of its relevant system components. Furthermore there are other constraints: technical, physical, and philosophical.

Understanding

These last statements reflect an ambiguity in the meaning of the term "understanding". We can understand a written description of the known components of the climate system provided we are familiar with the language. We can also understand a system of differential equations, i.e. we understand a perfect model system described with the help of these equations, whereas the precondition is the knowledge of all system properties and possible states. Most real life systems will never be understood in such a way. We can only describe them as model systems as every verbal description refers to a model system as well as a description using sets of equations. Since we do not know all system properties our modeled and computed systems are imperfect. Only model systems are ready to be understood. If understanding is understood as computing this is a fact that is easy to see but it is also true if understanding just stands for description since it is not possible to describe a real life system in exactly the way it is in reality. There will always be small processes or properties that escape our attention or are simply unobservable. Every understanding will be the understanding of an approximation which means: understanding is simplifying.
All of these deficiencies in understanding are especially important in climate physics where the underlying real life system is a very complicated nonlinear system.

Approximation As long as the approximation is a good one it is no principal constraint to understanding since we can only truly understand approximations instead of

real life systems. A good approximation is one that includes important processes and time scales. The crux of the matter is therefore the identification of important processes. If understanding in the form of computing is desired, processes must not only be identified but also be mathematically described.
A really good approximation is a goal that will never be reached in climate science, at least not for the system as a whole. This chapter will deal with the most important reasons for this statement.
Until now an understanding of the climate system has not even been attempted for good reasons. One always just tries to understand certain aspects of it, but there are fundamental limitations to even this attempt.
That means we in fact simplify whenever we describe parts and aspects of the climate system as a system itself. If this works we did indeed understand the system in a very weak sense. But it would only be an extreme simplification and thus not even a fairly adequate approximation. This weak understanding is possible if every *working* simplification is a form of understanding. *Working* in this context means that the simplified system is consistent in itself[2]. Considering this, the impossibility of understanding the climate system only holds for understanding in terms of computing and predicting. But as the latter is especially requested by society it is important to analyse why this desire must remain unfulfilled, at least with respect to accuracy.

Nonlinearity

When describing the climate system as explained above by describing the earth as a system dividable into subsystems where different laws hold and several processes and feedback mechanisms take place, do we understand it as a model system? Is the model system a good approximation? Typically the quality of a model system can be checked while carrying out experiments in the real system but due to its uniqueness and integrated whole this is not possible in case of the climate system. Besides that, the fact that the climate system is a nonlinear system makes it difficult to figure out the important processes as they are often interactive feedback mechanisms between subsystems and different scales. Nevertheless, our "resort to 'geo-cyberspace', where virtual copies of the ecosystem can be exposed" to virtual experiments (Schellnhuber and Kropp (1998)), does not equal a computer game, as we have observed data to compare with virtual ones. This permits an evaluation of our approximation. As the data are also used to develop a climate system model such an evaluation is not without problems but definitely possible. But as we as human beings are unable to think in a nonlinear way we will not be able to fully grasp a nonlinear system. Thus its nonlinearity is one of the fundamental constraints in understanding the climate system.
In addition to the above, this notion of understanding is such that the fact alone that we do describe the climate system is a form of understanding which avoids the somewhat artificial differentiation between real world and the climate as a system. But this is of course not the form of understanding needed to predict the future evolution of a system.

[2]This is not a claim for consistency with respect to established theories to which Feyerabend (2010) for example objects, but for internal consistency. However, in chapter IV it will be argued that this type of consistency is also desirable in climate modelling.

Not being able to understand a system, as said before, prohibits its computation. Inevitably this reveals uncertainties in handling models as well as the original system. Our lack of understanding can be differentiated into three categories of uncertainty. This is a more complex differentiation of uncertainties than the very rough one I give in section 3.1. It sets a different focus in terms of understanding and accounts for the relevance of human kind as part of the climate system. Basically the first two categories correspond to fundamental and model uncertainty, respectively.
Using the terminology of Schellnhuber and Kropp (1998)[3] there is *removable cognitive uncertainty*, *irremediable cognitive uncertainty* and *voluntative uncertainty*. The first category entails those gaps in our knowledge that are due to unrevealed facts but which can principally be obtained with the help of better observation methods and modelling skills. The second class of uncertainties contains final uncertainties, i.e. all aspects of the climate system that are principally determined by natural laws but due to their nonlinearity and complexity are not computable for a given point in time. These types of uncertainty are fundamental and hold independently of civilization on Earth. But the last and also irrevocable uncertainty is caused by the 'freedom of will' that all actors exhibit. The impossibility of predicting the behaviour of nations, companies and single persons is a 'fundamental indeterminacy of future' climate system dynamics.
The fundamental indeterminacy of the future is not limited to climate physics and only becomes relevant insofar as this discipline is largely preoccupied with prediction. It is no fundamental constraint in understanding the system. Thus I do not count it as a fundamental constraint in the context of the climate system and its understandability. But it is indeed important for conceiving climate prediction, climate change, and prediction making in principle.

The nonlinearity of the climate system is accompanied by its enormous complexity, which makes it even more disturbing. Many climatic processes originate in internal feedback mechanisms of the system's dynamics that interact in a nonlinear and thus unforeseeable way. Nevertheless the climate system as a physical system can be described by mathematical equations. As explicated in section 2.1 in more detail, the basic equations are the equations of motion, the first law of thermodynamics as the principle describing the conservation of energy, and the continuity equation representing the conservation of mass. The equations of motions are represented via the Navier-Stokes equation which arise from the application of Newton's second law to fluid dynamics. There, in the dynamics, the nonlinearity is represented in the equations. The Navier-Stokes equations are nonlinear partial differential equations, where the nonlinearities drive the turbulences which are an essential driver of the climate system. But because of them it is impossible to solve this equation analytically for the climate system or parts of it. Moreover, not only the solution but also the correct setup of the equations is impossible because of the enormity of influencing factors, such as e.g. the highly scale-dependent friction. Thus many assumptions are necessary to simplify the Navier-Stokes equation and make it numerically solvable.

[3] The following citations in this paragraph are also taken from Schellnhuber and Kropp (1998)

Furthermore, some equations are not even known. It is for example impossible to give the equations of motion concerning sea ice and also shelf ice is only included into modelling using relatively crude approximations of equations of motion, namely the shallow shelf approximation and the shallow ice approximation (i.e. Mangeney and Califano (1998)). There are many such subsystems of the climate system that are at least mathematically not adequately describable.
Nonlinearity is a limitation to understanding insofar as 'understanding something' is defined as 'the ability to compute it'. But our inability also to setup the equations shows that the nonlinearity limits even our intellectual ability to grasp the single processes as it principally prohibits a knowledge of at least their quantity, thus not only computation but even more basic forms of understanding the climate system are limited due to its nonlinearity.

3.2.2. Fundamental constraints in understanding the climate system: observational constraints

The second fundamental constraint in understanding reveals itself again. In order to understand or describe the climate system completely all factors that influence the energy balance must be known qualitatively and quantitatively. The latter in particular is a very difficult task, as it would not only be necessary to know all components and mechanisms but to have measured them in detail. The gaps in knowledge of quantities are fundamental limitations to our understanding of the climate. An example are our lacks of knowledge of the magnitude of the poleward atmospheric heat transport, constraining the understanding of large and small-scale processes related to the zonal heat balance which is a key driver of the climate of mid latitudes.
Regarding the feedback mechanisms discussed above it becomes obvious how limiting undetected or poorly understood climate mechanisms can be. In addition to these feedback mechanisms scientists lately published an overview of possible climatic tipping elements (Lenton et al. (2008a)), which are discussed in section 3.8. The earth's climate system is a complex, nonlinear system, which means that even small disturbances in the system, as for example caused by greenhouse gases, may lead to large effects. Within the climate system there are some processes and regimes that are especially sensitive to climatic changes. These so-called tipping elements can be disturbed in such a way that they may 'tip' into a completely different state. As these elements are related to important processes within the climate system, their tipping over will have serious impacts on the system as a whole. This includes the possibility of an irreversible shifting of the system into a completely different state. Most tipping elements are likely (Kriegler et al. (2009)) to play an important role if the global temperature increases by more than two degrees.
The existence of such phenomena shows that a qualitative and quantitative comprehension of climate processes is required in order to prevent further dramatic anthropogenic manipulation of the climate. But it also reveals the second fundamental limitation to our

understanding, which is the restricted amount of observational data. This is especially a limiting factor concerning the oceans and upper troposphere and stratosphere where direct measuring is difficult and only intermittently possible at all. Observational data are not only needed to investigate mechanisms of the climate system but also to tune and evaluate climate models. Thus the impossibility of measuring data on a global grid sets a fundamental limit to our understanding and to possible models.
To do this properly the purchase and interpretation of climate data is needed. Before being able to simulate the climate system a huge amount of data is necessary to tune and validate climate models. This task is especially difficult concerning prehistoric climate data and oceanic and atmospheric data or data from uninhabited landscapes. Besides the difficult observation the preparation for modelling purposes of these data is another challenge not possible without computer simulations. Only via modelling can the data be made comparable to modelling data and without complex computer programs the amount of data would not be possible to grasp at all. As it is not possible now and for the foreseeable future to measure all the needed variables on sufficient climate system points, an analysis of the observed data to interpolate missing data is necessary. This happens with the help of so-called reanalysis models. If carefully done the practices do not necessarily lead to equal systematic errors in the reanalysed data and modelling data. But due to this, climate physics must cope with an additional source of error compared to observations in classical physics.
However, this is no replacement at all for truly measured data, as an interpolation can only be as good as the knowledge of the mechanism. That in turn can only be good if enough data exist. Thus the amount of observed data must be seen as a fundamental constraint: sufficient measurement of all variables is impossible.

3.3. Boon and bane of parameterisation

The wide spacing of even the most ambitious regional GCMs (see section 2.4) creates a problem of scale, as several important climatic processes occur on scales smaller than model grids. The most famous example of such sub-scale processes are clouds. They normally spread across a few kilometres whereas the typical length of a GCM grid box is a few hundred kilometres. Furthermore, clouds highly influence the energy balance of the atmosphere as they reflect or absorb radiation, depending on the cloud type. The whole scale of convective motions is basically beyond the grid scale of models but on that scale nearly all processes determining precipitation take place, which in turn is not to be neglected when considering temperature development. Thus individual clouds and other convective motions are important factors considering especially the impacts of green house effect and temperature development but they are not resolvable in climate models. Clouds are the most uncertain effects but their influence is small on a global scale. Nevertheless clouds and convective-scale motions are essential in GCM modelling. Other examples of processes in need of parameterisation are the transfer of radiation into the atmosphere, transport processes in boundary layers, surface energy exchanges and subsystem processes of neglected subsystems (Hack (1992)).

To represent such processes they are modelled as a parametric representation, a parameterisation. The parameter which gives such representations their name is not a fixed value, one for cloud and zero for clear sky, but rather a proportionality factor linking the cloud cover to model computed variables like temperature or humidity. To find such a factor several climatic variables resolved by the grid and their relations to non-resolved variables determining or influencing the cloud cover are analysed statistically in observational and reanalysis data. Furthermore, relations known from theoretic fluid dynamics are taken into account. The temperature and humidity over a grid box can now be used to predict the average cloudiness of that box for given temperature and humidity values. But this empirical method is not adapted for the prediction of individual clouds (Schneider (1992)).
Thus parameterisation tries to express the contribution of sub-scale processes to the time evolution of resolved motions on the one hand and on the other hand as functions of the large-scale fields (Hack (1992)). The connection of resolved variables and non-resolvable process variables is normally not a simple linear proportionality but a complex interconnection. Thus the parameter to represent this dependency can also vary from very simple to highly complex. But a fundamental characteristic of parameterisations is that the higher the level of physical sophistication, the higher the computational capacity needed for the parameterisation technique in the model simulation process.
With the help of parameterisations it is thus possible to simulate the effects of small scale processes on large scales. Modelled large-scale processes, theoretical knowledge of analogue processes, and observed correlations in basic variables lead to simulated large-scale effects of small-scale processes.

Figure 3.1 gives a simple graphic to show a parameterisation in principle. T_{t+1} and T_t symbolise observables resolved within the modelling approach. The graph f characterises the relation between T_t and T_{t+1} in the time evolution given by the physical equations the model includes. The dots stand for observed data points. Obviously the theoretical model output does not fit to the data. They would better be described with a function similar to regression line g. If a modeller observes such behaviour when comparing model physics and observed data it is a sign of an unresolved small-scale process that affects the large scale of model output observables. To better simulate the large-scale effect the unknown small-scale process can be parameterised with the help of a map, shifting f to g, displayed in figure 3.1 as arrows $p(T_t)$. If the parameterisation is successful the next time step of the model can be calculated fitting the observations around g without knowing the physics or even the scale and sphere behind p. The next step is then:

$$T_{t+1} = g(T_t) = f(T_t) \circ p(T_t);$$

The $\circ$ symbolises the fact that all mathematical operations are possible in this place. Parameterisations need not be motivated by physics because the modelling approach deliberately omits a known process. Parameterisations often only close a gap between modelled equations and observed data.

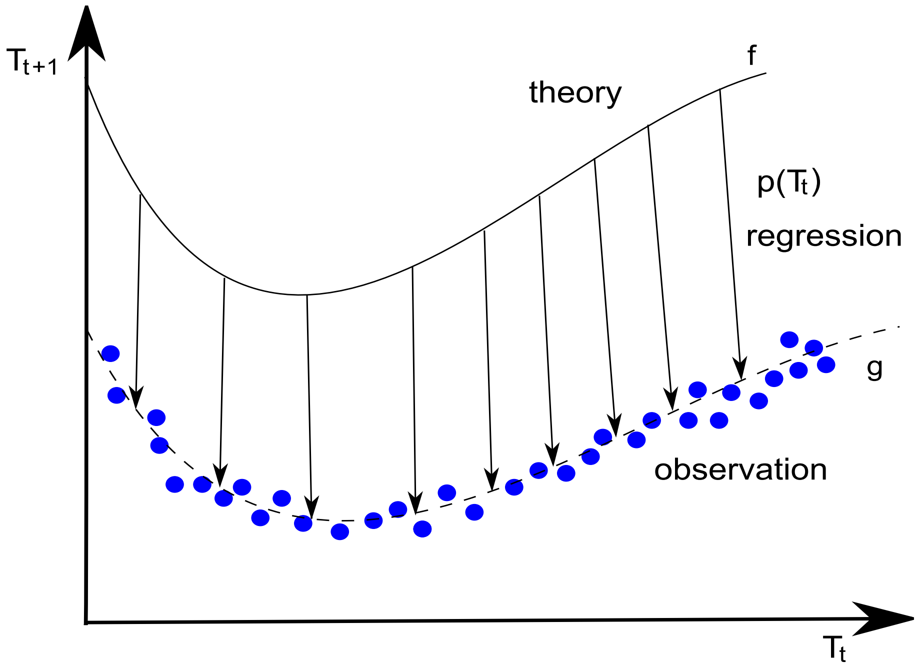

Figure 3.1.: Parameterisation in principle. T is an observable in time evolution; f stands for its theoretic relation whereas the blue dots with regression g give the observations. The arrows $p(T_t)$ denote the parameterisation.

Another characteristic of parameterisations is the need for independent calculations of every single parameterised process for every single vertical column on the basis of the large-scale state information for that column. For serial computations this is a very time and resource-consuming task, but the increasing use of parallel computing environments fits to such tasks.

Nevertheless an ideal model would be a model without fixed parameters apart from continental topography. Without fixed values besides the initialisation values the internal dynamic generated by the model alone from the represented physics of the climate system would compute all variables. Nowadays even the most sophisticated GCMs depend on fixed parameters somewhere in every represented physical process. Thus *the* problem of climate modelling is parameterisation, which is the solution to the problem of scale. All modelling attempts rely on the so-called "closure approximation". This is the postulate that "small-scale processes can ultimately be represented accurately in terms of the large-scale variables available to the models" (Edwards (2001)). To what degree this approximation violates reality is relatively unknown to us as our inability to model small-scale processes corresponds with our lack of knowledge of their importance for climate development.

The problem of scale is the central problem of climate modelling, but there are several other problems that result from our lack of understanding as well as from the trade-off between complexity, resolution and consumption of computational power.

Even if they are referred to as primitive the equations in principle represent all important aspects of the climate system. Their complete consideration for all spatial directions on small grids consumes enormous resources. And these equations alone do not build up the climate system. All the processes that are not represented here have to be considered as well. Thus, depending on the desired complexity, a huge amount of parameterised processes must be calculated as well. Therefore many models do not solve the whole set of equations for all spatial dimensions in the atmosphere and the oceans but replace or totally omit at least one dimension. Also the whole ocean or the whole atmosphere respectively can be replaced by parameterised incoming fluxes to the remaining sphere. Looking at the model spectrum it becomes obvious that the omission of a complete set of equations is also an option to simplify the model and accelerate the simulations. Studies that only represent the energy balance normally do not consider the equations of motion.

But of course all these simplifications are paid for by the 'realisticness' of the model. The problem of such simplifications results from the interdependency of climate processes and our lack of knowledge about it. The modeller knows which part of the climate system is deliberately poorly represented, but what he does not know is which other processes are affected by this simplification.

Uncertainties are of course nothing entirely unknown to physics as well as working with already falsified theories and models in the classic meaning of this word[4]. And the work with assumptions and approximations known to be incorrect is also a method often used in the history of physics. The most famous example of a theory known to be false in terms of being limited is Newton's mechanics, which is still in use even if falsified by quantum mechanics. But in this case the Newtonian way of calculating and predicting is rescued for marginal cases of quantum mechanics which includes almost every mechanical process beyond microscopic scales. Nevertheless the application of Newton's physics is always an approximation, as normally point masses are the replacement of real world bodies.

Even if point masses are theoretical objects the situation is different to that in climate modelling. We know that planets are bigger then point masses and we know with comparably high accuracy to what degree. Additionally we do have a theory about real bodies even if we cannot compute it. However, in climate modelling the situation is quite the opposite. We do not have a theory about the processes we replace with parameters in our models. In some cases we have some basic ideas but often we only know the scale of the processes. The magnitude of the parameter is calculated numerically, by trial and error, resulting in a small set of parameters to represent this one special process. There are commonly accepted parameterisations widely used in different modelling processes as well as specific parameterisations for specified models. As the former are most important for this branch of science, finding new parameterisations is one of *the* achievements in climate model development. Processes are not replaced by simplified

[4]Section 3.7 is dedicated to a detailed analysis of the problem and the meaning of falsification in climate modelling.

theoretical models where the simplifications are known and can be evened out to some degree if necessary, but rather are replaced by numerical schemes and parameters which are theoretically based or only motivated by the fact that they "work" in modelling setups. Normally we do not know how realistic they are as experiments are impossible to realise. Furthermore, the theory frame is not computable, at least for GCMs grounded on the Navier-Stokes equations or the primitive equations (see section 2.1.3).
Experimental measuring and theoretical predictions of certain values are never identical in classical physics, which is due to imperfect measuring conditions. Even such a discrepancy is not existent to the same degree in climate modelling. As measured data are not used directly to tune or validate climate models but edited in reanalysis models to be comparable to model output, the modeling conditions of reanalysis models modify data in a different way than "normal" measurement devices do. A reanalysis model suffers basically from the same need for parameterising important climate processes as prediction models, thus most of the general problems apply here as well. At least they entail systematic errors from the computational setup that add up to biased measuring. The basis of reanalysis models are numerical weather prediction models, fed with observed data. After one time step of integration the output data are compared to observed variables and changed accordingly. This is possible as numerical weather prediction models use statistical modelling approaches which allow for probabilistic variation of initial and boundary conditions, as well as development paths of the model. This fact guarantees at least a considerably large distance to deterministic GCMs.
Nevertheless agreements between climate model output and reanalysis data may partially be due to equally biased modelling approaches to generate reanalysis data on the one hand and the climate model data sets on the other hand. The most important source of such biasing is again parameterisation.

3.4. Using and abusing model tuning

Model tuning is not calibrating the climate model and therefore not adjusting parameters that are observationally well constrained. In contrast to that tuning is the adjustment of internal model parameters which are not representations of physical parameters or physical parameters without corresponding observations. Especially in economic modelling the term tuning is not often used but internal parameter adjustment is called either validation or calibration. With respect to the traditional use of these terms in the history of science this is a rather misleading handling of words.
Every instrumental setup in the experimental sciences needs to be calibrated before starting measurements to be able to analyse results in comparison to already gathered data, thus to a norm. However, climate models "cannot be meaningfully calibrated because they are simulating a never before experienced state of the system" (Stainforth et al. (2007)). Even if intended to simulate past climates, the model climate system is necessarily extremely different from real climatic systems of the past. It is a different nonlinear system. Calibrating a chaotic system would be senseless even if the model

was a perfect model. This is due to the crucial role initial conditions play in nonlinear systems. They are of central importance but cannot be known. Thus calibration is practically impossible and theoretically senseless. A calibration is the standardisation of an experimental setup in comparison to an accepted norm. Due to the very nature of the climate system and nonlinear systems in general, a 'normal' state of the system does not exist. The motivation to tune climate models is thus different. In a manner of speaking the running into a stable equilibrium state of a climate model could be taken as finding a 'normal' system state, but this is quite different from searching for a norm. Nevertheless, tuning is an adjustment of model parameters to achieve agreement with observations, which means that parameter values that are weakly restricted by observations are adjusted to generate good agreement with observations for those variables that are better restricted or even known by observation (Bender (2008)). Parameters are chosen in such a way that the simulated variables are fitted to observational data, or, if not all important variables are known, they are chosen to simulate a process that qualitatively represents an observed process. Or, if observations and understanding are very limited the model is tuned to show more physically plausible behaviour, whatever that means.[5] The more complex the model is and the more different independent and dependent parameters it comprises the more necessary tuning could be for several model parameters. Tuning is a task most important and especially difficult for modelling approaches such as GCMs and EMICs. As general circulation modelling without parameterisations is not possible (see section 3.3) it is just as impossible without tuning. Tuning is necessary to represent observed climatic processes. But tuning climate models is not an easy task in particular because of the interdependencies of different model parameters.
An example of the influence of tuning as well as parameter choice is given by Bender (2008). A GCM is tuned against two different satellite observation data sets to show radiative balances at the top of the atmosphere in agreement with the respective data set. The tuning is carried out through alterations of parameter values in cloud microphysics. As this is a field that is highly parameter-dependent and hardly restricted by observations, parameters may be adjusted independently of 'physicality'.
The latitudinal distribution of surface temperature is hardly affected by the tuning in both attempts. An (non)effect that counts as necessary condition for good tuning because this variable is thoroughly restricted through observational data. In cloud properties the two approaches provide differences, which is not an unexpected behaviour as the tuning was done through cloud parameters. The variations in cloud water paths are significant but small compared to discrepancies in observations.

Figure 3.2 illustrates the example. The first graph shows two different sets of observed data representing the radiative balance at the top of the atmosphere. The second diagram displays the observable 'surface temperature' of the same two different data sets. The discrepancies in the observations give certain degrees of freedom for the parame-

[5]Tuning is a very crucial but subjective part of model development and will be put in a more philosophical context in part IV.

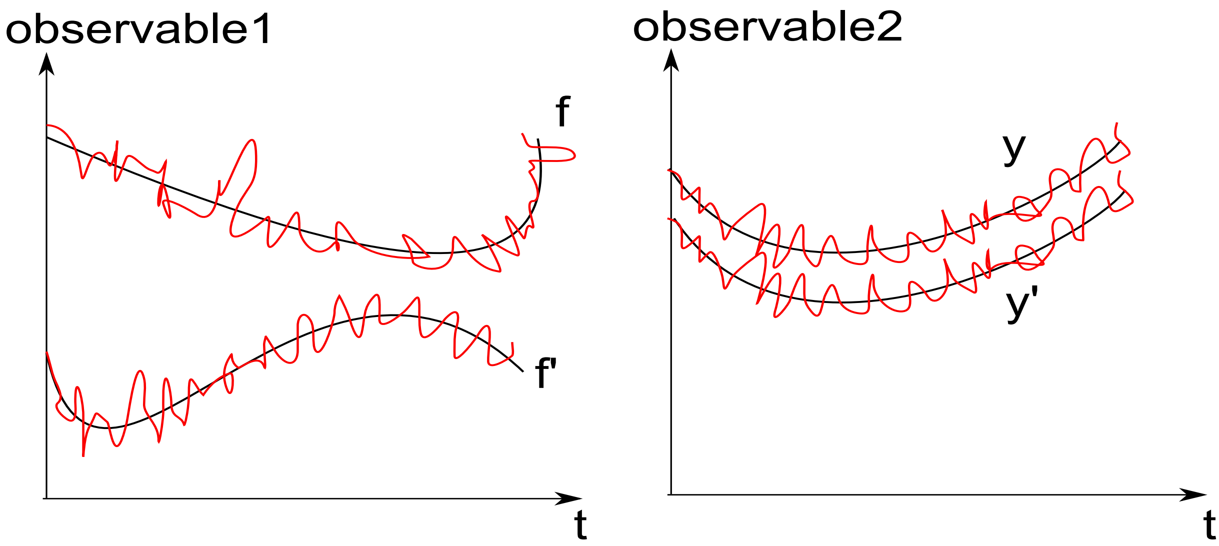

Figure 3.2.: Two different observables of two data sets are displayed. Observable one varies heavily, f and f'; whereas observable two seems to be good constraint by observations (red), y and y'.

ter adjustment within the parameterisations defining the observable 'radiative balance at the top of the atmosphere'. The graphs f and f' can be represented either with parameter set $p1$ or $p2$:

$$[f] = m(x, t, p1); \qquad [f'] = m(x, t, p2).$$

$m(...)$ displays the model dynamics given by physical model equations with initial and boundary conditions x, time t, and the weakly constraint parameters $p1$ and $p2$. At the same time the same relations must also be true for the observable 'surface temperature' represented by the virtually identical graphs y and y'. Thus the parameter adjustment is only meaningful if

$$[f, y] = m(x, t, p1); \qquad [f', y'] = m(x, t, p2)$$

holds as well. That is, the tuning must not effect good constraint parameters and observables. This constrains the degree of freedom in parameter tuning. But there is no decision criterion at all to decide between model set up $[f, y]$ and $[f', y']$. From the point of view of all known physics and available data, both approaches seem to be 'optimal'.

In the two differently tuned modelling approaches the equilibrium of climate sensitivity is calculated. It differs but the difference is of degree not of kind. But in a similar study by Stainforth et al. (2005), where the tuned parameters are not only varied between two different 'optimal' setups but within the range of estimated uncertainty of the tuning parameters, the resulting range of climate sensitivity is of several degrees Kelvin, which means that different tuning of cloud parameters highly influences key variables of climate change in GCMs. And it is impossible to say which set of parameters is best.

Tuning manipulates physics

Considering this example an apparent problem of tuning is its influence not only on the varied parameter but on all directly and indirectly dependent parameterised processes. An obvious impact of this fact is that the model output cannot be used to make strong statements about the processes only represented due to tuning. This is a straightforward implication but entails many consequences and is not always considered when analysing model output. In particular, it is hard to detect whether a certain process shows stability due to parameter tuning or for dynamical reasons. For example, an ocean model that is tuned to have a thermohaline circulation(THC) is likely to have such a THC more stable than real oceans, thus the behaviour of the modelled THC cannot be taken entirely seriously. This is a fact scientists developing the model are more or less aware of but it is not published or discussed within the peer group, which raises profound doubt whether scientists who are only using the model are equally aware of it. The only model description briefly mentioning this problem is given by Weaver et al. (2001) in the description of the UVic earth system Model of intermediate complexity.
An example is Hargreaves et al. (2004) where two ocean diffusivities, a sea-ice diffusivity, an Atlantic-Pacific moisture flux, two parameters controlling wind-driven circulation, and six parameters controlling atmospheric heat and moisture transport, in total twelve parameters are used to tune an EMIC to simulate realistic climate variables. The model is furthermore used to examine among other things the stability of the THC under simple global warming scenarios, which was found to break off in one third of the simulations. The THC is to a great extent driven by heat and freshwater fluxes and dependent on wind-driven currents. Of course these twelve tuned parameters influence everything which is possible to find in the model. In examining and interpreting the model output this fact must be included in model uncertainty analyses but disregarded otherwise. But stability analyses of processes only occurring due to tuning seems overconfident concerning tuning technics. Only those processes also occurring without the specific tuning are meaningful to analyse according to their stability under climate change scenarios.

Considering these aspects of tuning it becomes clear that tuning is a tool to improve model performance but can also effect the dynamics resulting from underlying equations in an undesired way. And therefore tuning is a method that violates the physics of a climate model. The difference between using and abusing tuning is a very small one. And if abused in some way or other the tuning process forces the model to provide the right results but for the wrong reasons. This is in the words of Collins (2007), who argues that a model which has a good present day simulation of, for example, surface air temperature trends, may have it owing to some cancellation of errors in closely related parameters.

3.5. The need for numerical solutions

Numerical algorithms are of interest in two cases: on the one hand if no explicit solution for the the mathematical problem exists, and on the other hand if such a solution exits but is complex to solve and prone to errors. There are also two methods of numerical solutions: a direct one which provides an exact solution in finite time if calculation is infinitely exact, and an indirect or implicit approach to get iteratively better approximations.
The analytical solution of the three dimensional Navier-Stokes equation belongs to the millennium problems of mathematics, as declared by the Clay Mathematics Institute. That is, as yet no exact solution is known. Analytical solutions only exist for special cases, for example in two dimensions.
Thus fluid dynamics which is virtually based on the Navier-Stokes equations is in need of numerical solutions for its problems. There are different methods to arrive at numerical solutions, all of them with advantages for special problems and general disadvantages (Batchelor (1967)). Direct numerical solutions are only possible for problems hardly showing turbulence which is only given on very limited spatiotemporal parts of a climatic sphere. Thus numerical methods in climate modelling are indirect methods, that is, approximations.
Most of the problems in climate science are fluid dynamical problems, as pointed out in section 2.1. The solution of the equations affords a discretisation of the normally continuous fluid. In climate models this is commonly done through the application of grids in atmosphere and ocean. The nonlinearity of the equations requires that either the spatial resolution is very fine or the time steps are small in order to get a meaningful numerical solution. The former is hardly possible due to computational power limits and insufficient data on fine grids to initialise the model (see section 3.2.2). Thus climate models including dynamics and hence numerical algorithms to solve them must be integrated on small time steps to guarantee stable solutions. Again, due to computational limits only short simulations with complex climate models are possible to run. Especially for GCMs this comprises the risk of not reaching an equilibrium state. An equilibrium is reached if all components and parameterisations of the climate model work together so that a stable climate is simulated. As far as all processes are initialised and the model and parameterised feedbacks co-operate as planned several decades of model years go by. The shortness of GCM simulations is thus one of the major problems in handling GCMs.
If a model did not reach equilibrium state before starting the real experiments a 'climate drift' can be observed. Normally such a drift manifests itself in an artificial trend within the modelling results. A prominent example of incorrect modelling results due to a climate drift is unintentionally given by von Storch et al. (2004). They made a run of 1000 years with the coupled AOGCM ECHO-G starting 1000 years ago and ending with nowadays climate. Unfortunately, the model was originally initialised with present day climate data, including a CO_2 concentration of 372 ppmv. But to start a simulation run in medieval times the concentration must be around 280 ppmv. Von Storch and his team adjusted the CO_2 concentration accordingly and started the simulation after a

transition time of 50 years. Their published results showed a strong cooling trend in the first few hundred simulation years. This caused a lot of political trouble as the results suggested a stronger internal climate variability as all other simulations and temperature reconstructions. Politicians used this pseudo-result not to act against climate change as the human influence seemed to be less than it actual is. Osborn et al. (2006) could easily show that these spectacular results are only pseudo-results as the trend was due to a climate drift induced by a much too short transition phase. It was the same mistake one would make if one took a thermometer from a warm room outside and immediately started to measure the outdoor temperature. One would observe that the temperature decreases, but only because the observation was begun before the thermometer was in equilibrium rather than because of real cooling outside.
This example, which is discussed in detail on www.realclimate.org[6], shows impressively model simulations that are too short violate physics.

The need for numerical solutions are manifested in the scale problem. The scales where climatic motions occur range from atomistic to global and from nanoseconds to millenia. The energy transfer interacts between all these different scales and triggers climatic motions on any scale and may or may not lead to large-scale changes. Thus the internal dynamics of the climate system show chaotic behavior which is the reason for the impossibility of predicting weather for more than 10 days in advance. A scale separation is therefore impossible due to the nonlinearity.
But climate not only depends on internal dynamics but on forcing mechanisms and feedback processes. While external forcing parameters such as the annual difference in incoming solar radiation and the cyclic changes in the solar constant are relatively easy to foresee, internal feedbacks are not. Additionally differences in land use or in the chemical composition of the atmosphere caused by humans, or unforeseeable externally modelled forcings like volcanic eruptions, are intrinsically unpredictable.
From a statistical point of view, Hack (1992) describes weather as the statistical noise on the climate signal, whereas the climate signal is triggered from the above mentioned forcing processes. But in contrast to other statistical processes the weather noise is not to be neglected as a distracting signal. On the contrary, it is of great importance as it is strongly interconnected to the climate signal. The weather noise also includes the internal dynamic of the climate system which often initiates changes in forcing parameters as the dynamic is strongly coupled to the forcing mechanisms via known and unknown feedback mechanisms.
Thus the second great challenge climate modelers face is the fundamental interconnection of short term weather with long-term variabilities in the climate signal. This too is a manifestation of the scale problem but this time not in space, but in time.
Even if in principle the atmospheric or oceanic motions are to be described with the help of fundamental laws of mechanics and thermodynamics, their solution is not analytically possible and therefore poses many numerical problems.

[6]"A mistake with Repercussions" 27. April 2006

3.6. Theory ladenness of climate models

In an inductive interpretation of theory building, generalisations and theories are drawn from observation, from facts. But facts should be independent from theories. This is the desired asymmetry for inductive theory building as described by e.g. Ludwig (1974), but unfortunately it is not true. Without a hypothesis or at least an idea we would not even know what kind of data to observe. Thus it is inevitable to invent hypotheses before collecting data. This so-called theory ladenness of data is therefore a fact of scientific research. This fact shows that pure inductivism seems not to be a grounding for scientific theory building at all, which is in opposition to a branch of philosophy of science and will be discussed in the following section.
Another aspect of theory ladenness is the charge against measuring devices. The accusation is that all of them are constructed along the lines of the theory they are used to confirm. Thus observed data is influenced at least twice by theories it should be the basis of. Dependent on how narrowly the term theory is interpreted this charge holds.
Hence the question is not whether theory ladenness exists or not, but how strong it is, and whether it is so overwhelming that meaningful experiments are impossible to conduct. The fact alone of influencing measurements by structural theoretical ideas or instruments does not mean automatically that it is impossible to make open-minded observations. But it does mean it is necessary to be careful with data. As data and observations are hard to get in climate physics theory ladenness seems to be a challenge climate science has to face.
According to Ludwig (1974), theory ladenness is omnipresent in physics. He freely admits that the accusation of theory ladenness applies entirely regarding physics theories. But in contrast to philosophers like Kuhn he does not see this as a problem. On the contrary, he deplores the independence of some theories. Ludwig identifies the possibility of errors in the perception of the given phenomena as crucial. According to him it seems not satisfactory to build the exact science of physics on the seesawing grounding of the more or less given phenomena. He proposes two ways out of this dilemma. The first one is to ignore it as it does not belong to the realm of physics to question the abilities of our perception. Thus, it is not a physics problem. The second escape could be to establish solid rules for the whole process of theory building.
If not for theory building, common sense[7] rules for climate modelling will be given at the end of this work. Some of the general accusations of theory ladenness can be omitted, if such rules are taken seriously. However, as the following paragraphs will show, there are principal problems in data assimilation and modelling that can be contained within the concept of theory ladenness.

Of course the charge of theory ladenness as described above also applies in climate physics but compared to other areas of physics it is less overwhelming. Climatology as

[7]This term is chosen to underline that no set of methodological rules could be meaningfully given for climate modelling for very similar reasons as discussed for science in general in the introduction (section 1.1). When discussing common sense rules in chapter 7 it becomes apparent that common sense rules apply also and are very central to ethical standards.

a field based on various theories of physics and chemistry uses a wide range of theories and is not said to be emerging straightforwardly from few axioms.
In a different context of theory ladenness, the classical interpretation of theory ladenness, climate physics and especially the modelling part of it, has advantages. The commonly raised charge of theory ladenness, that theories or scientific frameworks are corroborated with data gained with measuring instruments which are constructed on theoretical assumptions from the theory to confirm, is relatively easy to refute. This is due to the fact that the methods of measuring climatological data are far away from modelling approaches concerning their theoretical equipment. Measuring instruments like a thermometer and a measuring pitcher for precipitation use comparatively simple theories in their construction, particularly the latter. But the data compares to data from climate models.
So far theory ladenness seems to be no problem for climate physics. But most of the data used today is not gained in such a way. Satellite data in particular has become increasingly important since the 1970s. Satellites are, unlike a thermometer, dependent on technical data processing. But in this there is no difference to other areas of physics.
The history of the spectrum of climate models should reveal the kinship between different modelling approaches. This is the dependency of the models within the spectrum which includes in particular the historical dependency of climate models. That is explicitly the question of whether there are relationships between the models which are narrower than the theoretical origin alone provides.
The apparent horror scenario arising from this question would be that there was one single historical climate model which serves as a mother of all today's climate models. As there are different approaches in modelling and at least two different lines of climate model development, there will certainly not be one model. But also three or five would be a disturbingly small number. The philosophical disturbance is given only if the kinship is close in theoretical and technical details. Of course theoretical closeness is to be expected to some degree as global models are to represent the climate system in total, which is grasped on an integrative theoretic approach. If this is really a problem is challenged in the next subsection.
But even so, the danger of copying systematic errors is not only a technical matter as the theory of the climate system is nothing fixed but a work in progress. The process of modelling climate scenarios is the most important factor in analysing the climate. Thus all consequences drawn from a climate modelling approach would be biased in the same way.
It is indeed the case that we can find thirty-year-old pieces of code in today's climate models. This is a fact that led science researchers like Gabriele Gramelsberger in Gramelsberger (2006) to consider an archeology of climate modelling, using such pieces of code as the origins of information.
Thus a very close kinship between different GCMs actually exists. But it is not a disturbingly close one. This is due to the characteristics of code. The code is the most exact part of a climate model, only code reveals all parameterisations and thus simplifications and assumptions made within the process of climate modelling. Therefore every piece of code integrated into a new context of modelling obtains a new local value. The

relationship between models is therefore a relationship in ideas and technical tricks as it can be found in every history of experimentation. Their manifestation in code does not add deeper interdependency.
Considering a procedure described for example by McGuffie and Henderson-Sellers (2001), another aspect of the possibly problematic relationship within the spectrum of climate models evolves. They describe an upwelling-diffusion energy budget model to evaluate Kyoto Protocol implications on large scales. "[S]uch a simple climate model relies on climate sensitivity and ice-melt parameters *obtained from a full GCM*" (McGuffie and Henderson-Sellers (2001))[8]. The dependency here is not in the theoretical setup of the model but in the initializing and tuning of data.
In contrast to code based relationships of climate models this is a problematic approach in climate modelling as the model output appears to be the result of an independent approach but is not. With an addition of GCM data to a simple EBM the spectrum of models will be violated as two types of models are merged together without calling it by a name. Such an action would be rational only if the spectrum of models were a true hierarchy with the best models at its top. But this is not at all the case. Models of all levels of the hierarchy are constructed for different purposes and serve these well or badly independently. GCMs do of course include most processes but they likewise include parameterisations in need of heavy parameter tuning, have short simulation runs which cannot guarantee that an equilibrium state will be reached, and are poorly validated. Thus the hierarchy is not actually a hierarchy as such which is the reason for starting to speak of a spectrum of climate models instead (Claussen et al. (2002)).
The same problem occurs not vertically in the model spectrum but horizontally if parameterisations are copied from one modelling approach to another. This is of course not to be dismissed, as the community is lucky if there is at least one meaningful parameterisation. Nevertheless it adds up to the theory ladenness of climate modelling approaches.
Despite these epistemically important reasons for not using GCM output to tune models of lower resolution there are practical reasons to do so. GCM output can normally be generated in exactly or nearly exactly the grid and format needed for the tuning. To get observational data of the same quality is usually more costly and too often not possible at all.

If such a practice is unacceptable or not, strongly depends on the purpose of the modelling approach. If it aims at projecting future climates GCM data is not appropriate to initiate the model. But if the model aims at understanding climate processes GCMs can deliver good initiation data.
For future climate projections alternative sources of initial data must be used.

[8] Italics are from me.

3.6.1. Reanalysis data

Reanalysis data is highly theory laden as the methods which comprise it rely solely on modelling but the intention differs (see the beginning of this part). The modelling approach is thus different which makes the data useful despite the theory ladenness but basic modelling uncertainties are the same as well as parameterisations.

Observed data is available only for very specific observables, which must be at least global and of good quality. Temperature and upper-air mass fields are of such high quality. Others, such as the moisture content of air, are not that good and observables like precipitation or surface fluxes are even worse. But even given that the measured data is of good quality, such as temperature, an energy balance model can be compared to 300 years at the most, on exclusive grid points only: nothing more. Thus data used to develop models, that is, to tune or validate, needs preprocessing before being complemented for comparison. The process of gaining reanalysis data is called data assimilation. Observed data from land surface, ship, radiosonde, balloon, aircraft, satellite and other sources needs to be assimilated to a global data set (Kalnay et al. (1996)). To do this a fully developed GCM is fed with this data and compared to it after every time step in the model simulation. Before this procedure is meaningfully possible the observed data from very different data sets must be put together into a single data format. Errors in dates or longitudes must be eliminated and most importantly, gaps in the data must be filled with other data or via interpolation. Not until then can the actual reanalysis model be executed.

The reanalysis data later used in sensitivity studies with all kinds of climate models is thus heavily loaded with theory. According to Kalnay et al. (1996), only these output fields related to good quality data are reliable and give the state of the atmosphere. Others are only partially defined by observations or completely dependent on the model characteristics as no usable observations are available. To account for that fact reanalysis data is classified according to its reliability. Temperature fields belong to highly reliable class A while precipitation and most fluxes belong to class C, which is as good as the data of a GCM can be, but does not give the true state of the atmosphere or any other resolved sphere.

Researchers using reanalysis data must be aware of the fact that reanalysis data for most system variables is of class C, which is the best data available, but it is GCM data and not observational data. Of course it is not the very same GCM as they want to test or tune but it is a GCM and the kinship between theoretic structures in GCMs is strong enough to be identical in some parts, for example parameterisations. In particular data assimilation models do suffer from the same shortcomings as poor validation and parameter tuning as GCMs for other devices do.

Thus the degree of theory ladenness in reanalysis data depends on the data class the data field in question belongs to. Class A data of reanalysis models, due to the additional statistical interpolation, provide an even better estimate of the state of the climate system as observations alone. But for data from other classes the charge of theory ladenness is well-founded.

The quality of input data strongly determines the model output. Thus data classes should be borne in mind for output data interpretation too.

3.6.2. Paleo data

An alternative to the use of current data sets is to use historic and prehistoric data. The situation is even worse for paleo data compared to modern data. Reanalysis data for the last 30 years has many different radiosonde, aircraft, weather station, and satellite measurements as a grounding, even if these observations are biased accordingly and the data is modeled within the same framework of theory as GCMs used for climate predictions. A paleo historical data set mostly stands alone for its exact period and region and most importantly does not contain temperature records but a certain concentration of water and air ingredients if the proxy is an ice core. Preparation of paleo data thus suffers from the same theory ladenness as reanalysis data, plus the theoretical input used to compute central climate variables from the measured concentrations of air or other available measurands. That is, in reanalysis data of historic and prehistoric periods a good quality data class A is not existent.

Theory ladenness is therefore a central problem in climate modelling if we are concerned with preprocessed observational data and modelling of higher complexity, which is is the case nearly always and everywhere. Thus the climate modeller ought to take great care not to exacerberate this problem through careless handling of data.
This means in particular not to using observed data from one source in more than one step of model development, such as model validation, model tuning, parameterisations, pattern analysis, and perhaps even more. It is a very hard task as there are many steps in model development which depend on data but few data sets, at least few good and complete sets of observed data.

3.7. The impossibility of falsifying climate models

Karl Popper gained fame for his insight that the verification of a theory cannot be achieved. From a logical point of view only the falsification of scientific theories is possible.
In his famous work "The logic of scientific discovery", Popper (1959) developed the term *falsifiable* along the following lines. A scientific theories is falsifiable if a crucial test can be made within the framework of this theory. The classical example for such a test is the prediction of the existence of the planet Neptune and its detection (see Chalmers (1999)). It was detected that the orbit of the planet Uranus was not as expected according to Newton's theory of gravitation. Instead of abandoning Newton it was predicted that another planet, later identified as Neptune, disturbs the orbit of Uranus. With help of the theory of Newton, the mass and position of Neptune could be computed. If no other planet had been found Newton's theory would have been falsified. As Neptune was in fact detected Newton's theory was corroborated.
Theories that lead to tests like that are falsifiable theories and thus scientific. Being falsifiable or not is Popper's demarcation criterion between scientific and not scien-

tific theories. Deliberately Popper neglects the scientificness of huge fields of research, among them most prominently economic sciences and sociology. Part of his motivation to present a strict criterion was the possibility to declare theories of e.g. Marx as not scientific. Popper's intention was to be able to distinguish between science and pseudoscience (Popper (1998)) in his interpretation of the latter.
Given this essence of Popper's work, falsificationism in this logical sense has not lead to a revolution in scientific work. Chalmers (1999) gives a concise definition of what falsifiable means in context of Popperian falsificationism: "A hypothesis is falsifiable if there exists a logically possible observation statement or set of observation statements that are inconsistent with, that is, which, if established as true, would falsify the hypothesis." Being falsifiable is thereby a property of the theory. Thus he not only provides a philosophy of the falsification of theories but he gives a definition of scientific. Only theories containing that property are to be taken as scientific, according to Popper. Only falsifiable hypotheses contain information about the world, if the world represented in observational statements cannot falsify them, they cannot be about them. This latter interpretation of his work is what makes it unfit for everyday scientific work.
Nevertheless, the idea that scientific assumptions should be falsifiable is important. Logical falsifiability cannot not be aimed for but new ideas in science must be testable. Otherwise the question of the scientificness would arise after all. That is, if no logical or technical tests are to be conducted at least the ideas and assumptions resulting in models and theories must be traceable for other scientists. With this consideration we actually come back to Popper. Despite his demarcation criterion Popper denotes in the preface of his "Logic of Scientific Discovery" rational discussion as *the* method of science. This statement puts Popper's whole theory in a completely different light, especially so as he additionally claims the critical eyes of our peers in a scientific community to be the only possibility of identifying errors in theories and gaining objectivity.
Interpreting falsifactionism in this weak way the falsifiability of scientific assumptions should be given for these assumptions to be meaningful. If errors cannot be identified models and theoretic assumptions are useless. With any common sense this claim is obvious.
But there are parts of science where errors are hardly traceable, or not traceable at all.

Climate models belong to the realm of science not open to satisfactory falsification mechanisms. This is not true for very simple, conceptual models consisting of analytical equations only. But the more complex a model the less falsifiable it is.
This is not due to the nonlinearity of the climate system but due to the complexity and the need to parameterise. That is, due to our lack of understanding, errors in modelling are untraceable. Every climate model consists of physically meaningful equations which we are able to test and a huge amount of assumptions are known to be insufficient. The latter is the reason for the impossibility of falsifying them.
Nevertheless it is possible to test the performance of a climate model. Chapter 5.1 is dedicated to the many different methods of climate model testing. There are several approaches to check the reliability of climate models and climate projections but none are able to identify the reason for bad performance. The source of errors may be any

of the equally plausible parameterisations, just as it may be indistinguishable tuning processes, or any other approximation within the modelling process. Due to the very nature of their nonlinear interdependency, their influence is not possible to test.[9]

If we choose not to take Popper's narrow definition of science into account, this fact does not make climate modelling a useless or unscientific undertaking. However, it does mean that modelling results are special scientific findings which cannot be interpreted along the lines on which other physics findings are.
One way to deal with this deficit in the process of climate scientific work is to assign probabilities to parameterisations and model output. How this is done and whether it really solves a problem is discussed in chapter 6.
One conclusion of the impossibility of falsifying climate models that can already be drawn at this step of analysing climate science is the absolute need to put one's cards on the table. Without knowing approximations and assumptions within the modelling process no sensible interpretation of model output is possible. But this implies that a meaningful interpretation *is* indeed possible, and we come back to Popper and the scrutinisation of peers.

3.8. The singularity of anthropogenic climate change

Today climate science is a field of research highly and controversially discussed in politics and society. The reason for this high level of interest is an unprecedented fast and global climate change caused by human fossil fuel burning. As described in section 2.2, this anthropogenic cause is not reasonably subject to doubt and has serious impacts on life on earth.
Climate change in general is nothing new in the lifetime of our planet but at least within the last 800,000 years a comparable high CO_2 concentration has not existed. It is also very much worthy of doubt that the rapidity of today's CO_2 increase has prehistoric predecessors. Thus today's climate science is confronted with a singular climate change. This singularity of current climate development highly influences research in climate science and contributes to the list of difficulties in this field of research. It is not a new fundamental problem of research as it somehow evolves from the fundamental problems of nonlinearity and observational constraints. However, it is to be expected that such a singular event leads to singular effects. That is, current climate development requires research not only to understand equilibrium climates with their forcing and nonlinear effects, but in particular to focus on singular effects and transient climate development.

In a nonlinear system like the climate system small disturbances within the forcing or feedback mechanisms may have extreme impacts. During the Holocene, the warm period we have been living in for 10,000 years, the climate has been extremely stable. But recently a set of effects has been identified by Lenton et al. (2008a) which could potentially

[9] In the philosophy of science these problems are discussed under the topic "unrealistic assumptions", which will be briefly discussed in section 7.2.

shift the climate system into instability. Thus especially the magnitude of this potential is of great interest for politics and society. This is a heavy burden on climate scientific research, as the assignment of meaningful probability distributions on climate change impacts is hardly possible (see section 6). Nevertheless, the identification of possible singular effects was a great success of climate science.

These processes that are particularly sensitive to (abrupt) climate changes are called tipping elements. Climate change may disturb them to such a high degree that they tip into a fundamentally different state. As the potential tipping elements are important processes within the global climate system, their tipping would have serious effects on the overall system. This includes the possibility of shifting the total system into a different state. Besides its impact on the global system the existence of positive feedback mechanisms is a criterion for identifying tipping processes, at least according to the paper of Lenton et al. (2008a). Recently the feedback mechanism was taken from the list, to include all elements of the climate system with the potential of a sudden change which has large impacts (see Lenton et al. (2008b)). Nevertheless, the existence of feedback mechanisms increases the tipping, thus they are of importance in context of tipping elements.

An important example for a positive feedback mechanism is the ice-albedo effect as described above (section 2.1). Therefore the melting of ice sheets of Greenland and western Antarctica belong almost certainly to the set of tipping elements. Sometimes the Arctic sea ice is also taken as a tipping element, in particular one in which tipping has clearly started already. Due to the ice-albedo feedback, among others, global warming is highest at the poles and becomes even higher at the Arctic with proceeded melting of sea ice. The system has tipped when summers are accompanied by an ice-free Arctic Ocean. Among the obvious consequences for the Arctic ecosystem an ice-free ocean would lead to heavy changes in global or at least hemispherical circulation, in particular, the North Atlantic circulation changes that basically determines European and North American weather. The changes are already observed today.

Another example for tipping elements already in the tipping process is the melting of the Greenland ice sheet. Complete melting would cause a sea level rise of seven meters. But this is also an example of the problem of assigning probabilities. Assessments of the timescale of the melting, given a global warming of 2°C, stretch from few decades to a millennium. The probability problem increases with the uncertainty in the identification of tipping processes and serves as an example of imprecise probability assessment (Kriegler et al. (2009)) in section 6.

Even if several identified tipping elements are highly uncertain in terms of their tipping potential and impacts, their existence and the fact that some have begun to tip are indicators of the singularity of anthropogenic climate change. Whether this is truly irreversible cannot determined yet but it is possible. Scientists such as the nobel laureate Paul Crutzen (2002) proclaimed the beginning of a new aeon: the Anthropocene. That we have indeed pushed the climate into an unprecedented state is beyond dispute.

If this change is irreversible it would have the incalculable consequences discussed in section 2.2. A new climate will not necessarily be fit for human life.

The described singularity of current climate development shows that climate modelling is affected by a special aspect of the well discussed problem of induction. A problem of climate modelling is not the confirmation of theoretical assumptions with observed findings; even less inductive theory building. Especially the latter is almost irrelevant for climate modelling in particular as observations are much too fragmentary. Also the problems of confirmation in climate modelling are different from that as discussed in classical inductivism. Chapter 4 is dedicated to the validation of findings from climate models.
The aspect of classical inductivism which is indeed a problem in climate science is underdetermination. The findings of climate science are underdetermined in every sense of the term. The problem is less that inductive inference is by definition underdetermined by evidence and rules of deduction as Lipton (1998) for example describes, but there is almost an absence of inductive inference. Paleo data is rare and laden with theory as explained previously and even nowadays observations are sketchy, but for almost every question of research concerned with anthropogenic climate change answers must be given without observed data. That is where climate models become of importance.
In section 2.2 the facts of global warming are presented. We do know with very high certainty that our fossil fuel consumption and CO_2 emissions lead to global temperature increase, but the impacts of these facts are uncertain with few exceptions. That is, climate science is forced to simulate a climate with an extremely high and extremely fast green house gas forcing in order to make realistic climate projections. Paleo data is only available for climates that existed in the past. Thus it is not possible to validate climate modelling results with observed data. There is paleo data for example of a melting of the western Antarctic ice sheet which caused a sea level rise of 3.5 meters but it took 1000 to 7000 years. However, past temperature increases of 5°C took around 5000 years, triggered by changes in the earth's orbit for example. If society does not act in emission reduction we may witness such an increase in 100 years. In a nonlinear system cannot be expected that impacts of changes in forcing are independent of the timescale of the forcing. Validation of future climate projections must therefore be supported without the help of comparable past climate changes, as there are none.

Even if anthropogenic climate change contributes to the problems of climate research it is also the reason for society's urgent need to get climate predictions. Politics and society demand concrete predictions of climate change impacts. On the other hand climate research would need a good idea of the future development of global emissions, at least to be able to make realistic projections.
Society demands predictions but climate science delivers uncertain projections. To grasp this dilemma the term 'prediction' needs some refinement.

3.8.1. Prediction in climate science

A prediction in every day language is nothing more then a statement that a certain event will occur in the future. In the etymology of the word it is nothing other than a

forecast. In the context of mathematical probability theory a prediction is a very well defined forecast of a future event insofar as a probability distribution of the event in question can explicitly be given. In a sense this fact ends the prediction discussion for climate modelling as true probability distributions are impossible to reach for future climate states.
Eagle (2005) gives a definition for predicting processes:
"A prediction function $\Psi_{P,T}(M,t)$ takes as input the current state M of a system described by a theory T as discerned by a predictor P, and an elapsed time parameter, and yields a temporally indexed probability distribution Pr_t over the space of possible states of the system. A prediction is a specific use of some prediction function by some predictor on some initial state and elapsed time, who then adopts Pr_t as his posterior credence function."
This definition represents in principle the prediction account of statistical mechanics. It underlines that a prediction is, in contrast to common sense approaches, a clearly defined forecast. Thus predictions by oracles are not predictions at all in this natural scientific sense. Furthermore it shows that a prediction depends on several aspects, which are the current state of the system M, knowledge about the system under consideration, here denoted as theory T, and the predictor P, its relationship to the system and its knowledge about it. Implicitly this definition also shows that the input state must be so widely known that fixing of a posterior probability function is possible.
In other words the quality of the prediction depends crucially on the capacities of the predictor as the required theory is derived by the predictor and initial states are measured by it. Therefore the prediction cannot be better than the observational capabilities of the predictor and the theory it builds. The constraints of the prediction are thus those of the predictor. Therefore predictions made in climatology suffer from the same limitations as our understanding of the climate system. In addition to epistemic and computational constraints our predictions are pragmatically constrained, which basically refers to the insufficient theoretical assumptions we have about the earth system and technical limitations of computing.
Within this context an important difference should be pointed out, the difference between qualitative and quantitative predictions. The latter is defined as depicted above while the former is not that formally fixed. Quantitative predictions are what we are longing for in climate physics. Even if it is at least a start to know that the climate will change considerably in the near future, such a qualitative prediction is not fit to base detailed decisions upon.
Quantitative predictions of density functions for future climate variables are impossible to give with current scientific knowledge. However, fortunately, there are at least two sorts of forecast that lie in between qualitative and quantitative predictions. That is a trend or directional forecast which does not predict a density function for the future evolution of a variable but forecasts if its magnitude will increase or decrease. We are talking for very good reasons about global warming thus at least a trend forecast for global mean temperature is possible.
The second kind of prediction in this context is not a mixture of qualitative and quantitative predicting but can be either the one or the other. This is a conditional forecast. A

conditional forecast is a prediction that something in the future will happen if a certain condition applies.
Every statement of the future made with the help of climate models is if anything a conditional prediction, where the conditions are that the assumptions made to initialise the model apply. That is, according to Eagle (2005), a model prediction always entails the presupposition of conditions, thus the assumption of a theory. The prediction includes the second prediction that this theory which also includes background knowledge holds in the future and leads to the conditions constraining the variable's forecast.
Whether conditional forecasts reveal something about reality depends heavily on the accompanying condition; they need not tell anything of the world at all.
In contrast to climatology, predictions considered by philosophers are not only quantitative ones. In the context of theory testing, time independent qualitative predictions, which are called ontological predictions (Betz (2006)), play an important role. An example is the theoretically predicted existence of the planet Neptune as discussed above. Such predictions are what Popper (1959) had in mind when highlighting their importance for theory falsification. In Popper's opinion theory testing will be most successful if the prediction is bold because a false prediction also allows us to learn a lot about the theory under consideration, whereas bold in this context is a prediction that would put the theory as a whole into question if it should proof false.
The fact that climate models, especially GCMs, are basically constructed to make rational predictions, predictions to base decisions on, forbids such a theory testing practice. It is not impossible to use models to forecast events hitherto unexplored. On the contrary, many models, in particular less complex ones, predict abrupt changes in climate processes. For example, a transition of the the Indian monsoon cycle as found in a minimal conceptual model by Levermann et al. (2009) in which they showed that two stable climate states in India exist, with one being unable to provide the monsoon rainfall necessary for Indian agriculture and the threshold between both not far from the actual climate. The disruption of the thermohaline circulation (Rahmstorf (1994)) with its enormous consequences for life in Europe is another example of sudden climate change and is thus bold in a Popperian sense. But such predictions of abrupt climatological changes differ in an important way from the ontological predictions Popper referred to since they are not time independent and depend on the very strong assumptions and simplifications such models necessarily make. Thus they are useless for testing the climate models and when the model is proved correct concerning the predicted event it is too late, in an extensive sense.
A more basic argument against the possibility of ontological prediction based theory testing is the absence of theory in climate modelling science. The inconsistency and incompleteness of the theoretical assumptions the models are based upon make it impossible to infer time independent ontological predictions. Thus for testing climate models their predictions are of no significant use.

The success of a prediction depends very much on the circumstances in which the prediction was made and on the type of prediction. Therefore the process of predicting depends also on these circumstances and in particular on the reason for prediction making. Our

reasons for making predictions should determine their type and outcome decisively.
Salmon (1998) shows three different reasons for making predictions. The first one is just curiosity about the future without wanting to wait for it. The second one is the testing of theories. The third one is the necessity of acting in a situation depending on the near future: an example of this type of decision making is wagering. In classical physics the reason for predicting is mostly a mixture of the first and the second, scientific curiosity and testing of theories. In today's climate physics we are also confronted with the necessity of decision making which takes into consideration prospective global warming. Thus the necessity to act is clearly given, the question of how to act depends on the outcomes of predictions. As the ability of making quantitative predictions is very low in climate physics the need to make decisions which nevertheless exists, governs important parts of climatological research. In particular the research field of integrated assessment models the climate according to different natural scientific and economic developments to provide grounding for political actions (see below). The task is to combine the quantitatively unpredictable future developments of the climate and the world wide economy to give instructions for how to act in the near future to secure a future environment worth living in for humans and as many other species as possible. Methods to cope with this intractable task are manifold. Even wagering becomes important in that context.

Considering these different reasons it becomes apparent that not only the outcomes of predictions are dependent on the reasons of their genesis but the whole process of predicting differs according to our reasons for predicting. Karl Popper suggested, e.g. in Popper (1959) as mentioned above, that it is best to falsify and thus test theories by making bold predictions. Therefore predictions to test theories will probably be such that they concern fields other theories cannot talk about or predict behaviour contrary to observation. If predictions are made instead to base decisions upon this method becomes difficult as the future development depends on human actions and decisions, which are in principle unpredictable.
Chalmers (1973) also identified a principal problem of Popper's method of making bold predictions in the context of theory testing. If we are busy making bold predictions and disregard the testing of cautious predictions we are running the risk of taking too much for granted. But if we test a very cautious statement evolving from a theory which was regarded as unproblematically true and the test falsified it, we really learned much more from the falsification as some wild guess. Chalmers also realised that the confirmation of a bold prediction is always accompanied by the falsification of cautious statements. This is due to the definition of the term bold. A bold prediction is one that is contradictory to accepted knowledge, the background knowledge, which naturally collapses if a bold prediction is confirmed by observation.
Interpreting bold predictions in such a way it becomes apparent that such a method is of no significant use for climate physics. Besides, the theoretical framework of climate physics is not sufficient to lead to meaningful bold predictions even if they were desired. There is not that much background knowledge to contradict. If there was the theoretical possibility and the will to do so the modelling know-how would be the limiting factor in such a task.

Bold predictions are of course not the only means of testing theories. Salmon's reasons second and third reasons for making predictions are therefore not as contradictory as they may seem given Popper's testing method.

If predicting the climate in order to make decisions leading to actions, the prediction and thus the modelling setup is arranged to represent a climate leading to possible interventions. That is, scenarios are modelled trying to represent possible worlds in a very narrow sense. The climate models used to produce grounding for economic and political decision making are integrated assessment models, which consist of climate as well as economic components. In particular they represent CO_2 emission forcing using different mixtures of energy sources. The technologies included in these scenarios are those we can now think of as available immediately or very soon (e.g. Meinshausen et al. (2009)). This normally includes regenerative energies, fossil energies and nuclear energies but excludes technological hopes such as nuclear fusion. Most scenarios of this type available today include carbon capturing and storage (CCS) as well, a technology not yet available. If it turns out that CCS is not ready for commercial use within the next decades the very basis of many recent political decisions will prove much too optimistic if no replacement of CCS shapes up alternatively.
Emission forcing of climate scenarios depending on energy use are extrapolated from current usage for worst case scenarios and from ambitious political proposals for best cases. In short, the scenarios are based on the current state of climate, economy, and politics as they should be the basis for political decisions today.
Predictions, to have the possibility of being successful, are made not to base an urgent far-ranging decision upon but are grounded on current knowledge, although not on today's possibilities of acting. Thus the range of outcomes of prediction scenarios is wider as the input frame is not that limited.

Predictions of the latter kind could be those that test theories. That is, to be able to speak of climatology, to test slightly different theoretical assumptions or to model processes not yet modelled at all.
Generally predictions made in order to corroborate or falsify certain theoretical aspects are set up with special respect to the process in question. A biasing towards that part of the system will be hard to omit.
Nevertheless, that kind of prediction is important for scientific progress, at least if many different theories are tested in a branch of science. Thus there is a variety of possible worlds generated that may answer pressing 'what if?' questions in that field.

In climate modelling sciences the making of predictions in order to test certain technological assumptions generates very different climate scenarios, even if the term 'prediction' in such cases is even less applicable than in climatology in general. To model a climate, including a new theoretical conception of how a certain process' physics could look, the model at hand is tuned to show the process while the new physics is included in the mathematical part of the model. This is a testing of this physics insofar as the tuning process can appear impossible or the model output reveal a world so very different to

the real world that the modelling setup is abandoned. But the tendency to show what was hoped for is clearly given due to the tuning process. Other physical mechanisms playing some role in context of the newly designed process in particular are very likely to be underestimated in such a testing approach.
Such predictions are bold in the sense of being partly in contrast to background knowledge but as their failure is relatively unlikely they do not serve as a crucial test for theoretical input or even the model itself.

The first reason for making predictions identified by Salmon (1998) is curiosity, which is very likely the motivation for predicting that least biases the outcome of the prediction in some way or the other.
Not being willing to corroborate a certain theoretical assumption or needing to give advice for a realistic action on the basis of the prediction allows for free choice of input. The guide for setting the prediction up could then be trying to answer the questions of what is possible and realistic as input knowledge. That is, the current state of the system is assumed as realistically as possible and the describing theory is chosen to account for all available theoretical knowledge.
Contrary perhaps to the term 'curiosity', chosen by Salmon, this motivation for prediction does not arise from science at all, rather all analysis conducted to understand scientific correlations whether involving predictions or not, are based on scientific curiosity. Curiosity and the will to understand is what drives science. And, if exploring a new branch of science curiosity is the first reason leading to predictions. Only if there is already some understanding do other reasons become of importance and actually become possible. But not so in climate science, as anthropogenic climate change forces us to act now and base our action on predictions.

To summarise this section the following can be said.
It is not only inherent to climate predictions that they exist on very small scientific or inductive grounding. However, the whole problem of induction becomes pressing in the context of climate modelling given the fact that we are confronted with a singular global event, climate change. That is, even ignoring the fact that inductive inference is philosophically questionable there is hardly any data for inductive inference in the context of climate change. Ice core data reveals (Lüthi et al. (2008)) that at least during the last 800,000 years there has never been a CO_2 concentration in the atmosphere as high as what we have reached today. Thus predictions, or better projections, made with climate models cannot be grounded on some prehistorically comparable event.
To make an assessment of future climate development it is necessary to have future climate forcing as model input. Thus model simulations of the future are based on scenarios of this forcing, whereas the most important forcing in times of climate change are CO_2 emissions. It is thus only possible to make projections of future climate development dependent of the scenarios. The projections strongly depend on the emissions suggested for the future, which in turn depend on political decisions and corporate actions. Therefore the uncertainties of the different scenarios change with every global conference, national election, or economic development. The uncertainty of future emissions enor-

mously increases the underdetermination of climate change research in comparison to climate science considered with understanding of the climate.
The underdetermination of our knowledge of possible climate development, given the fact of 380 ppmv CO_2 and more is singular, since there has never been an identical situation in the known history of the climate. This singularity is thus a singularly pressing problem, for our models and human race.

Part III.

Handling of Problems

4. Validation in natural sciences

The previous part of this work was dedicated to the problems the external and internal features of climate science research is confronted with. Not all of these problems are straight forward to identify. It is especially difficult to find out how they influence the results within a modelling process, thus a comprehensive evaluation of climate models is also a rather challenging task. Additionally, not only the model itself as represented by the model output but also the input assumptions must be tested, which presents us with a dilemma as the only method to test the theoretical assumptions of climate science is to model them. It is therefore impossible to truly trace errors back to their source in either the theoretical assumptions or model implementation.

Climate models, as defined in section 2.3, are not fixed entities but distinct and complex processes captured within the term of a climate modelling approach. Testing of such an approach is therefore testing predictions, numerics, theoretical assumptions, technical approaches, programming skills, parameterizations, emission scenarios, tuning abilities, and much more. While public attention is almost exclusively focused on forecasting, scrutinizing is similarly important for all other aspects of climate modelling. Moreover, only carefully evaluated climate models could provide successful predictions as described below (section 4.1.2). The core of the problem concerning predictions in climate physics or climate modelling is that climate models are not possible to falsify. This is due to the fact that the hard criteria which a climate model must fulfil in order to either stand or fail a test is lacking, as is discussed in detail in chapter 5. Every prediction is furthermore strongly dependent on highly speculative scenarios for the future development of the world's society, including CO_2 emission paths. In this sense Thomas Kuhn's analysis (see below) of scientific practice also holds true for climate science.

Thomas Kuhn is different to his predecessors. In contrast to them he realized that what he called "normal science"[1] is considered with puzzle solving on varying scales, and to cite Hacking (1983), is not in the "confirmation, verification, falsification or conjecture-and-refutation business at all". For Kuhn (1996) all theories are imperfect and thus show anomalies and phenomena that are not covered by the theory. Of this category are the puzzles scientists are normally occupied with solving. Only in those cases where the puzzles remain unsolvable within the active theory frame and anomalies pile up does the discipline get into a crisis which will likely result in a scientific revolution. Instead of the imprecise notion of theory frame, Kuhn uses the term paradigm which is open to even more varying interpretations. Therefore a scientific revolution is identified with a

[1]For Kuhn Popper's description of science is merely a description of scientific revolutions but not of scientific every day work. He uses the term "normal science" to describe the latter in contrast to scientific revolutions when one paradigm is replaced by an other. It is used here accordingly to depict what scientists, especially natural scientists, do in their every day work.

paradigm shift which can be interpreted as the rejection of theory. But this rejection is not the result of a severe theory testing process, it occurs as a necessity when too many pieces are missing for the puzzle to be solved. In Kuhn's opinion every theory contains integral parts that will lead to its own destruction. Thus a scientific revolution is nothing special, rather it is something that will happen sooner or later in every theory frame. Following this train of thought, the testing of theories becomes not only superfluous but also senseless. Every theory will be obsolete sooner or later. Thus a theory is good if it serves one's current needs consistently. This can be interpreted for climate modelling: the core for their own destruction lies in the models as it does in Kuhnian theories. What is different is that we put in place these cores deliberately without knowing their effect. Nevertheless, perhaps even especially so, they do need validation, evaluation, corroboration, or whatever term proves to be adequate in terms of a thorough judgment of climate models. However, as even theories are not confirmed through strict tests, a binary type rejection or even the term false in the context of climate modelling is inappropriate. Instead a confirmation on partial agreement between observation and model output or probabilistic methods are the aim of actual climate model testing.
The following chapter shows that the different model testing methods used nowadays are so comprehensive that they show many features modern climate models do have that cope quite well with some of the epistemic problems. But before this is possible the next paragraphs show that actual testing in physics is relatively far from philosophical interpretation. It is thus a refinement of to what is captured by the term testing.
In the context of testing, predictions play an important role. At the same time predictions are fundamental in climate modelling but for a different purpose. To distinguish these prediction approaches from each other and discuss their relevance in climate model evaluations section 4.1.2 discusses criteria of prediction making.

4.1. Testing in natural science

This section will not consider philosophical ideas of how to test scientific findings. First of all as this is *the* subject of epistemology during the 20th century it would be far too much to discuss. But second, and most important, it would not shed any light on what scientists do, as the philosophical research on theory testing is very much independent of actual climate modelling practice.

Nevertheless, the two already mentioned influential philosophers Karl Popper and Thomas Kuhn identified crucial aspects of everyday research, although Popper in a footnote rather than in his actual work.
In reply to Popper and in opposition to him, Kuhn worked out that testing of theories or even just parts of them is impossible. It is merely the scientist who is tested instead of the theory. A theory cannot be tested independently of its whole scientific context.
Another very important problem Kuhn (1998) refers to in examination of Popper is the fact that testing in order to falsify is something that normally does not happen regularly

and purposefully in normal science. This of course need not imply that it would not be better if it happened but at least reveals that scientific practice can obviously do rather well without. Therefore Kuhn moves Popper's demarcation line so far as to lie between natural and other scientific undertakings. The latter is open to testing while the former is a puzzle solving business. But in the context of Kuhn, testing is interpreted very differently than in Popper's definition of scientific practice.
Looking at laboratory practice, theoretical and model based every day work, Kuhn's analysis seems correct. It is not the case that scientists are testing broad hypotheses or even theories. Every day work is concerned with very small problems in neat contexts, rather than with theories. Scientists try to implement new ideas in some context or the other. These ideas if thoroughly analysed can be regarded as part of a hypothesis, but normally they are not analysed at all but are attempted to be realised. This idea focuses on a small aspect of a theoretical or technical problem. The solution of this problem is guided almost completely by practical questions and possibilities. According to this practice the idea might be refuted as impossible to implement or even false within the current context. In a manner of speaking the impossibility to implement small ideas in the wider context may lead to a refutation of this context and thus can be called testing of hypotheses. But these kinds of hypotheses are not strictly theoretically derived assumptions that e.g. Ludwig (1974) termed hypotheses. The every day work is indeed concerned with refuting and corroborating ideas but they normally touch only very small bits and pieces of the context, as is exemplarily shown in section 2.3.2. Thus in every day work theories as whole coherent stories are not tested at all. They are not even considered as framing the daily puzzling.
This fact shows that a purposefully undefined term like paradigm, which Kuhn uses in 22 different interpretations according to Hacking (1983), is meaningful to display reality. If asked a scientist can hardly name one theory frame of his work but several physical theories and assumptions effective for his specific questions.

Very similar to Popper's note in the preface of his "Logic of scientific Discovery" that rational discussion is *the* method of science, Kuhn identifies this discussion as testing but only in non-natural sciences. Similar to Popper's claim, he claims that the critical eyes of our peers is the only possibility to identify errors in theories and gain objectivity. For Kuhn such a discussion exists only within philosophical and sociological contexts and within the humanities but works in a comparable way. In the natural sciences he does not see that inter subjective control but collective and competitive puzzle solving. This is how Kuhn comes to claim that Popper's demarcation criterion is a line between natural and other sciences.
Both philosophers made an important point here: Popper with the identification of rational discussion as *the* method of science and Kuhn in claiming puzzle solving as scientific practice. However, Kuhn is incorrect in his analysis that natural science is free of intersubjective control, because not only does critical discussion exist within natural science, but it is institutionalized in the peer reviewing practice of publishing scientific papers.
In interpreting puzzle solving as the scientific method, Kuhn (1998) says that rules are

needed according to which puzzles can be solved or regarded as solved. Only if such rules are available within a field of research is scientific practice possible and is the very field considered as science, although the rules are individual for each branch of science.

Looking at the interpretation of science of these two famous and influential philosophers, more scrutinizing provides more ideas but no more consensus. Yet there are certain crucial criteria marking scientific work, which are the critical discussion of scientific results, a community of peers able to judge the results, the publication of methods and a precise handling of data.

All of this exists in climate science. And given the fact of how important this branch of science is for policy and society, most scientists are extremely careful in the interpretation and publication of their findings. Additionally, the climate science community is very heterogeneous as physicists, economists, mathematicians, sociologists and more work together on one goal[2]. Thus the critical discussion takes place from different points of view.

Nevertheless, Kuhn's idea of individual rules for branches of science would be helpful if given explicitly for climate modelling. The rules Kuhn that claims exist are hardly explicit rules but there are specific implicit rules for several branches.

Whether such rules do explicitly or implicitly exist and what role they do play, could play, and should play in climate science is the subject of the closing chapter of this work.

4.1.1. Climate models as experiments

In chapter 2.3 a climate model is defined in terms of a modelling approach. But it is also explained that climate models belong to the category of experiment. They especially belong there as they are similarly used to test assumptions. Their epistemic status includes the fact that climate models are not a one-to-one replacement of experiments in classical physics. Crucial experiments are impossible to perform, because a consistent chain of causes and effects is lacking. That is especially important on the input side of the model which is not governed by a consistent theory or hypothesis, but depends crucially on parameterizations and simplifications. A model experiment as a computer experiment is often called an in silico experiment (Gramelsberger (2010)) which implies a close relationship to an in vivo experiment which also lacks some criteria of scientific experimenting. In computer experiments the modelling replaces the mathematical modelling of an experimental setup, whereas the simulation is the experimental phase. The implied analogy to in vivo experiments is misleading. For three main reasons model experiments are not experiments in a traditional sense and also not analogous to them. Firstly, the models contain parameters whose values are highly uncertain or not constrained by real world phenomena at all. Secondly, models are imperfect representations of the underlying system and thirdly observational data measured in the real system used

[2]One goal is certainly not entirely true as will be discussed in part IV, but understanding the system is the central aim to achieve others.

to initialize and tune the models are imperfectly measured. Furthermore, the nonlinearity of the model needs to be underlined as it prohibits in most models the identification of a closed chain of cause and effect. But it is this cause and effect structure we capitalize on in traditional experimenting. Accordingly models cannot be used in analogy to physics experiments which test theoretical assumptions. If anything a model experiment can be compared to a statistical test which is discussed in section 6.2.

Besides the discussed shortcomings in climate modelling which prevent it being compared to an experiment, the spectrum of climate models presents another dilemma. Simple models show approximate causality but no resemblance to the world. Thus they can be used to test the consistency of theoretical ideas but not whether such hypotheses describe any aspect of reality. Complex models instead are inconsistent. Highly complex climate models like GCMs are used in a more classical method of theory testing via implementation of new 'spheres', like stratospheric chemistry. Such a new sphere comes as a new component of the model and is a model in itself. Coupling it to a GCM is a kind of experimental theory testing insofar as the new part is added to an already evaluated climate model. But using complex models as experimental setups the charge of underdetermination holds even more than in traditional natural science. Thus, if a newly developed coupled GCM fails to represent expected behaviour this failure could be for a million reasons, and due to nonlinearity and complexity the reason cannot be identified. Furthermore, such new components are developed with the aim of better representing natural processes than older representations of the system component in question. If and when such a component is implemented it is expected to be better than its predecessor. Thus, also typical of underdetermination argumentation, other explanations of unphysical model behaviour are exploited instead of questioning the newly assumed theory.

To merge the findings from above models are essential for theory development and the development of understanding theory, but not for theory testing. This underlines the special status of climate science as there are several widely accepted theoretical assumptions but there is very poor knowledge of what this implies for the system they describe. Climate models are crucial in enlarging this knowledge but not particularly in testing it, which is also not their aim. Larry Laudans famously claimed that "accepting and rejecting theories is a rather minor part of science" (in Hacking (1983)), which is all the more true for climate physics.

4.1.2. Successful predictions

Whether a forecast is true or not will be seen when the predicted time is the present. This is the ultimate reality check but if it were the only one, predictions could not be meaningfully included in scientific work. However, the are. In the history of science many predictions were made that were credible when they came into existence and became true within the given error ranges. Thus there is a history of successful predictions and most of them did not become true by mere chance.

The identification of a successful prediction depends on the modality of the prediction in question. Quantitative predictions bear other attributes than qualitative ones, although all predictions have in common a fundamental dependency on the theoretical surroundings of their emergence. If the theory is weak the prediction can be everything. This is comparable to the syllogism ex falso quodlibet. In contrast to this, predictions made in the context of and that are consistent with a strong theory bear a high probability of being successful. Most easily this is to be seen concerning ontological predictions. The French mathematician Le Verrier calculated the position and magnitude of an eighth planet from a disruption of Uranus' orbit. Only taking the existence of Neptune into account could the orbit be calculated exactly on the basis of Newton's established theory of mechanics (example from Hacking (1983)). The theory used to calculate Uranus' orbit and predict Neptune's existence was highly elaborated which made it possible to calculate the orbits with high precision. Hence the prediction of the planet was a very precise prediction. The success of a prediction depends on such precision since only precise predictions contain information. This last point is an ultimate criterion for all predictions. A prediction must entail new information to be successful. However, this is of course not a sufficient condition as false information is new as well.

The demand for precision applies in particular to quantitative predictions whereas qualitative predictions contain information by definition. Neptune's existence was, for example, predicted with it's precise orbit and mass, but qualitatively also the prediction of the existence of a thus far unknown planet is informative.

A prediction is to be accepted as credible if it is consistent with its theoretic background and with the data it was derived with. This implies that the data used for the prediction is also reliable. Whether we trust the theoretical background or not will then finally determines our judgment of the prediction at hand. This is especially true in the case of qualitative predictions. If all this is given, we strongly believe in the theoretical background, we have good data to base the prediction upon, and it fits consistently into this background, it is possible that a qualitative prediction is so reliable that we can call it successful even before being able to make the reality check. Such a belief is applied to consistent theories with closed and solvable mathematical theories that have stood the test of different testing methods. The scientific community believes in quantum mechanics but also in Newton's mechanics even if this is definitely not true, because it describes our perception of reality.

The best example of a successful qualitative prediction in the context of climate science is global warming. As described in detail in chapter 2.2 the predicted increase of global mean temperature rests upon the very basic principles of physics which are not reasonably questionable given the available knowledge. Therefore global warming is to be taken as a successful prediction even if up until now empirical evidence for its truth is small.

The case is completely different if it comes to quantifying global warming. To make meaningful quantitative predictions a certain level of precision is needed as not to state hardly more than that the variable will have any value, which is not a prediction but a trivial statement and therefore not successful either.

This implies that the success of a prediction depends not only on the quality of underlying theories and its precision but on the predicted subject as well as on the reason for making a prediction. If our knowledge of a certain topic is already high, only narrow ranges of values can be taken as new information and thus successful predictions. On the other hand, there are branches of science and especially fields and subsystems of climate science where the credible prediction of the order of magnitude of the value of a certain variable will have is a great success. To apply an informative degree of precision to climate predictions is thus a great goal of today's climate modelling approaches, yet a difficult balancing act. To accommodate the demand of new information predictions of climate variables with confidence intervals or even probability density functions are attempted, but at the same time the level of precision must be low to allow for the fundamental and model uncertainties of climate modelling. In chapter 6 the assignment of probabilities in climate modelling is discussed in detail.

Given the last paragraphs successful prediction making seems to be a very hard task for climate system modelling as they can neither emerge from a strong theoretical background nor be given with precision. To nevertheless make meaningful predictions of the future the prediction of past events has become established as a very fruitful method of validating climate models. The hope of this so-called hindcasting is that climate models able to represent past events contain correct theoretical assumptions which are therefore applicable to predictions of the future. This concept of hindcasting is the most successful tool of climate model testing and is further discussed in section 5.1.2, but an important problem using successful hindcasting to forecast is mentioned here: there are only three major climatic periods that we do have data for. There is considerably good data for the last glacial maximum and for the post-industrial period, i.e. the present, furthermore there is data for the Eemian, which was a comparable stable interglacial period starting 130,000 years ago. To truly test the climate model using data from these periods the modeller must "forget" that they have this data and the information it contains in order to avoid building the model in such a way that it will show the events portrayed by the data. Otherwise the model needs tuning, which in turn needs data and as complex models are unable to simulate several thousands of years tuning data and validation data necessarily comes from the same period, which increases the difficulty of actively forgetting about the data.[3]

Predictions of future climate development are thus biased towards a climate and show perhaps more stable future climate as would be correct since the available data is of stable climate periods.

To summarise, it can be said that predictions of climate change impacts are normally extremely uncertain and hence imprecise and are built on uncertain data thus the criteria for successful prediction cannot be met, whereas the forecasted global warming fulfils every criterion for a successful prediction. But the fact that most impacts of this warming are not exactly predictable does not mean that they are not true. It implies that we have to live with high levels of uncertainty which makes climate projections

[3]This problem strongly relates to the philosophical discussion of new data as depicted in the introduction. In this context it will be discussed in part IV

even more disturbing as we will never know in advance how bad it might get. To bet on the low probability of harmless impacts would be an act of high levity which greatly violates the precautionary principle.
Furthermore, the successful prediction of global warming can be used to corroborate climate modelling results insofar as climate model outcomes showing a global temperature increase represent a future reality and are thus preferable to those which lack this ability. It is even a better criterion than hindcasting ability since a model correctly representing past climates need not be comparably good for the future. The ability to simulate global warming is only a negative criterion to falsify climate models, but the next section shows that it is one of very few, or maybe the only, hard criterion in climate model testing.

5. Model Testing

The first vindication of climate modelling practice from the community itself is mostly the argument that models are 'the best we have' to answer scientific and societal questions; especially if confronted with deficiencies of modelling approaches. From an epistemic point of view this is no vindication at all, which is true for a natural scientific view as well. But even if it is the case that the actual spectrum of climate models and their handling is the best practice available there is no need to be so defensive, despite the manifold problems discussed in part II.
Within scientific practice a lot of work and thought is put into validation and vindication of the models in use. There are several different methods to evaluate modelling approaches and their performance. This chapter will show the range of actually applied testing methods whereas the next part is dedicated to hint at methods of modelling and model testing to improve modelling practice not comprehensively applied yet.

5.1. Model testing

The forth IPCC assessment report starts its eighth chapter on model evaluation (Randall et al. (2007)) with the "confession" that only certain predictions can be demonstrated to be right or wrong but the models themselves must always be regarded critically. That means that we can gain at the most some confidence in a model but never any certainty in advance whether the results from model simulations reflect reality. In short, model testing seems to be impossible. Whether this interpretation proves to be the only possible one will be evaluated in the following paragraphs.
What is certainly true is the impossibility of verification or validation of a model. This is due to the fact that model results are never unique and natural systems never closed (Oreskes et al. (1994)). The former is a result of the underdetermination of models by available data while the latter is quite obvious as only pure logical systems can be verified. This and the facts that models are normally too complex to grasp in every detail and contain nonlinearities not open to human analysis make only partial confirmation possible. But even such a corroboration is not to be gained easily.

5.1.1. Testing in general

The straightforward way to get confidence in a climate model is to successfully simulate present and past climate states using the model in question, which is introduced above

as hindcasting. The most common way of doing this is through modelling time series comparable to observed data, reanalysis data, and paleo data.
A widespread method of hindcasting is to perform the same preprocessing with reanalysis data as well as model output. An example is the calculation of climatologically important fluxes out of state variables. Even more basic is statistical analysis of the state variables directly. Time series of model output are compared to time series of reanalysis data using statistical tools. Statistical methods are for example the principle component analysis and computation of empirical orthogonal functions. This is just one of several statistical methods but as it is very commonly used the method is described here to illustrate the general principle.
The principal component analysis (PCA) allows us to structure, simplify, and visualize huge data sets. A number of statistical variables, e.g. state variables of climate model output, are approximated by a much smaller number of hopefully significant linear combinations (the principal components). Possibly correlated variables are thus transformed into uncorrelated ones. The calculation of empirical orthogonal functions (EOF) is mathematically comparable but instead of PC analysis the computation of EOFs allows not only time series in the data to be found but also spatial patterns. Therefore performing of PCA and EOF is not only of statistical value but allows for the detection of typical climatological patterns in model data. Such patterns are for example large atmospheric circulation regimes such as the north Atlantic oscillation (NAO), which is basically the interaction of Azorean high, Iceland, and Aleuten low, or the El Niño/ Southern Oscillation (ENSO) phenomenon.
Statistics like this may thus also help to judge the performance of the model without direct comparison to reanalysis data. It is equally useful in inter and intra model comparison. Tests of this sort are to be made on the component level as well as on the whole coupled model.
The comparison of time series and EOFs is possible even by looking at them. This seems to be an easy method but there is no common quality criterion governing which degree of agreement is good, in which variables and on which time scales. Therefore it is always open to interpretation whether the inability to represent some pattern or the other counts as a model failure. There will always be scales and variables that disagree and there will also be agreement for the wrong reasons, namely the appearance of a certain pattern due to parameterisation only. Such a simulated phenomenon is independent of the model's physics and thus carries no information concerning the ability of the model to well represent such patterns in climate projections. This problem is quite common in comparison with 20th century observations as especially GCMs are constantly fitted to represent this very data and are too complex to trace the source of patterns.

In sophisticated evaluation processes the dynamics of the model data comparison is used in addition to component analysis to analyse statistical errors. These errors are calculated from different model variables and the matching observed variables and are compared afterwards. This is basically done by calculation and comparison of absolute error measures. The thus computed mean absolute error (MAE) is the error of the model and not of a single simulation by summing up a sequence of errors e_i from the comparison of single hindcast simulations for the predicted variable and measured data for that variable.

$$MAE := \frac{1}{n} \sum |e_i|$$

But this error takes only the bias into account not the variance, thus every error counts the same independently of its magnitude. To measure errors according to their weight the mean square error (MSE) can be calculated.

$$MSE := \frac{1}{n} \sum (e_i)^2$$

The MSE can be seen as the variance plus the squared bias. As this is not of the same magnitude as the underlying variable the generally computed statistical error measure is the root of MSE the RMSE:

$$RMSE := \sqrt{\frac{\sum (e_i)^2}{n}}.$$

Computing the RMSE larger errors have stronger weight than small ones thus the RMSE seems to be an appropriate measure in the context of model data comparison. Mathematically the MSE and RMSE are the first two moments of the distribution of the variable if the distribution can be approximated as Gaussian, which is often done, it is sufficient to compute only the mean and its variance.

Besides these very simple and common but nevertheless most important statistical methods there are several approaches of model evaluation using standardized tests of numerical methods within the model. Finding such testing methods and making them comparable is undertaken for example in regular workshops. The huge amount of data necessary to develop and conduct such tests is based on an enormous number of case studies concentrating on parts of the system (e.g. Randall (2000)).

An example of an alternative measure that does not compare two states as the common error measures is the metric relative entropy proposed by Shukla et al. (2006). This entropy is a distance function between two distributions and is found to be smaller if the model better represents present climate patterns and phenomena. Therefore Shukla et al. (2006) suggest relative entropy as a metric to measure the reliability of climate projections and applied it exemplarily to climate sensitivity projections. Since they found that those projections with the highest values of climate sensitivity reveal the smallest values of relative entropy they concluded that the actual climate sensitivity is to be expected to lie near the upper boundary of the state of the art value from $3 \pm 1°C$. There are comparable ideas of better metrics than those normally applied but as all of them, including those of Shukla et al. (2006), are supposed to require more time and effort and are not proved to be really superior to standard methods, they are not comprehensively applied.
An exception is perhaps, as the near future will prove, a recently successfully applied

metric to test and improve climate modelling using so-called transfer functions, a term the authors took from control engineering literature. MacMynowski and Tziperman (2010) show that this method improves GCM simulation of the ENSO cycle which is unsatisfyingly represented in state of the art GCMs. The transfer functions are found by dividing the ENSO model into simpler models, representing only some of the possible drivers of the ENSO cycle, with separate input and output whereas the latter can be traced back to dynamical processes in the simple model. The GCM is then retuned to simulate not only observations but also the transfer functions which could omit the right results for the wrong reasons in the GCM.

Conclusion for general model testing methods No matter how simple or elaborate the mathematical methods and pattern detections are, they do not provide independent criteria indicating which testing results count as passing the test and which as failing.
The latest IPCC report's chapter on model evaluation addresses this problem of the lacking criteria. The testing of model performance against observed data is only meaningfully possible if practical adjustments are chosen in equal measure. The length of the times series as well as the influence of forcing are mainly chosen due to the researcher's needs. That is, there are mathematical testing methods for model performance but their results depend very much on the testing conditions and on the researcher's personnel demands. As the methods do not provide independent criteria there is no standard as to the maximum of acceptable errors or minimal agreements between simulation and observation.
A judgment about whether a simulation is to be accepted or not on the basis of commonly shared rules would be desirable but not comprehensively possible as becomes clear in the discussion in the following chapters. An exception is the very weak rule that climate models projecting the near future are to represent a global temperature increase independent of the emission scenario to account for the inertia of the climate system in response to the emissions already sent into the atmosphere. Yet there are scenarios conceivable that must not fulfil this criterion, namely the simulation of a climate that is engineered to have less radiative forcing than today's natural conditions provide.

To nevertheless have some kind of common ground for model testing Randall et al. (2007) argues for negative criteria instead of positive. That is, according to the IPCC there are at least three common rules to identify differences between simulation and observation which do *not* denote bad model performance. Climate models are not recommended to reveal the following features in order to pass the test. These points are:

1. unpredictable internal variability,
2. expected differences in forcing,
3. uncertainties in the observed fields.

A model may be considered as performing well even if there is no commonness in patterns concerning these fields and errors are huge.

The first item refers to untypical events in observational data as the accumulation of extreme events. The second point says that a model run from 1850 is obviously not going to represent climate data from 1990 under CO_2 forcing. Whether such a comparison is meaningful at all is a different question. The last item is self-explanatory: if there are uncertainties in data measurement it cannot be represented accordingly.
However, if the model is not expected to cover the climate states above climate models also fail to reproduce past observations in detail (Stainforth et al. (2007)). Here detail means to reproduce observed data within the observational uncertainties. As climate models do not include many processes known to be important, this disability is not a big surprise. But it leaves the decision to accept or disregard a model simulation to the researcher alone.
Taking these facts into account we have to come back to the conclusion drawn from the comparison of Popper's and Kuhn's analysis of scientific testing (see section 4). If there is one method applicable it is the critical discussion in peer communities. So far it seems to be necessary to find criteria for good model performance for single cases in presenting the results of mathematical model testing to a review by peers, which is more or less actual practice.

5.1.2. Paleo data

While the forcing is well known for the present day climate it is very uncertain for paleo climates. Thus testing model performance against reconstructed paleo climate data has greater limitations, yet to corroborate scenarios with fundamental climate changes a paleo climate data comparison is the only possible data comparison.
Paleo climatic data has the great advantage of showing climate states generated under different external forcings, such as a different angle of the earth's rotation axis. The external forcing factors of climate models can be tested in comparison. If they reveal the 'right' influence on the simulated climate system the paleo climate should be reproducible. But due to the different system conditions results of such data comparison are not fit to corroborate a model if the model and proxy data matches. Only the failure of reproducing proxy data can give hints of modelling incapabilities. Mostly GCMs are indeed able to qualitatively represent paleo climates but fail at quantitative representations. But inconsistencies in quantitative data reproduction is often not taken as failure in the model but in the paleo data (Betz (2006) p75).
An additional problem of paleo data model data intercomparison is the time scale on which historical climate changes occurred. This scale is normally longer as complex models like GCMs can simulate with admissible computational cost in an acceptable time. But the most important problem in using paleo data is its sparseness. The data we have is gained from very few ice cores and tree rings from very few regions of the world, thus time series from paleo data contain only sparse spatial information. As data is also needed in the modelling and tuning process there is often no independently measured data left for comparison, which does not mean that scientists use the same datum twice but rather that they have to split up the data set. The residual data set maybe

very small and the question already asked in section 4.1.2 discusses whether this data is really independent. Thus confidence in long-term behaviour of climate models can only be gained through the internal consistency of the models and understanding of physical processes.
As consistency in climate models is especially testable within small modelling approaches on the less complex end of the model spectrum it is the more crucial for the model governing sets of equations. Modelling approaches do not result from coherent and complete theories but of simplifications of equations of motion and thermodynamics. Whether such simplified equations are consistent in themselves and when compared to accompanying conditions can best be checked by modelling them. If the resulting climate variables bear any resemblance in magnitude and scale to climate observables, consistency is given. Of course consistency checks like this are best conducted via models on minor hierarchical levels. As consistency is a big advantage of conceptual models it is also more important on that scale. For complex models consistency in the governing assumptions cannot be checked via the model as the causes of unphysical model behaviour cannot be identified thoroughly.
Thus at the lower end of the model hierarchy climate models are used to test theoretical models and vice versa. This means that unexpected behaviour of the climate models may lead to new theoretical models which, if consistent, may explain real world phenomena and deliver a greater understanding of the climate system.

5.2. Model comparison

Apart from hindcasting with climate models, comparing models to present and paleo data, comparing different modelling approaches is most important for model testing. Comparing GCM model output with modelling results of an energy balance model is exemplary for intermodel comparison. An intramodel comparison is given if only one or a few parameters of a modelling approach are varied. But the border between both approaches is not fixed. On the one hand both methods can be combined. And on the other hand no criterion can be given as to how much model variation marks a different modelling approach.
The main aim of model comparison studies is to understand differences in the modelling approaches and to trace cause and effect chains. If such chains are identified and demonstrated to hold true for repeated modelling approaches they become an important criterion for the reliability of new climate models of the same level of the hierarchy.

Model comparison is normally not the comparison of two modelling approaches but of model ensembles. Ensembles are large groups of parallel model simulations. The varying results across the ensemble members give an estimate of the spreading of results. A very common error of scientists is to take this spreading as the real uncertainty but it actually represents the variability of the models. There is no such thing as *the* uncertainty of a modelling approach, as the space of possibilities of variables is unknown. It is thus

only possible to assess the uncertainty of certain aspects of climate modelling and some variables.
Ensembles of models are used for studying the range of plausible climate responses to a given forcing and can be generated either by collecting results from a range of models from different modelling centres, so-called multi-model ensembles, or multiple model versions within a particular model structure are generated. Variability resulting of multi-model ensembles refers to uncertainty in model internal climate variability and modelling differences. An ensemble of several versions of the same modelling approach, a so-called perturbed physics ensemble, is generated by systematic variation of internal model parameters within plausible ranges with the aim of estimating modelling uncertainty. The uncertainty associated with model internal climate variability can be analysed in particular if an ensemble is generated using the same model but under different initial conditions.
Comparison of large ensembles has only recently become possible as consumption of computational power is high. Multi-model ensembles of computationally intensive modelling approaches are still not possible for more then ten models at the most, which is not enough for statistically relevant model testing.

5.2.1. Intramodel comparison

Apart from one-to-one model intercomparison within the spectrum of models the perturbed physics approach is a very commonly used method of model testing but as discussed below it is not a probabilistic method and thus not stochastic climate model testing.
That is, different modelling approaches are not used within an ensemble but rather slightly differing simulations of the same model and also model version. A perturbed physics approach is an ensemble of one climate modelling approach with a range of parameter values simulated. This multi-parameter ensemble allows for the detection of good parameter values. The samples of parameters labeled as good are chosen according to the physical consistency of the model and again in comparison to observed data. This way it is possible to see which climate variables show the greatest uncertainty. Sampling the whole range of physically possible parameters gives a comparable complete range of possible values of state variables in the climate model and important climate variables as climate sensitivity (see for example Schneider von Deimling et al. (2006)). That is, in the first step parameters can be constrained to a physically meaningful range, while in the second and most important step it is possible to show the range of uncertainty of key climate variables, at least theoretically. In real life computational power and time limits the actually sampled parameter ranges substantially.
A very famous approach to stating the parameter uncertainty in predictions of global climate change was documented by Stainforth et al. (2005). It is so far the largest ensemble of simulations of an individual modelling approach. The experiment allowing for this huge ensemble is the 'climate*prediction*.net'. The model generating the ensemble is a version of the Met Office Unified Model. It is a GCM based on the AGCM HadAM3

developed at the Hadley Centre in the United Kingdom. With the help of many volunteers, allowing the model simulation to run on their private computers, it was possible to obtain 2017 unique simulations, which is more than 100,000 model years up until 2005. As the project still goes on there is data of 44,548,488 model years from seven different ensembles available today. There are also other models included now but all based on the Hadley Centre models HadAM3, HadSM3, and HadCM3.
Stainforth et al. (2005) show how very important large ensembles are to really check out the range of model uncertainty taking climate sensitivity as an example. In the ensembles with physically meaningful parameters the climate sensitivity ranges from $1.9°C$ to $11°C$ within the confidence interval between 5 and 95%. This is much broader than the usually given range, e.g. from IPCC, of $1.5°C$ to $4.5°C$. Even if the mean value in this perturbed physics approach lies at $3.4°C$ there is no physically valid reason for assigning a higher uncertainty to a high value of climate sensitivity, at least within this ensemble study. Knutti et al. (2006) suggested in a study of the influence of seasonal cycles in this perturbed physics project that modelling approaches with high climate sensitivities overestimate the amplitude of these annual cycles. Knutti et al. (2006) additionally analysed the results of the perturbed physics ensembles comparison in sensitivity studies. Only the combination of both testing methods allows for a redirection of likelihood towards the IPCC climate sensitivity of $3 \pm 1°C$.
This example shows that one testing method alone is not sufficient for a meaningful assessment of variable and parameter ranges. And it also illustrates that the term likelihood, used for want of a better, is misplaced. Perturbed physics ensembles cannot provide objective probabilities for climate projections and even less for model parameters. Whether subjective probabilities could result from such an approach is the topic of chapter 6.
Even in this ensemble the parameter ranges are not so broad that all possible values for all important parameters are included. On the contrary, a few parameters are perturbed within a range the modeller thinks more or less likely. That is even if a comparably small field of parameter uncertainty is checked within this approach the uncertainty it reveals for climate projections is enormous.
This fact illustrates the dilemma climate science is confronted with. Ensemble runs sufficiently large to test parameter uncertainty are impossible to carry out for large GCMs. Thus the projections given with the models have unknown parameter uncertainties.
One way to make the dilemma less pressing is to test the parameter uncertainty with respect to key variables as the climate sensitivity and to consider the resulting range within future GCM approaches. Testing of these models will then include a broader range of parameters. The next chapter will show that this is especially hard to consider if using deterministic climate models, which is usually the case. This is not to speak of the additional amount of time and power consumption.
The second way is to use less complex models within the ensembles to quantify parameter uncertainty and compare them to more complex models. The following section is mainly dedicated to this topic.

5.3. Model intercomparison

In model intercomparison studies different models and model versions of the same hierarchical level are used as an ensemble to provide a range of simulations for an experiment. A very famous way of getting really big ensembles is done in the project climate*prediction*.net where single simulations of different model versions are sampled. The boundary between intra and inter model comparison is not fixed here. But as the aim of this study is not a perturbed physics ensemble but to collect a huge amount of distributions of model futures it may count as an inter-model comparison project. The resulting distributions are scattered around a mean value for each variable. But they do not provide probability ranges of variables as ensemble means have no connection to the real world as they are inter-dependent and all do suffer from systematic errors.
The range such model intercomparison ensembles provide is the lower boundary of the maximum range of uncertainty (Stainforth et al. (2007)). Thus multi-model ensembles do not provide a method of labelling models as superior or inferior as they are all empirically inadequate. Such studies do not compare models to reality but provide insights into model performance and are thus an important mean of understanding differences in models. For example, models can be judged according to their representation of the physical processes they are designed to represent, as model ensembles run with comparable physics can reveal inconsistencies in their approach if simulated distributions are inconsistent with respect to important variables. Right results for wrong reasons can also be identified in such studies if models represent patterns they are not expected to show given their physical basis.
Model intercomparison projects are especially suited for identification of systematic errors in climate modelling approaches describing the range of uncertainty. Systematic errors are basically differences in statistics of model variables compared to observations of that variable, that is, e.g. a significantly higher or lower simulation of state variables like surface temperature. Intermodel comparison allows for detection of systematic errors in single modelling approaches as well as in all models. In single models the source of the error may lie in the numerics, a special parameterization or representation of dynamics or the inclusion of chemical details. Systematic errors detected in all models are of special importance as they are only detectable in multi-model ensembles and hint to a lack of climatological understanding throughout the whole climate modelling community. Such an error may occur due to an undetected but important dynamical or chemical process, an overinterpretation of the influence of another, or a mistake in a globally applied parameterization.
The definition of systematic errors shows that model intercomparison approaches are also not independent of observational data. Model intercomparison projects can only be seen as an additional step in sophisticated model evaluation. For those variables of the model that relate directly to observable and observed data a model data comparison should be undertaken thoroughly.
Comprehensive model intercomparison projects are suited for but not especially designed to test a GCM modelling approach, but rather to make climate projections. Huge undertakings of model intercomparison are made previous to the IPCC assessment reports to

provide its scientific basis. The initiator of the most important intercomparison projects is the World Climate Research Program's Working Group on Numerical Experimentation. It provided the first Atmospheric Model Intercomparison Project (AMIP;Gates (1992)) in 1992 and are now undertaking the fifth phase of Coupled Model Intercomparison Project (CMIP;Meehl et al. (2000)) to deliver the facts for the fifth IPCC assessment report to be released in 2013. The basis for the latest IPCC assessment report (AR4) is CMIP phase three. Additionally there are other types of multi-model ensembles used within the research for IPCC AR4 and, most notably, the Special Report on Emission Scenarios (SRES; Nakicenovic and Swart (2000)) that provides the time evolution of the climate system under different climate forcing scenarios.
In contrast to that the CMIP scenarios do not have a strict and realistic time evolution in the model output. Thus they are used to study the types of climate change and the range and differences of model response to an idealized CO_2 increase of 1% per model input year scenario and three forcing scenarios based on SRES simulations. The increase of CO_2 of 1% per year is one of the two experiments to measure a 'standard' that should quantify the reaction of a global GCM to other forcings. One of these measures is the transient climate response (TCR) that is the global mean surface temperature change in a 1% per year increase of CO_2 until a doubling of preindustrial values is reached. The other measure is equilibrium climate sensitivity(CS) which must by definition be calculated with a doubling of CO_2 equivalent forcing. Both measures are for temperature response and can be computed within the same simulation approach where the TCR gives the immediate temperature response of a CO_2 doubling and CS the equilibrium response.
Such ensembles aim at cancelling the individual model errors in adding up all simulations to compute mean variables. It could be seen (Meehl et al. (2007)) that averages across structurally different models do indeed represent observed means better than ensembles of individual or identical models.
As the models are not weighted in any form a model intercomparison like this is not a stochastic ensemble, therefore it is not possible to quantify the uncertainty of models and variables within such an approach. For that kind of testing model intercomparison approaches are invented using Bayesian learning strategies (see section 6.2) or perturbed physics ensembles. Stochastic aggregation of model ensembles would be desirable but so far the only method is expert elicitation which is questionable as the range of uncertainty given by experts varies extremely (Zickfeld et al. (2007)) and experts able to provide founded views on several models are rare.
Nevertheless multi-model ensembles allow for statistical analysis of the ensemble results and thus decrease uncertainty. Smith (2002) invented a formalized test to quantify how well a model is doing and thereby to underline the advantages of big statistical model ensembles. He proposes calculating a temporal credibility ratio τ_{cred} that holds for a variable in an ensemble of comparable simulations. The temporal credibility ratio is defined as

$$\tau_{cred} = \frac{\Delta t}{\tau_{ave}} \; ; \tag{5.1}$$

whereas Δt is the smallest time step in the model and τ_{ave} the smallest duration over which a variable has to be averaged before it compares favorably with observations. For single climate models τ_{ave} will normally be big and shortens with increasing number of simulations. Computing the temporal credibility ratio permits an easy overview of on which spacial and temporal time scales the model performance is good and which most urgently need attention.
While the temporal credibility ratio can give information about the uncertainty of certain variables in a simulation grand ensembles using different models and model versions can be used to learn about uncertainty of a distribution of the state of the system. The seriousness of uncertainty of this kind can be gained via contrasting climate variable distributions of various modelling setups. The more these distributions differ the bigger the uncertainty. If they differ only in terms of detail larger ensembles can be used to find out differences and thus get information for model improvement. If the difference is so overwhelming that details are irrelevant, comparative data can be used to disqualify modelling setups.
The temporal credibility ratio given by Smith (2002) is a very simple example of formalized inter model comparison. Actually applied formalisms are more complicated but the idea is comparable.

A different mean to corroborate the results of model simulations is to compare output from complex models with those of energy balance or other box models. If the projected distributions for selected variables are comparable they are more reliable for both models but especially for GCM output as the right results for the wrong reasons seems improbable if identical parameterisation is excluded. But such a comparison is of course only possible for thermodynamic variables like temperature as there are no dynamics represented in energy balance models. Thus it gives no hint at all of the dynamical performance of the high resolution model.
Simple climate models play an increasingly important role in model intercomparison approaches. Whereas the model forcings taken from SRES are used equally in the idealized approach, to find out types and ranges of climate change, the reliability of these forcing scenarios can be tested within this different type of model intercomparison. Interhierarchical model intercomparison can be used to get information on the realisticness of the scenario choice in the forcing, that is, by sampling a large amount of SRES scenarios with simple models and EMICs.
The role simple models play in such an intercomparison is the simulation of interaction of global variables. As this type of models consists of a set of large scale boxes only, simulating the atmosphere, the oceans, hemispheres, or simple biochemical cycles they do not need comparable computational power. Thus, initializing the model with observed or AOGCM data and implementing the forcing they give the range of possible development of the main variables, based only on the energy balance of the represented system. Accordingly they are also used to apply output data of higher models to a wider range of alternative forcings. Within such a comparison the influence of just the energetic development of the AOGCM can also be depicted. Simple model and GCM should take comparable paths in overlapping time periods with similar initializing and forcing.

If they strongly differ it hints to structural errors in the GCM modelling approach. An example is perhaps Rahmstorf (2007), even if the comparison is a bit lopsided, he uses a very simple modelling approach to compute sea level rise. In paleo data he found a proportionality between global mean temperature and sea level with the proportionality factor 3.4mm per year. With this simple linear approach he is able to predict the actual sea level rise of the last few years much better then sophisticated GCMs are able to. This is a clear indication that GCM dynamics lack some important processes and are also unable to cover it by parameterisations. The simple model is not conceptual but semi-empirical, thus it is not based on first principles[1] which would make the evidence even more compelling.

Another method of testing AOGCM modelling approaches with simpler models is done via emulating the results of GCMs with simpler models or EMICs as for example done by Meinshausen et al. (2008) with GCM results used for the IPCC AR4. Reproducing the results of complex climate models using simple models is called emulation. The simple models simulate the results by abandoning physical interpretability of it. If the emulation models are unable to represent the AOGCM data for a wide range of scenarios it is again a hint of modelling errors in the GCM approach.
The emulation of the trouble causing ECHO-G results (von Storch et al. (2004)), discussed in section 3.5, with the EMIC MAGICC, helped Osborn et al. (2006) to identify the climate drift in the AOGCM simulations.

In contrast to simple models EMICs include the main processes as represented in AOGCMs but on a larger scale, which makes quantification of uncertainties almost impossible using EMICs (Meehl et al. (2007)). But EMICs are of enormous value concerning parameterizations, one of the main sources of uncertainty in climate modelling. Within modelling approaches of intermediate complexity it is possible to sample the space of the parameter and thus to develop better parameterizations and sampling of the parameter space makes true probabilistic model projections realisable. How this can be reasonably done is the topic of the following chapter.
Furthermore EMICs can be used to investigate the large-scale effects of coupling earth system components. As AOGCMs are the only climate models which include the emergent internal variability of the climate system, interhierarchical comparison of models allows for measuring the influence of that variability on global climate development.

[1] First principles is a term, commonly used in climate science, to refer to physical laws we believe to be true, e.g. conservation of energy, and to result in parameter constraints, e.g. velocities which are not negative. A model not based on such principles need not to violate them but they did not govern the model development.

6. Probability in climate modelling

6.1. Reasons for applying probability

The socially most important reason for modelling the climate system is not the understanding of the system in itself but to be able to make prognoses of its future development. A perfect understanding of the system is impossible and so is forecasting future impacts of global warming with certainty. The next thing to a certain prediction that might be achievable would be a set of future scenarios with assigned probabilities of occurrence. Thus probability plays an important role in public discussion of climate science.

Predictions made with climate models, the published output of such models, seem to be probabilistic but normally they are not. As results are often given as ranges they appear to have confidence intervals. But this is normally not the case.

A probabilistic model projection is thus a prediction of state variables and their associated probability which is a statistical weighting of climate scenarios or ensemble members. There are several methods to do this weighting of which some have proven valuable for climate predictions and will be discussed in detail, while others entail disadvantages for a broad implementation and are thus only applicable in special modelling setups.

But is it actually possible to talk meaningfully about probabilities in consideration of climate modelling results, given the degree of uncertainty? At least for objective probabilities in terms of relative frequencies the question is easy to answer. Relative frequencies are only possible to measure if the range of possible values is known; independent of the actual variable. Given the analysis of part II this is imperfectly known for observables and unknown for most parameters. Therefore it is impossible to assign objective probabilities to simulated climate development paths.

But this is not necessarily true for subjective probabilities. And the fact that objective probabilities are not to be assigned to model output does not mean that it is meaningless or impossible to apply probability functions on model input. As statements about the probabilities of possible climate change impacts are very important for society it is especially important to be serious in their assessment.

To be able to judge probability statements the possibilities of their coming into existence are reflected upon in the following: to finally answer the question how to take probability statements in climate modelling contexts.

Before discussing the different ways in which probabilities enter climate modelling some terms important within this context must be fixed, especially the difference between deterministic and stochastic models and their relevance in climate modelling.

6.1.1. Deterministic vs. stochastic modelling

A simple example for a linear deterministic system is:

$$f(x+1) = 2f(x) + 3.$$

From a system like this it is quite straight forward to build a deterministic model. For example:

$$g(x+1) = ag(x) + b.$$

To use such a model to simulate the system and make predictions the parameters a and b must be fixed. Furthermore initial conditions are to be prescribed to start modelling. Configuring parameters a and b is possible via fitting on measured time series. Initial conditions for prognoses come from currently measured data. Having done this successfully the system is understood and predictions can be made using the model. Whether it was indeed a success can be tested on the basis of observational data which can falsify the model output.

Normally systems we are confronted with in our physical environment are not simply linear deterministic. There are basically two ways of making the system more complicated: the system can be stochastic or nonlinear, or it can be both.
Nonlinear dependencies within a system make it harder to find parameters to build models whereas stochastic behavior can only be modelled adequately by using stochastic tools within the modelling approach.
A stochastic system could be for example:

$$f(x+1) = 2f(x) + 3 + \epsilon(x);$$

with $\epsilon(x)$ being a random variable drawn from a distribution at every time step x. Thus the time series produced by the system also depends on randomness. Now, the problem for modelling purposes is that only one of many possible time series is realised in the observations. To nevertheless simulate such a system one can build a model including a noise term:

$$g(x+1) = ag(x) + b + \epsilon'(x).$$

A simple fitting is not possible due to the noise. That is, for every set of parameters (a, b) many random realizations are to be calculated while counting the frequency of hitting the actual time series. This leads to a probability distribution of the parameters which in turn, with current data as the initial conditions of a prognosis, leads to a probabilistic prediction.

A deterministic system as described above is in the words of Laplace a system where the epistemic features of some ideal predictor cohere perfectly with the ontological features of that world (Eagle (2005)). In such a world it would be possible to predict with certainty the development of a physical system as it would be possible to know the current state of that system and its dynamics exactly.
Our world and human epistemic abilities are not like that. We are unable to ascertain the current state of a system. This is principally due to measuring inabilities as every measurement changes the state and pragmatically due to the impossibility of measuring with arbitrary precision. But it is also due to our limited knowledge about relevant variables. At least this is true for micro states, on a quantum mechanical scale but especially the latter is also a fact of macro states which is relevant for climate science.
For Eagle (2005) the latter facts untie the band of prediction and determinism in such a way that the question of determinism becomes irrelevant in terms of predictability. To put it in more simple words: whether the climate system is deterministic or not is unimportant concerning climate predictions. At least it shows that we cannot find out whether the world is strictly deterministic or not, which has the same consequence said of strong determinism, we definitely do not need a perfectly deterministic world to be able to make successful predictions.

But independent of the question of whether the world is deterministic or not we must decide on a type of model that is preferable for climate prediction making.
If we consider complex systems instead we cannot predict the full trajectory of a system but we can make predictions about chaotic systems using statistics and stochastical tools. These predictions are not perfect but the better the system is analysed and understood the closer we come to giving probability distributions. This is again independent of the system being deterministic or truly stochastic.
In stochastic systems the knowledge of the initial state would not suffice to accurately predict the system development even if it was perfectly understood, for deterministic systems this was possible. But as understanding the system is impossible the question of systems needs not to be answered.
Nevertheless, there are important differences in modelling this world deterministically or with stochastic models. Especially concerning the quality of predictions there are hints that stochastic elements in modelling approaches improve their behaviour. Both on the input side and in postprocessing model output, stochastic process representations improve process simulation. Stochastic methods, namely statistics, also do not need truly stochastic models but can be applied to deterministic model output. The use of statistical methods would only be useless for predicting if confronted with perfect randomness which is almost nowhere and never to be found. At the most, climate patterns and regimes are overlaid with practically random noise, which can make life harder but is not a fundamental limitation to predictions and does not shed light on the question of whether the system is deterministic or not. In this context the difference between noise and internal system and model variability respectively is important to highlight.

Statistics as a method used to make predictions is also a method which depends on measurement abilities. The usefulness of statistical analysis rests on the availability of high quality data. Thus statistically produced predictions highlight again the crucial role of measurement and how it separates determinism from predictability.
Our predictions cannot be any better than our measurements of the system variables that are subject to prediction. In practice they are worse. As O'Connor (1957) puts it, "our predictions can do no more than specify a class of possible events." This is true for any event being measurable in a metric system but even more so in systems without metric. The precision of predictions depends on the language available to describe the system within which the prediction is made. The less precise the description of a system and therewith the measurement of variables the less precise the prediction can be. Predictions about variables not able to be measured at present are if anything open to prophecy but never to prediction. That is, there is always a whole class of possible events that will correspond with the predicted event; given the prediction was a success at all.

The situation in climate modelling is that complex climate models are deterministic but the observed time series are noisy, due to unresolved processes. To be nevertheless able to tune deterministic models with the help of noisy data, the noise is eliminated using filters. The easiest way is just to take the mean. This is pretending the data is of a deterministic system or pretending that nature realised only the expected values of the variation.
This may be acceptable for linear systems and models but for the second case, nonlinear systems, it is not. Nonlinear systems are such that small changes in data may result in big differences. Therefore the filtering is inadmissible due to the assumption that in spite of average building over time steps or areas in space neglecting noise makes no difference.
Mathematically this is easy to comprehend. Filtering noise is as much as using the expected value of all possible noise realizations via average building over the data.
If E is the average then

$$\begin{aligned} E(f(x+1)) &= E(2f(x)+3+\epsilon) \text{ equals only} \\ &= 2f(x)+3+E(\epsilon) \end{aligned}$$

if the noise goes in a linear way into the system, which is definitely not the case concerning the climate system.
Thus it is in a sense illegitimate to filter the noise from data to compare it afterwards with model data from a deterministic climate model. The proper way round would be to add noise to a deterministic climate model to compare the output with equally noisy observation data.
This method is indeed applied in climate modelling science under the name of 'detection and attribution' but the previously explained method is more common in complex climate modelling.

To summarise it can be said that it is irrelevant for climate modelling whether the climate system is truly stochastic or not. Nevertheless climate modelling needs to deal with noisy behaviour on a macro scale, may it be induced by internal dynamics within complex climate models or by limited resolution of climate data.
The following section is about different methods to relate modeled climate variables to probability functions. It can be seen that the applicability of statistical methods in particular is more sophisticated within stochastic modelling approaches. The increasingly important statistics subsumed under the term Bayesian inference is especially only possible in noisy climate models.
However deterministic climate models are of great importance for climate projections. GCMs are impossible to realise as a stochastic approach due to the enormous computational power such a model would consume. But concerning probability predictions of climate impacts deterministic climate models may be given too much relevance.

6.1.2. Uncertainty

In chapter 3.1 I discussed the uncertainties related to climate modelling. There are two main types of uncertainties determining climate modelling, fundamental uncertainties and model uncertainties. While the former uncertainties cannot be overcome in principle, the latter can principally be overcome. But due to insufficient knowledge and technological skills we have to live with model uncertainties as well.
In the context of model projections the uncertainties directly linked to the climate model in particular are of special importance compared to general uncertainties resulting from our lack of understanding of the climate system. Model uncertainties are in particular parameter uncertainty and structural uncertainties in the modelling approach. Both of them are highly interlinked but constitute uncertainties that are reduceable using different methods. Parameter uncertainty relates to all uncertainties concerning model parameterizations, in particular the choice of parameters. Structural uncertainties relate to uncertainties in model design. This includes the modelling concept: which processes are modeled, which are neglected, what kind of resolution and grid the model has, and what types of numerical schemes are chosen within the modelling approach.
As there is normally no obviously best way to model a certain aspect of the climate system uncertainties, it is especially difficult to assign structural uncertainties with probabilities. Nevertheless the assignment of probabilities to model output variables is mainly to adequately present some uncertainties of the modelling approach. This is done via ensemble modelling on the one hand and statistical methods on the other.
The structural uncertainties are untestable within a modelling framework as they account for the approach. But model intercomparison ensembles can reduce uncertainties of that sort. A shortening of parameter uncertainty can be undertaken internal to one model, using statistical methods and perturbed physics approaches.
The following sections give an overview of the different ways of dealing with uncertainty. In conclusion it is argued that especially ensemble approaches are indeed appropriate tools to deal with uncertainty but hardly adequate methods to assign climate variables with probabilities that deserve the labelling.

6.2. Statistical modelling

One method to enable climate models to make probabilistic predictions is to assign probabilities to the parameters of the model and thus propagate those probabilities to the output by running the model. In deterministic modelling the parameter values are taken to be the best guesses of the modeller, or they result from some kind of tuning procedure, as best values to represent some given feature in the model output. When turning to probabilistic projections, probability distributions need to be assigned to the model parameters. In a first step those probabilities can stem from expert judgments and represent subjective beliefs about the true parameter value. Once such an a priori probability measure is appointed the information incorporated in all types of data or measurement can be used to update the knowledge about the parameter represented by the probability density function according to the Bayesian approach. Bayesian inference is thus a sophisticated method but not independent of the classical statistical method of inference, the frequentist inference.

The classical statistical method of inference is statistical hypothesis testing which subsumes several methods of statistical analysis. In principle this is a method of confirmatory data analysis used especially to assign confidence intervals to model parameters on the basis of observed data.

The idea is to have a null-hypothesis and an alternative under the assumption that the null-hypothesis, for example equal distribution of all parameter values in a given range, is true. Before trying to confirm this assumption a critical number of confirming data points must be set. This critical number determines the probability of a so-called Type I error within the testing process, that is a false positive confirmation. This is the tricky part of the procedure as it is often difficult to assess a number that is captured by data and delivers a sufficiently small probability of Type I errors. There are many different methods to assess the critical number and calculate the corresponding error. Before the test is actually performed, the desired probability of a Type I error is determined. But errors of Type II also exist, which is a non-confirmation of the hypothesis although it is true. Normally it is not possible in parameter testing contexts to get small probabilities for both errors.

The null-hypothesis of the testing procedure could theoretically be a subjective probability density function but according to the confirmation data available only confidence intervals can plausibly be confirmed or rejected within frequentist inference.

Bayesian Learning and Bayesian Inference, which also subsumes several methods of statistical inference, is instead a method of sharpening subjective probability functions.

The principal method of Bayesian Learning is the application of Bayes' Theorem:

$$p(h/e) = \frac{p(e/h)p(h)}{p(e)}.$$

This theorem marks the central basis of empirical hypotheses corroboration in Bayesianism (Carrier (2006)). Datum e corroborates hypothesis h if e leads to an increase of probability of h. $p(h/e)$ is thus the probability of the hypothesis in terms of the reliability of hypothesis h given datum e as conditional probability of h. It is the level of corroboration. $p(h)$ is the prior probability of h before observation of e. The prior probability is a subjective measure of reliability of h. $p(e/h)$ denotes the likelihood of e on the basis of hypothesis h. That is the likelihood of observing e, assuming h is true. In contrast to that $p(e)$ denotes the probability of e independent of h.

Translating the theorem from a philosophical to a more applied vocabulary, the following example (Lorenz et al. (2010)) shows an application of the method.
Using the Bayesian approach to constrain model parameters α the crucial input is the prior probability $p(\alpha)$ on the one hand and the likelihood function $L(\alpha/y)$ on the other. $L(\alpha/y)$ denotes the relative plausibility of parameter value α in view of the evidence y. $L(\alpha/y)$ equals the probability $p(y/\alpha)$ for observations y and the parameter assumptions α.
Within the application of Lorenz et al. (2010) the parameter "ocean diffusivity" α and others are constrained using the duration of the 8.2 ka cold event as data y. The 8.2 ka event was an abrupt cooling of $6 \pm 2°$C, at least in the northern hemisphere, which took place 8200 years before present with a duration of approximately 160 years. The event is well reported in Greenland ice core data.
The EMIC CLIMBER 2.3n was used to carry out this study. The prior weighting of α is done from physical assumptions. Both negative and very high values of the model parameter representing ocean diffusivity can be left out due to the fact that ocean heat transport is neither negative nor arbitrarily fast. As the Bayesian approach affords not only a parameter range but a prior probability function normally a Gaussian like distribution with a rather flat maximum is assumed to represent sparse knowledge of the prior distribution which is also called the non-informative prior (Lorenz et al. (2010)).
The likelihood $L(\alpha/y)$ of correctly representing the duration of the cold event for a given parameter value α is not known a priori but had to be established numerically. This was done by simulating the cold event with the noisy version of CLIMBER for many different realisations of noise stemming from freshwater forcing for each model parameter α. From these noise-ensembles of model runs for each parameter a relative likelihood of the correct duration is established by simply counting the number of "correct" runs, those that represent the event more or less as the ice core data. The resulting likelihood can then be applied to the Bayesian formalism to give a sharpened probability distribution of the parameter α instead of the mere guessed one used as a priori probability for α.
This method is especially good if the sampling is done more than once in light of independent data using the resulting a posteriori probability as the new a priori probability of the parameter in the next step.
Carrying this procedure out for every parameter, probability distributions can be assigned to the output variables of the model.
So far the Bayesian approach is a method to include reasonable subjective probabilities in modelling approaches. Furthermore Bayesian inference is a method used to test climate models as follows.
Stochastic models that represent noisy processes with certain shape of noise and a cer-

tain memory can be tested via formal Bayesian Learning. Via dicing the noise and fitting model parameters to time series from observational data the probability of reproducing the time series correctly is computed. In this way statements can be made of the form: given a model of structure X and a time series Z the probability that model parameters Y meet values Y' is Q. Additionally plausible parameter ranges of different models can be analysed which leads to conclusions that model 1 happens to be able to reproduce the time series well enough for a larger range of parameters than model 2. Thus model 1 is better in terms of more being general and more compatible compared to the time series.
Therefore a model can be considered as better tested if it compares to more data and thus longer time series; especially if the data comparison is relatively direct.
Again the quality of the observational time series is decisive as said above in the context of paleo data. The length of the time scale is crucial as longer time series allow for the detection of slower dynamics and longer memories. But there is a dimension of information quality that is not proportional to the length of the time series. The fact that an infinite time series represents the complete phase space does not correspond to longer time series representing more parts of the phase space.

But there are many constraints, especially technical ones, for a general applicability of the Bayesian approach in climate science.
First of all the model must not be too expensive in terms of computational requirements in order to be run many times for estimating the likelihood and sampling the Bayesian approach. Thus, the method is inapplicable to GCMs.
Secondly the number of parameters must be manageable and an a priori assessment of the parameter ranges and probabilities within the ranges must be available.
Furthermore the model must be stochastic. Sampling is only meaningfully possible if the forcing is noisy; that is what the n in CLIMBER 2.3n stands for. Within deterministic models, the probability of meeting certain data is either 0 or 1. This does not preclude the usage of Bayesian updating but only leads to a cropping of the support of the parameter distribution instead of a change of the distribution within the range. Even if classical testing of hypotheses is still available for deterministic models this large group of climate models is excluded from Bayesian probability assumption.

Concerning probabilistic predictions the Bayesian approach is superior to classical statistical inference as the procedure results in probability density functions instead of confidence intervals. Aiming at probabilistic assumptions of climate variables a true probability density function is preferred to such confidence ranges. This fact is not narrowed by the probability being subjective. Subjectivity in scientific contexts must not be understood as being in opposition to objectivity, but as related to the mind. According to Rougier (2007) all probabilities are subjective and there is no such thing as *the* probability but only the probability of oneself. To nevertheless obtain an intersubjectively plausible probability the scientist must be able to justify it. The scientist has to make plausible that he used all relevant information in a meaningful way to assign a probability density function to a parameter.

Recently Annan and Hargreaves (2011) have shown exemplarily for climate sensitivity that not to use all relevant information to determine prior probabilities lead to extreme pessimistic probability functions for the climate sensitivity. They could explain uniform priors to be a poor choice as they do not lead to robust probabilities for the climate sensitivity. Robust in this context means that the representation of uncertainty is "not strongly dependent on contentious assumptions or expected to change significantly in the very near future as incremental scientific progress occurs" (Parker (2010a), p.270). These conditions of justification and robustness become of even more importance in contexts of probabilities in GCM modelling contexts as described below.

6.3. Ensemble Modelling

Climate model predictions are made to investigate what climate would be like under different emission scenarios. Predictions for decision making discussions are almost always based on GCM projections. Thus the statistical methods discussed above are inapplicable to make probabilistic predictions as GCMs are not stochastic.
To nevertheless represent modelling uncertainty ensemble studies are carried out. As described above and in chapter 4 there are two categories of model ensembles: multi-model ensemble studies and perturbed physics ensemble studies.
Multi-model ensembles are used to reduce structural uncertainty. The more different modelling approaches are used to assemble the model intercomparison study the more different model structures are probed. Using multi-model ensembles instead of single models has proved to provide more reliable (see chapter 4) representations of state variables. Thus the model uncertainty is reduced, simply due to the fact that the range of sampled parameters is bigger and different modelling concepts are compared to each other. At the same time also the uncertainty in the initial conditions can be diminished if every model is run several times with perturbed initial conditions. In contrast to the assessment of some parameter ranges, for example the ocean diffusivity described in section 6.2, such an assessment is principally impossible for structural uncertainty as an adequate way of setting the right equations and numerics is not known.
In comparison to observed data, multi-model ensembles give better results than single modelling approaches, thus it is an adequate method of testing models and reducing uncertainty in a trial and error manner but without even sampling ranges of possible approaches. Moreover differences in models are not even chosen systematically but depend on which modelling groups agree to participate in a study like CMIP3 described in section 5.3 (Meehl et al. (2007)). Therefore the word probability is out of place in the context of multi-model ensemble output.
In contrast to state variables, multi-model ensembles do not remarkably increase the reliability of strictly parameterization dependent measures such as clouds or precipitation. Of course no model intercomparison study can straighten out systematic deficiencies in all models. Nevertheless the still bad representation of parameterized variables shows that a modelintercomparison is not the most appropriate approach to decrease parame-

ter uncertainty.
The standard approach to representing parameter uncertainty adequately is a perturbed physics ensemble, either in multi-parameterisation ensembles or multi-parameter ensembles.
A perturbed physics ensemble of a General Circulation Model consists of a large number of model runs obtained by varying the GCM parameters within their physically acceptable range (multi-parameter ensemble) or by applying different parameterisation schemes (multi-parameterisation ensemble). In principle, a multi-parameter ensemble could be comparable to a probabilistic ensemble but the input is just a range for parameters, not a probability distribution of this range. Thus it can not be decided whether a scenario is more likely than another. Perturbed physics ensembles also do not lead to true probabilistic projections.
However, perturbed physics ensembles are also carried out with simpler climate models. Simple models are less computationally demanding, thus a more complete range of parameters may be sampled. According to Parker (2010a), such a comprehensive parameter sample resembles a Monte Carlo estimation. Probability distributions that reflect uncertainty regarding parameter values are randomly sampled many times and a simulation is produced with each set of values selected via the sampling in order to estimate uncertainty in output variables. Sophisticated examples of such approaches combine Bayesian analysis to get input probability distributions. An example is that of Meinshausen et al. (2009), in which the advantages of a perturbed physics ensemble are combined with Bayesian inference which can be taken as an example for state of the art probabilistic modelling. Nevertheless it suffers from the limitations Bayesian analysis does, and it is also impossible to carry out for regional climate variables. Parker (2010a) additionally argues that rapid changes in global climate due to missing linear feedbacks in the modelling approach are possible but ignored within this study.
Looking at published output, as for example the probability distributions for global mean temperature changes in Meehl et al. (2007), the probabilities differ remarkably, although the emission scenarios are identical. The reason for that lies in the fact that perturbed physics ensembles do not account for structural uncertainty. Thus reliable probabilities are assignable to output variables of a certain model but not independent of the modelling structures. For the time being, no method is known to account for structural modelling uncertainties in terms of probabilities.
The only possible consequence of this analysis is that a reliable probability assignment to climate predictions cannot be justified. Only conditional probabilities are meaningful in the context of climate modelling.
Probabilities are inadequate measures to represent uncertainties in GCM climate modelling. It implies a precision of knowledge that is not given in these contexts.

Part IV.

Conclusion

7. Rules for dealing with climate models

Throughout this work the idea of guiding rules that my be used within the process of climate modelling is mentioned. In the introduction to the philosophical debate (section 1.1)this case study touches upon some arguments for why methodological rules for science cannot be meaningfully justified. Most importantly, this is because a teleological view of science in general is not an adequate description of science. However, the arguments against such a view might not apply to science in particular. There are certainly goals and aims of scientific undertakings, for example climate modelling, which many scientists would find they have in common, some very practical and some of a more ethical nature. A set of common sense rules, so to speak, which could make the modelling process more transparent for other scientists and increase the possibility and reliability of model comparison would account for these aims which are specified throughout this chapter. In particular such rules could avoid a misuse of data and thus hard the theory ladenness (section 3.6) of climate models. Kuhn, for example, even claims the existence of individual rules for each branch of science as a sign of true science (section 4.1).
In terms of explicit rules this objective may be too ambitious but such rules would certainly benefit modelling practice. It is a meaningful undertaking to formulate vague rules as they might at least point to the crucial steps in model development. It will perhaps be possible to refine such rules if working with them in practice. Rules are advantageous in dealing with criticism and hostility by demonstrating that rules are applied. It is then upon the critic to show that the rules have been breached, thus the burden of proof is, so to speak, placed on the opponent because rules for good practice provide arguments in favour of the methodology. Nevertheless, to discover rules to improve climate modelling in all mentioned fields is a task difficult to achieve. This is due to the fact that the prevention of data misuse especially involves a very detailed and technical analysis of the practice. Thus a comprehensive set of rules cannot be given here but rather an attempt at a starting set of rules, which might be improved by use.
However, in every climate modelling process there are error-prone steps and typical questions arise and an aim most scientists would agree upon is certainly to prevent such errors, of which some very basic ones are discussed in part II. In section 5.1.1 is described how the IPCC in its forth assessment report (AR4, Randall et al. (2007)) gave a set of negative rules to answer the question of whether a specific climate model test is to be regarded as a failure or not. These rules are neither detailed or technical but on a common sense level for scientists from within the community. Appealing to this very level of common sense, Petoukov et al. (2000) referred to some common sense rules for model tuning. As tuning is a very crucial step in modelling it is taken in the following

section as an example of identifying a set of rules in the following section. The subsequent section takes these rules and analyses whether they reveal some more general common sense rules applicable for climate modelling more broadly and if so, whether it would be good scientific practice to maintain them. To what extent and under which circumstances climate modelling is good scientific practice is then explored by reviewing an IPCC guidance note to extract implied common sense rules of climate modelling.

7.1. Common sense rules of model tuning

There is such a thing as good scientific practice which refers to a professional standard and involves avoiding trivial mistakes and using common sense. But generally this is an agreement of general, mostly ethical, standards which are not written down. Thus rules of good scientific practice are more implicit rules. The German Research Foundation (DFG), for example, published a memorandum concerning recommendations for good scientific practice (Deutsche Forschungsgemeinschaft (DFG) (1998)). The second recommendation is a request for universities and research institutes to shape individual guidelines for good scientific practice. Recommendation one describes what such rules have to be about, which are basically the "fundamentals of scientific work" including the documentation of results, honesty, and constant questioning of scientific findings. This is why good scientific practice is a more or less implicit agreement on ethical standards and not a set of methodological rules. Furthermore, these rules are to be stated in general and, if necessary, specified for individual disciplines.
The latter is especially not done comprehensively but only in ethically difficult disciplines as medicine. Nevertheless there is common agreement that manipulation of results, dishonesty, invention of data, etc., characterises bad scientific practice. But there are less obvious rules to follow to avoid the latter, and the following section will show that it is not always a straightforward undertaking to avoid manipulating data and modelling results.
Common sense rules are neither purely ethical nor methodological rules but guidelines to follow to avoid those activities listed above under the term bad scientific practice. Good scientific practice should in this context be interpreted accordingly as a practice to achieve exactly the same goal as common sense rules. Both practice and rules are necessarily individual for every small part of scientific endeavour. The common sense of an experienced climate modeler gives quite different advice to that of scientists from other disciplines or people from outside science. Hence if some basic rules were expressed it would simplify communication particularly between different scientific communities and at the interface of the scientific community and the public. Beyond that, the climate science community could benefit from more specialised rules. The remaining part of this chapter identifies some starting points for guidelines for good modelling practice in general and in particular for the example of model tuning.

7.1.1. Rules of model tuning

The fact that tuning is an ambiguous tool raised the question of whether there is a line to be drawn between using and abusing tuning according to certain rules. Be aware as well that tuning is a delicate but indispensable tool for climate modelling, the fourth IPCC assessment report includes a section (Randall et al. (2007), IPCC Working Group I, Chapter 8) on that topic. The IPCC AR4 suggests that tuning is justified, in the sense of not being abused (see section 3.4), if the following requirements are met by the applied tuning strategy. These are:

R1 Observationally based constraints on parameter ranges are not exceeded.

R2 The number of degrees of freedom in the tunable parameters is less than the number of degrees of freedom in the observational constraints used in model evaluation.

The first rule (**R1**) seems to be easy to adhere to since an important aspect of tuning is to vary those parameters not constrained by observed data. The second rule (**R2**) is an important one even if it is common sense to follow it. In doing so, overtuning is prohibited insofar as in terms of quantity not every output variable can be result of a specially tuned parameter. The predictive capacity of a model is practically zero if the number of the degrees of freedom of the model is equal to or exceeds those of the output. An interpretation of this rule that may also be of value for the complex process of climate model building is that the available theory must be used. Those components and processes of the climate system that are known to scientists and are computable within the modelling approach in question should not be parameterised. The latter dramatically limits the possibility of truly using all theoretical knowledge. Nevertheless the concrete wording quoted above gives an essential rule for climate model tuning but does not guarantee a model output resulting from physically well-grounded relations. These two rules presented in the fourth IPCC assessment report are not sufficient to prevent the abuse of tuning.
An alternative set of conditions to be met in order to tune responsibly is given by Petoukov et al. (2000) where four common sense rules for good tuning practice are given:

P1 Parameters which are known empirically or from theory must not be used for tuning.

P2 Wherever possible parameterizations should be tuned separately against observed data, not in context of the whole model.

P3 Parameters must relate to physical processes, not to specific geographic regions.

P4 The number of tuning parameters must be much smaller than the degrees of freedom predicted by the model.

This set of rules is ascertained as a kind of sidenote in a model description paper to prevent tuning abuse (see section 3.4) which is explained by the authors as avoiding "right results for wrong" reasons. The implications of this central modelling aim are discussed in the following section.

The latter set of rules entails the former more or less while emphasising the underlying physics. This is explicitly done in rule three (**P3**), but the first (**P1**) and second one (**P2**) also relate to that. Relating parameters to physical processes is difficult insofar as tuning is done in cases where the process cannot be modeled or is unknown. Thus it is not easy to make sure that a parameterisation used for tuning really represents a physical process. What is definitely avoided in adhering to this rule (**P3**) is the invention of e.g. impulse or heat fluxes when it is known that no such fluxes exist.
The first rule (**P1**) underlines the need to carefully choose tuning parameters and data, and instructs not to tune more than necessary. It is different from the first rule (**R1**) of the former set of rules. It is phrased more strongly because it says not only to tune within parameter ranges known from theory and first principles, but also not to exceed observed parameter ranges. For well observed parameters this is certainly important but in terms of sparse observational data an extension could be well justified. The first first rule (**R1**) might thus be a more appropriate wording as in particular parameters that are almost untouched by observation or that have implications or related parameters for which rudimental data exists, are used for tuning. Both first rules (**R1**,**P1**) emphasise the thorough check of all available data not to manipulate parameters where physical or empirical constraints exist. Furthermore the existence of these two rules underlines the need to be very careful with parameter ranges and the necessity for the justification of an extension of empirically known ranges.
The second rule (**P2**) should also be observed in order to be able to identify and avoid if possible wrong reasons for right results, which becomes even more difficult in a coupled model. Adhering to this second rule (**P2**) would make it hard to tune a model using methods of 'package tuning'. With this term I denote methods as the application of filters to tune several parameters together (see section 3.4). In very complex models filter tuning is necessary as it is computationally and temporally impossible to tune every single parameter. But this rule (**P2**) is very important for understanding dependencies between different processes. To assess the model output properly this is in fact a necessary procedure at least for submodels. Even though modelling results might be better if the complex model is tuned comprehensively, 'wrong reasons' are undetectable afterwards. Thus this second role is clearly an important addition to the rules presented in the fourth IPCC report as it advises scientists to analyse the model in terms of deconstructing the modelling problem and to independently build tunable submodels whenever possible.
The last condition (**P4**) is evident as following it prevents the model from being overdetermined. If there are as many tuned parameters as degrees of freedom no dynamically computed variables could be treated as such, as the tuning would have constrained them. This last rule (**P4**) does not contain any other implications than the one presented above as **R2**.

These four rules presented by Petoukov et al. (2000) give a comprehensible grounding for good tuning praxis and an overview of possible errors resulting from it. Nevertheless they are incomplete with respect to the second rule (**P2**). If it is possible to adhere to this rule (**P2**) the tuning is on a firm ground given the other rules are followed as well. But if it is not possible to keep this one essential rule number two (**P2**) there is much room left for the uncontrolled growth of tuning mistakes.

In less complex models, the hierarchy downwards from EMICs, the physics is so comprehensible and computationally efficient that parameterisations can be tuned separately. In GCMs with their interactive dynamic this is impossible. For some EMICs it is possible to fully map model parameter space but for GCMs, which contain up to hundreds of uncertain parameters it is not. For such models parameters are tuned by sphere, that is, separately in submodels for atmosphere, ocean, sea ice, etc. but not separately for every parameter. Dividing the model into submodels for tuning is especially helpful in avoiding tuning errors if the submodels could stand alone and the performance of the submodels could be evaluated separately before coupling.
Due to the interaction of all variables tuning is a difficult task, changing one coefficient to proper values may push other parameter behaviour out of an acceptable range in the sense determined above. Thus the tuning process requires great computational power the more complex the model is. It is possible to automatise the optimisation procedure until the best parameter setting is found. But mathematical optimisation procedures and ensemble methods which are particularly used for tuning GCM parameters do not always result in optimal parameter settings. Such methods make GCM tuning possible but adhering to the second rule (**P2**) impossible.
Considering the call for rules to improve tuning, automatised tuning seems to be a good thing as it follows an algorithm and therefore explicit rules. It has further advantages as it makes tuning of complex models possible. Furthermore, automatised tuning methods, as multi-parameter tuning with so-called filters is more independent of individual preferences than tuning parameters on the grounds of subjective choices. But whether such methods are really better, in the sense of less arbitrariness[1], and a better reflection of reality, as Tebaldi and Knutti (2007) suggest, remains to be shown also in their interpretation. Choices of parameters made by hand are often related to scientific understanding of physical processes which is the reason for insisting on following the second rule (**P2**) in the first place. If made thoroughly they may be better then parameters automatically chosen only to fit observations. But the latter method may be better in cases where nothing is known about the process underlying the tuning parameters and it prevents overestimation of physical understanding and thus prejudiced parameter choices. Especially in very complex models automatised methods may not only be necessary but also better than tuning on expert intuition only. The problem of intransparent interdependencies of parameters remains in automatically produced tuning parameter sets.
The explicit rules given in automatisation algorithm are apparently not sufficient to avoid the abuse of tuning but they are effective in constraining internal parameters according to restricted variables. If climate models were based on first principles[2] and observationally restricted parameters only, such algorithms would be not only effective in constraining but could also be prescribed to avoid the abuse but climate models are not like that. It is not possible to tune a model on the basis of first principles and laws of physics, as tuning is per se unphysical. Nevertheless, the processes simulated

[1]Tuning parameter values are not chosen randomly but since they are by definition (section 3.4) unconstrained by observation individual choices are necessary to be able to tune at all and it is technically impossible to sample huge ranges of all parameters.
[2]see ftnt. 1 in section 5.3.

by the model should not be unphysical, which means they should not show unplausible behaviour as, for example, heat fluxes that violate mass conservation, negative velocities, or atmospheric velocities faster then sound. Heat fluxes and similar processes in the atmosphere are good examples of processes not constrained by observation thus no meaningful flux behaviour can be included in algorithms, nevertheless such fluxes can be plausible or not according to the common sense of climate scientists. Therefore common sense rules could be meaningful even if they are not rules explicit as algorithms. With respect to common sense rules it could also be justified to not adhere to them but this justification is necessary as otherwise it is commonly assumed by the community that the rule is followed.
The best tuning results are probably gained if automatically tuned parameters are checked afterwards according to the crucial points depicted in the rules and subjective criteria gained from an understanding of physics and experience. Anyway the possibility of abandoning tuning parameter sets that are not consistent with such criteria should be given. Common sense rules as presented here are not new but they make implicit rules explicit. This is not superfluous because many tuning approaches do not violate the model physics or the common sense of climate modellers but some do and in most cases this would not have happened if the modeller had envisioned these common sense rules before the tuning. In agreement with Petoukov et al. (2000) it seems easier to realise explicit common sense rules than implicit ones.

From the rules discussed above a set of five rules (**R1**,**P1**-**P4**) can be extracted, which when adhered to will support good tuning practice. It is five instead of six as the second rule of the first set of rules (**R2**) presented in this section is very similar to the fourth rule of the second set (**P4**) whereas rule number one of the first (**R1**) set is an extension, as explained above. **O2**, which is the essence of **R2** and **P4** as explained above, is no contradiction to **O3** because the latter is used for parameters which are known in very defined ranges, whereas **O2** refers to parameters that are constrained to some extent but only by first principles or in very broad ranges. **O6** is in a certain interpretation quite similar to **O3** but is additionally mentioned in particular to point to the possibility that tuning can violate the model subsequent to a “correct” setup, especially by filter tuning. It is additionally important in the very first phase of setting up a modelling approach to ask the question of what variables to implement and which to parameterise.
But these five rules are not sufficient as tuning guidelines. In the first place I will prefix this set with a rule that is very common sense but nevertheless important to prevent circular reasoning. In the absence of a theory of the climate system a climate model could not gain any corroboration if the same data used for tuning is used for corroboration. Rule number one (**O1**) must thus be to use independent data sets for model tuning and model evaluation.[3] The crucial question is then of what counts as independent data. This is a question that cannot be solved within this rule-setting approach because I do

[3]In the context of model tuning this looks like a triviality but in the context of the philosophy of science and confirmation of theories it is far from trivial but rather a controversy,the implications of which for climate modelling rules will be briefly discussed below.

not know what kind of criteria to apply, which is one of the reasons for being able to present vague rules only. A rule climate scientists could certainly agree upon is not to use the same value from the same source twice.
Another aspect of tuning that is highly problematic is that which is explained in the context of the THC example in section 3.4. Climatic processes only existing in climate simulations due to tuning parameters may just be there for the wrong reasons. Thus it may be acceptable make claims about their behaviour under changed climatic conditions but it is not acceptable to make statements with respect to the stability of such processes under climate change (**O7**). This is a fact often disregarded in practice and not addressed in the relevant literature. Only Weaver et al. (2001) admitted that stability might be affected by tuning in such a way that it is not realistically represented. If such an analysis were made, causalities would be assumed to exist which are only pseudocausalities implemented by tuning. Hence the seventh rule is intended to counter the common error to assume that tuned cause and effect chains are real causalities. It is not entirely possible to prevent such misinterpretation but for cases given in the example sticking to this rule is possible if modellers make clear which processes exist due to tuning only. This rule can also be taken as a special case to the claim of independent data in the sense that pseudocausal relations are to be used only once, within the tuning process but not again in model interpretation.
The set of rules is still incomplete especially due to the numerous cases where it is impossible to tune parameters separately regarding the physical appropriateness. But as no reliable method to circumvent this problem is known so far the set will stay incomplete for now. Tuning as a crucial aspect of climate modelling has not been subject to much scientific study yet. It only came into broader focus of research lately with publications by Collins (2007) and Tebaldi and Knutti (2007) and the attention from the IPCC but did not became the focus of a broader discussion. In the long run this will happen and the set of rules presented here is expected to be expanded and advanced.
A state of the art set of common sense tuning rules is the following:

O1 *Data sets used for tuning and evaluating models must be independent.*

O2 *Known constraints on parameter ranges are not exceeded.*

O3 *Parameters which are known empirically or from theory must not be used for tuning.*

O4 *Wherever possible parameterisations should be tuned separately against observed data, not in the context of the whole model.*

O5 *Parameters must relate to physical processes, not to specific geographic regions.*

O6 *The number of tuning parameters must be much smaller than the degrees of freedom predicted by the model.*

O7 *Pseudocausal interactions implemented by tuning should not be analysed in terms of causality relations.*

Sticking to this set of rules is likely to not be sufficient to guarantee a scientifically good tuning practice but is necessary to avoid undeliberate manipulation of model physics by tuning. But the fact still remains that tuning is a very subjective method of parameterisation and model adjustment. Thus it is a kind of scientific art where it is difficult and not always recommended to apply rules. The connection to physical theories and observational data adds up to a great variety.
The rules above are phrased to account for crucial steps in the tuning process but their importance is given for the process of model building in general. Especially in the more abstract wording of the **O**-rules they could be taken as model building rules with the main advantage of highlighting the crucial steps and thus showing which assumptions in the modelling approach must be made more transparent in the future. A breaching of this rules must be exposed, explained, and justified.

7.2. Common sense rules

A set of rules as depicted above could assure a scientifically valid modelling practice if given for all crucial modelling steps and observed thoroughly. But even if this was a desirable piece of work to do it would be a lot of work and it would never be finished as adjustment of rules would be needed whenever new technical or scientific methods are applied. But looking at the tuning rules in detail they entail more comprehensive modelling advice that could be called common sense. These are common sense rules that are not only of value for tuning but should be followed throughout the whole modelling approach. It is thus possible to extract comprehensive rules for the entire modelling approach from this specified set of tuning rules.
If meaningful rules for climate modelling were to be revealed it would be better if they did not contradict specific rules, but they would not be derivable from each other in a logical sens. Both sets of rules should enable scientists to reach their modelling aims and thus support each other on a common sense level. As said above, for climate science especially specified rules for modelling steps would be helpful to avoid errors. This is also true for unspecified rules as they point at crucial steps, open to the choices of the modeller. If these choices are transparent they can be judged and discussed by peers, which in the best case allows for a specification of the rule.

The first rule given above (**O1**) says that data should be independent, that is a datum known from observation must not be used more than once within the modelling process. The data can be used to assess a parameter, or to tune the model, or to validate the model. This rule seems to be trivial for climate science but in times of scarce data it is sometimes hard to follow and in the philosophy of science this rule is a matter of broad controversy. If interpreted as a claim for the use of new data for hypothesis testing only, it is not at all clear that it should be defended. The problem then is the difference between *logical* confirmation and *historical* confirmation as Musgrave (1974) puts it, where the latter accepts only data that was not known, or not even observed in a very strong form of predictionism (Curd and Cover (1998)). Time dependent confirmation is nothing you can meaningful claim for a climate model and not only in this context is this historical perspective questionable. A lot of climate modelling approaches are built on the hope of providing scientific understanding to make political decisions now that relate to a "better" future. If climate models in their current application are fit for this job is a question beyond this analysis. Perhaps they would be more adequate given their limitations and possibilities if they were used for worst-case projections. However, climate modelling would be impossible without the assumption that the observations used to build the model may also be used to confirm it, because there is no comprehensive theory to base climate modelling upon. Data is of greater importance for climate modelling than for theory development as the data is not only the basis of theoretical ideas but part of the model as initial and boundary data and in tuned parameters. Thus no confirmation can come from the very same data but if it could also not come from known data no confirmation would be possible at all. This is along the lines of Worrall (1978) who stated with regard to theory confirmation the simple rule: "One can't use the same fact twice: once in the construction of a theory and then again in its support." is to apply. This rule seems to be plausible for theories but it is true for climate modelling, at least given the interpretation of the "same" data from above.

It is of course always possible to split the available data set and use one half for tuning and the other for validation but this minimises data dramatically and the data is not truly independent. This problem is addressed in chapter 6 where it is explained that such a practice affords the active forgetting of the data you use for validation, which seems to be a rather hard task. This is especially so as a datum must also not be used in two different modelling approaches if the models are part of an ensemble. The latter is not guaranteed but within one modelling approach scientists must necessarily stick to the rule: never use a datum twice. Otherwise the model makes false presumptions and reasoning is circular. The data are biased, due to coarse spatial and temporal resolution and other observational obstacles, even if this rule is kept is out of question but there is no problem of climate modelling that could be solved within a rule.

Following the tuning guidelines the next two rules are not so fundamental as there exist arguments for the use of "best" parameters even if they are known to be false. Parker (2010b) gives an example: The average speed of falling ice crystals is $1 \pm 0.5m/s$ but the model gives better results, compared to observed data, if the parameter identified with this speed is much higher. A deliberately wrongly chosen parameter is thus to compensate for other model deficiencies.

Given the analysis throughout this work this practice does not seem to be a good one, as it makes it impossible to detect cause and effect chains within the modelling approach and thus in the climate system.

Despite the arguments that wrong parameters might increase the predictability of a climate model there is in principle consensus among climate scientists to avoid right results for wrong reasons because this very detection of causal relationships, which is in any event difficult for climate modelling, is impossible with such wrong parameters.

Unrealistic assumptions are pervasive in climate modelling, it would be impossible without these assumptions therefore most modelling results would be arrived at for wrong reasons if this term included all simplifying assumptions. Unrealistic assumptions ac-

cording to Betz (2006) are false with respect to the accepted theory but simplifications are also unrealistic, although they are not false in a contradictory sense, rather merely limited. An example of an unrealistic assumptions that is a brute simplification is the ideal gas law as illustrated by Betz (2006) which includes two unrealistic assumptions. The ideal gas theory takes gas molecules as Newtonian mass points and assumes the absence of inter-molecular forces. These assumptions are false as gas molecules are spacious and attract each other but especially for huge volumes and less pressure the molecules are very small compared to the volume itself and they do not meet each other very often. Therefore this simplification is meaningful and, importantly, it is an unrealistic assumption whose implications are quite traceable as they rely on a theory "above" the kinetic gas theory, which is that they are made before computations and are physically plausible and belong to an accepted theory, Newton's Mechanics. In the discussion of confirmation or corroboration not only inductive support from successful predictions is discussed but also support from "above", as Curd and Cover (1998) terms it, by an accepted theory that frames, in a manner of speaking, the hypothesis in need of confirmation.
Unrealistic assumptions, like those discussed with respect to kinetic gas theory, are in accordance with first principles, which cannot be said for many parameters to compensate for various model failure, and if they follow first principles it is necessarily not detectable. Climate models producing right results for wrong reasons with the definition of wrong reasons from above could be much better for prediction making then models with no or few such compensation parameters, but useless for learning about climate system processes and for learning about modelling for the purpose of developing better models. Without model development the aim of making good predictions cannot be achieved thus right results for wrong reasons prevent scientists from reaching their main modelling goals.
To conform to this wish, to avoid right results for wrong reasons, two fundamental principles of modelling practice should be acted upon. These should use all available theoretical grounding of the modelling approach and not take tuned climate mechanisms as causalities. The former, in principle, means not to exceed known ranges of parameters and especially not to violate first principles, whereas the latter could be interpreted as a very special case of the first rule: never use a datum twice, to avoid using pseudocausalities for scientific analysis.

From seven rules of tuning three principles for good modelling practice may be derived, whereas the latter two are the two-fold axiom to avoid right results for wrong reasons..

1 Never use a datum twice.

2 Use all available theory.

3 Do not analyse pseudocausalities.

These three rules reveal the fundamentals of good scientific practice and not only valid climate modelling rules. In this respect it is an unsatisfactory result not to be able to extract specific modelling rules from a set of tuning rules. But starting with tuning

guidelines there are at least three hard rules to follow in climate model development. If we had instead tried to give a set of model testing rules and generalise it we would very likely have ended up with nothing, as the three guidelines given above are not as relevant for testing as for model building, although the first one is concerned with to model comparison as described above. The last one provides no guidance for model validation but rather for the interpretation of results.
The IPCC came up with a set of rules identifying criteria for what does not count as model failure. But a positive list, to my knowledge, does not exist. There are good reasons for the absence of such a list, because what criterion would be included? For all practical purposes the representation of the present climate is a good criterion but, thinking it through, one cannot even claim this for a climate model as explained in chapter 6, much less other quality factors mentioned in chapter 5 such as consistency. This is for the more philosophical reason that models could in principle predict the future exactly but fail to simulate present day climates and for the very practical reason that it is impossible to give the degree of precision needed to count as "passing the test". The same holds for the rule that climate prediction models must depict a global warming trend in order to be acceptable. For practical purposes present climate and future warming are good guidelines to assess model performance, yet they are not fundamentally essential as there is the highly unlikely possibility of a negative feedback to counteract global warming in the near future and models able to represent the future must not necessarily give a correct simulation of the present.
This lack of rules that stand a thorough scrutinisation is basically due to the amount of uncertainty in our knowledge of the climate system and in modelling approaches. Also the many different modelling approaches concerning complexity, physical basis, but also numerical representations prevent comprehensive criteria being met. The insufficient availability of observational data closes the list of obstacles on the way to comprehensive model testing criteria.

The validation of climate models is not the only aspect not captured by these rules. Probability is also hardly touched by them but is nevertheless important in every step of modelling. But while model testing cannot be captured easily by common sense rules, there are criteria to be fixed in the handling of probability.
Even more than the arguments for whether parameter values should be true, the assignment of probabilities to climate variables is a controversial matter within the community. Some scientists (Murphy et al. (2004), Tebaldi et al. (2005)) apply probability distributions to climate variables from ensemble results despite the problems discussed in chapter 6, while others (Stainforth et al. (2007)) argue against it. The American philosopher of science Wendy Parker (2010b) tries to show a way to assign probabilities without ignoring the limitations of statistical and ensemble methods. As probability distributions are representations of uncertainty she gives three criteria the distribution must meet in order to represent it meaningfully. These are *ownership*, *justification* and *robustness*. Ownership characterises the fact of probabilities in climate science being strongly subjective. The scientists should be willing to claim it as their own representation of uncertainty. They must furthermore be able to justify shape and range of

the probability distribution while showing that this distribution represents all available knowledge about the predicted variable. This claim to represent all available knowledge is central and a mapping of the second modelling principle from above. Additionally the distribution's shape should not depend strongly on highly controversial findings and it should not be expected to change dramatically if minor model development takes place. If these criteria are met probability density functions of climate variables could be communicated, according to Parker (2010b). But Parker also admits in the concluding remarks of the cited paper that in most circumstances "something less than a full probability density function" (Parker (2010b)) might be appropriate.
The conclusion to be drawn given the analysis in chapter 6 is that probability distributions are generally out of place for GCM ensemble results. Nevertheless the criteria given by Parker are extremely helpful throughout the modelling process. One word to express these requirements is transparency. Scientists must communicate the assumptions, uncertainties and subjective beliefs they build their model upon. Consequently another rule of climate modelling could demand transparency. However, the model building guidelines depicted above are in particular necessary to keep for the modelling approach to be transparent for other scientists. But additionally it is exactly these facts, which theory is applied and which model behaviour resulted from tuning which must be laid open.
In the context of probability assessment of climate variables Rougier (2007) also provides a set of questions to be answered before probability distributions are given. These five questions are about measurements, model input, model imperfections, model validation, and computation. The main question for all these topics is basically whether the quantification of errors and naming of uncertainties at the relevant modelling step are given. Probability distributions of climate variables are the ultimate goal of climate modelling but given the state of the art modelling abilities are an aim that is too ambitious. The criteria such distributions must meet are nevertheless important guidelines throughout the modelling process. Only if every step of modelling is justifiable, will reliable probability distributions be a reachable goal in the future. To be able to justify it the whole process of modelling from building to validation and interpretation must be transparent and available knowledge must be used.
The criteria probability distributions must met, according to Parker (2010b), and the questions Rougier (2007) asks do not evolve from individual ethics but from general scientific responsibility. Good scientific practice, as mentioned above, is basically to stick to the scientific method, which is a similar loose term but with an even longer history. In the twentieth century the scientific method was the subject of study of several controversial philosophers, as described in connection to selected parts of Popper's and Kuhn's work in chapter 4.
Apart from philosophical debate there are several efforts from research institutions to give practical guidelines for good scientific practice. The, latest and most influential, for the time being, results of discussing the scientific method in Germany are the proposals for safeguarding good scientific practice (Deutsche Forschungsgemeinschaft (DFG) (1998)) mentioned above. No matter how controversial this topic is in philosophy, the recommendations of the research foundation, which are reflected in every university's

research guidelines, can be subsumed in the following imperatives. They also reflect Popper's and Kuhn's ideas of rational discussion, discussed in chapters 3.7 and 4.1, which underlie their individual philosophies and are discussed throughout their work, e.g. Popper (1959), Kuhn (1996).

1. Document your research to guarantee the repeatability of experiments.
2. Permanently doubt your results.
3. Be honest in the handling of data and towards your colleagues.
4. Every author is responsible for the contents of his published work.

If rules of climate modelling are possible, sticking to them must guarantee good scientific practice according to these axioms of empirical science. Crucial axioms connected with climate modelling research are documentation and honesty, where the latter is not a problem in terms of honesty about purpose but is concerned with uncertainty and especially lack of transparency. The controversy concerning the assignment of probability distributions to climate variables expresses that this is indeed the crucial axiom. Honesty in dealing with uncertainty is also the axiom that allows solely the conclusion in handling of probability distributions given above.
In the tuning example it becomes apparent that documentation is also crucial. In tuning and parameterising complex models used by several scientists, especially scientists who use a model as a black box, documentation is essential. Otherwise scientists not involved in every step of the specific model development cannot decide what the model is forced to represent due to tuning or what is the result of physics. Everyone using a model to do scientific research must be able to judge the uncertainties connected to their results. Therefore conceptual documentation of model development should be given in addition to today's technical reports of climate models. Complementing technical reports, model descriptions could include something like a summary for data users.

7.2.1. Dealing with uncertainty in climate modelling

To summarise the needs for a good scientific practice in climate modelling it can be said that transparency is the most important demand. If decisions in the modelling process are untraceable for other scientists no critical discussion and scrutinisation by peers is possible, which is the essence of science. The climate science community has the great advantage that it has an organisation of these peers where scientists of almost every institution concerned with climate science contribute their work. This organisation is of course the Intergovernmental Panel on Climate Change, the IPCC. Major findings of climate science are presented in the assessment reports of the IPCC, the fifth of which will be released in 2013, which means planning and modelling starts now. Due to the very broad contribution to the reports the panel is highly accepted throughout the climate community. In the IPCC there exists a body much more influential than national organisations, for example the German Research Foundation, that could provide rules to guarantee good climate scientific practice. Actually it does provide such rules, which are basically contained in the “Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties” (Mastrandrea et al. (2010)). The aim of these rules is to urgently recommend scientists to consider uncertainties and to provide a calibrated language to do this consistently. Eleven points are made about the treatment of uncertainty and six grades of uncertainty for climate variables are identified. Generally scientists are advised to consider uncertainty independently for every finding and to assess conditional uncertainties for findings in cause and effect chains. They are also encouraged to describe findings they consider unlikely and to provide information on the whole ranges of key climate variables even if they consider the values on the outer ranges improbable. In treating uncertainty IPCC authors are furthermore to take the following points into account.

First of all it is important to plan a strategy to deal with uncertainty at an early stage of assessment. The scientists working together are to know early which of their views might differ so that a range of them must be describe. Their findings are to be judged as expertly if possible or if necessary in a formal expert elicitation process and the judgment must be explained. To avoid the convergence of intrinsically different views every scientist is advised to make his judgment before discussing the matter as a group.

In displaying their uncertainty assessment scientists must have a psychological understanding of probability statements. A 90% chance of survival is interpreted more positively than a 10% chance of dying, for example. The authors of the guidance furthermore encourage scientists to describe findings with compelling evidence without uncertainty qualifiers. The latter are to be reviewed carefully to avoid incomplete assessment of all sources of uncertainty.

To evaluate and communicate uncertainty scientists are recommended to use three different measures. The most basic concept of uncertainty assessment is to use *summary terms* (1) to describe the type, amount, quality and consistency of a finding, on the one hand, and to independently evaluate the degree of agreement concerning this variable or event on the other hand. Even for climate processes that are very poorly known summary terms such as limited, medium or robust validity and low, medium or high agreement can be used. Climate variables which are at least known with their sign can additionally be assessed using the second measure: confidence (2). The guidance note proposes five levels of confidence from very low to very high, whereas findings of low and very low confidence are to be expressed only in fields of major concern. It is important to underline that these levels of confidence are subjective measures, not to be mistaken for statistical confidence levels. For variables which can be identified with ranges of values, not only confidence levels are assignable but likelihoods (3). These likelihoods may be displayed using seven terms calibrated for probability ranges from virtually certain corresponding to a probability range from 99% to 100% to exceptionally unlikely with a probability of between 1% and 0%, whereas the boundaries of the corresponding probability ranges are fuzzy.

To help to assess variables they can be classified into six classes, from ambiguous variables (1) over those known by sign (2), magnitude (3), or range (4) to variables that can be given with likelihood or probability (5), resulting from ensembles, multiple obser-

vations, or expert judgment, and also to variables determined from statistical analyses with probability distributions (6).
The author team summarises their guidance note on consistent treatment of uncertainties with the advice to communicate uncertainty carefully and to provide traceable account of their findings and their evaluation.

This guidance note describes a process of good scientific practice strongly related to the recommendations of such practice discussed above. In providing a consistent treatment of uncertainty the essential problem of climate modelling is addressed, which is, as said before, that decisions in modelling must be traceable for scientists. In my opinion, climate modelling would gain a lot of confidence if communication of uncertainty was standard in publishing results. To do this in a traceable way a consistent language is needed, and this is indeed a significant achievement of the IPCC and the guidance note. If editors of relevant journals accepted only such publications that deal responsibly with uncertainty it would be possible to assess research results from other scientists correctly. Until now only the research contributing to IPCC reports implement such standards, others do not address uncertainty at all or using uncalibrated language. Not only within but especially in communication outside of the community it is necessary in order to avoid the overinterpretation of modelling results which will, if results are contradicted by the next model generation, erode the general credibility of climate science. In a very brief outlook below it will be discussed whether rules could be given if findings of climate modelling approaches are given as advice to the public, and how this could be done. This means discussing the question of whether there are rules possible for advising on action to be taken, or differently phrased, whether rules can prevent the erosion of public credibility.
Taking the guidance note of the IPCC as a true guide scientists would reach a high degree of transparency for their work and thus confidence in research results from within and outside the community. But one point is misplaced in this guide as it would lead to an inconsistent picture of climate research results. That is the recommendation to present findings with overwhelming evidence without uncertainty measures. If all other results are presented with such a measure an unmeasured result could stir up mistrust due to the belief that uncertainty is not measurable in this case. As the concept given in the guidance note allows the expression of a high degree of certainty, for example, a likelihood of 99% to 100%, it is to be used in every case to guarantee consistency. Additionally, low probabilities are extremely difficult for human beings to assess. An even better recommendation would therefore be to also give certain events that are without probabilities or with low probabilities but high potential impact, without probabilities but as possibilities. Furthermore, the recommendation to assign probability distributions to variables where the distributions result from formal expert elicitation seems to overstretch the expertise of experts, likelihood assessment would be more honest with respect to the precision of available knowledge.

In sum, the guidance note of the IPCC concerning the treatment of uncertainty touches upon the main points preventing good modelling practice so far. If uncertainty is a topic addressed early in the modelling process, and not only addressed but considered adequately throughout the process, climate modelling will be a branch of science providing reliable results. Climate models will never have a measurable quality due to the fact that they are a modelling approach rather than being a fixed entity, and due to the problems of climate modelling discussed in chapter II. But addressing these problems carefully and always being able to justify assumptions and results, makes climate models a very powerful tools in the process of understanding the climate system. In almost every case conditional uncertainty would be assigned to modelling results, for example actual model projections are conditional predictions. It is now important to describe the conditions in a transparent way. The three rules of model development point to crucial steps of uncertainty within the modelling process.

7.3. Conclusion and outlook

Climate models defined as a modelling approach suffer from uncertainty in crucial steps of a modelling process, it is thus inevitable to assess the uncertainty within the process and in published modelling results to guarantee valid modelling results.
In principle there are two types of uncertainty: fundamental uncertainty and model uncertainty. The former is due to the nonlinearity of the climate system and its complexity which limits human understanding and possibilities to observe the climate comprehensively. The fundamental uncertainty cannot generally be overcome but deficits in understanding the climate decrease as research continues and modelling improves. Model uncertainties are not of a comparably fundamental nature but will also not dissolve either. Furthermore in fields of high uncertainty there is not only evidential uncertainty concerning processes or variables but also uncertainty in agreement about this uncertainty between scientists.
A crucial manifestation of the uncertainty is the parameterisation of processes and subsystems within the modelling approach. In the need to parameterise unknown principles in the climate system and processes on smaller scales as the resolution of model, and unobserved variables, the main problems of climate modelling come to a head. Tuning parameters is thus a method to make modelling possible on the one hand but is on the other hand extremely sensitive to unintended violation of physics if uncertainties in parameter knowledge are inadequately respected. Only if there is agreement about the uncertainties and they are treated accordingly is it at all meaningful to assign probabilities to parameters and therewith to modelled climate variables. An increasing use of stochastic models could improve probabilistic constraining of parameters and only hence offer true probability density functions. Instead of suggesting a very high level of precision, uncertainty would better be considered if calibrated terms of agreement and evidence of uncertainty were used throughout a modelling approach instead of probability density functions.

To serve all the different purposes that climate models are good for, a whole spectrum of models exist representing a hierarchy of complexity and various levels of understanding, all of them with different advantages and shortcomings. Due to this fact, it is not possible to describe comprehensive modelling rules that exceed general rules of good scientific practice significantly. Yet some crucial steps within the modelling process could be improved in terms of transparency and reliability if common sense rules are used for the specific procedure. This might be, for example, the tuning of climate models for which a set of rules is given above and the assignment of probabilities to predicted climate variables.
The degree of uncertainty in climate modelling steps and approaches varies enormously from virtually certain knowledge concerning the physical basis and thermodynamic projections to high uncertainty in the prediction of regional precipitation changes or cloud physics. If these differences in the certainty of climatological findings were communicated adequately, confidence in modelling results could be much increased. It would avoid an excessively optimistic level of precision in the assignment of probabilities but would also most importantly prevent the suspicion that all modelling results suffer from outside the community. It is virtually certain that the planet warms and that this warming increases and that this is accompanied by rising sea level and an increase in extreme weather events. Our understanding of the climate system and the modelling abilities providing these facts are sound.
Not only the communication of climate research and the dialogue with the public and politicians could benefit from the transparent dealing with uncertainty, but also research within the community. The climate is such a complex system that it is impossible to capture all relevant subsystems in a single modelling approach. Therefore there are research groups holding more or less monopolies over comprehensive models for specific subsystems. If such monopolies are unavoidable it will at least be necessary to review their findings thoroughly which is only possible if assumptions and their uncertainties throughout the modelling process are made transparent. Uncertainties in agreement of uncertainties can only be treated adequately if assumptions are discussed and shortcomings recognised. Three very basic rules of model building could be depicted that relate to typical model shortcomings and could help to avoid modelling errors. Furthermore, and in particular, these rules could be used to enlarge transparency because they give, if adhered to, scientists a basis to show that and which scientifically valid approaches are taken.

The IPCC has tried and tries increasingly harder to make the formation of presented climate modelling results more transparent. Rules for specific modelling steps as presented here for tuning and in very vague words for general model building, as well as rules for dealing with uncertainty will further this process of increasing transparency. No climate skeptic ever measured the uncertainty of his claims or laid his sources open. As an outlook this fact begs the question of whether the set of rules especially for model development could be complemented by rules used as guidelines in model validation, dealing with probabilities, and maybe in instructing actions according to climate modelling results. Given the analysis throughout this thesis the latter seems especially impossible, basically due to the fact that climate models and political actions are on very different levels of understanding and motivation, but this is certainly not the only possible conclusion possible from an analysis of this topic. Rules for dealing with probabilities instead will likely not be very different from those dealing with uncertainty, supplemented by guidelines referring to low probability, high impact events and virtually certain facts.

Within this paper I find it very much beneficial for climate modelling to have a set of guidelines assuring scientifically valid climate models. In the literature this is a highly discussed topic although in the context of theory building, not model building. It is thus an apparently interesting question to discuss whether the consequences drawn from modelling analysis for the development of model building rules could be applied to the theory building debate in the philosophy of science.

Glossary

Anthropogenic climate change is the change of the climate state due to atmospheric radiative forcing from greenhouse gases emitted by fossil fuel combustion and changes in land use. Leading signal of the anthropogenic climate change is global warming, the increase of the global mean surface temperature, section 2.2.

Carbon dioxide, CO_2 is a trace gas comprising 0.0035% of the atmosphere and normally given in parts per million in an air volume of one liter (ppmv). CO_2 is radiatively active as a greenhouse gas as it reflects long wave heat radiation back to Earth. It is the most important greenhouse gas secondary to water vapour but stays in the atmosphere for centuries whereas water vapour survives only for days at maximum. Carbon dioxide is built in plants by photosynthesis and released when biomass and fossil fuels are burnt. The preindustrial CO_2 concentration was 280 ppmv whereas today it is more than 390 ppmv (/co2now.org (2011)/).

Climate is defined as the average weather over a certain time span, traditionally 30 years, with average being the statistical mean and variability of relevant variables, which are for example surface temperature and precipitation, section 2.1.

Climate change is a statistically significant change in the mean or variability of variables and thus a change of the state of the climate system. Climate changes can occur naturally through variations in forcing parameters or changes in internal processes or can be induced by anthropogenic changes in the composition of the atmosphere, section 2.1 and 2.2.

Climate model refers to the approach of modelling the climate system or aspects of it by use of mathematical models commonly implemented within a computational software. A climate model is thus a numerical representation of the components of the climate system that are known and understood in such a way that they can be described by physical equations and parameterisations. Climate models are tools to simulate properties and components of the climate system. There is a whole spectrum of climate models referring to different purposes of modelling and representing varying levels of understanding and computational cost, section 2.3.

Climate Sensitivity is defined as the global mean equilibrium temperature change following a doubling of atmospheric CO_2-concentration from preindustrial 280ppmv to 560ppmv. The likely value of climate sensitivity is $3 \pm 1°C$.

Climate system refers to the nonlinear physical system of the earth's climate. It is composed of many subsystems and their feedbacks and interactions. The most important subsystems are the atmosphere, the oceans, the cryosphere, the biosphere, and the land surface, section 2.1.

Emissions in this text refers to anthropogenic CO_2 equivalent emissions of greenhouse gases released by biomass and fossil fuel burning and increased due to land use changes, agriculture and deforestation.

Ensemble denotes a set of parallel model simulations to give an estimate of the spreading of climate model output variables, section 5.2.

Experiment in the modelling context is a modelling approach with fixed boundary and initial conditions but variable parameters, part II.

Feedbacks can either be positive or negative. In the former case feedbacks are interactions of climate system components that increase these interaction while negative feedbacks decrease the interaction process, section 2.1.2.

Forcing as radiative forcing refers to the rate of energy change at the top of the atmosphere per area of the earth as it is defined in W/m^2. Forcing loosely describes driving mechanisms of the climate system whereas external forcing are drivers outside the system such as solar radiation, volcanic eruptions, or anthropogenic changing of atmospheric greenhouse gas concentration. Internal forcing mechanisms are basically internal feedbacks, section 2.1.

GCM is the abbreviation for General Circulation Model which is the most complex type of climate models. Coupled Atmospheric-Ocean GCMs (**AOGCM**) are the state of the art models for climate predictions as they seek to include as many relevant climate processes as possible to get a comprehensive simulation of the climate system, section 2.4.

Global temperature is, if not specified otherwise, the mean of the global surface air temperature.

Greenhouse effect refers to the ability of greenhouse gases such as carbon dioxide, water vapour, methane, etc. to reflect long wave radiation, while being transparent for incoming short wave solar radiation, and thus warming the earth like a glass roof in a greenhouses does, section 2.1.

Impacts (of climate change) are all events, like floods, hurricanes, droughts, etc. occurring due to global warming. Besides events directly linkable to climate change there are also indirect impacts as for example migration from areas uninhabitable after extreme weather events, section 2.2.

Nonlinearity is a property of a system if there are no direct and simple chains of causes and effects to identify within the system, as small changes in the cause may lead to unproportionally large changes in the effect; section 2.1 and section 3.2.1.

Parameterisation refers to the method of representing processes that cannot be explicitly resolved at the spatial or temporal resolution of the climate model because the process is inadequately understood or exists on a sub-grid scale. A parameterisation is to find a relationships between model-resolved larger-scale processes and unresolved processes in the climate system, section 3.3.

Prediction is a term used in its common sense if not explicitely stated otherwise. It is principally impossible to make climate predictions with the same precision and low uncertainty as weather forcasts, section 3.8.1.

Probability is normally given as a probability density function which is a function that indicates the relative chances of occurrence of different outcomes of a predicted climate variable. Probability density functions integrate to unity and are the most precise measures to assign to climate variables to account for the uncertainty in model and prediction, chapter 6.

Projection refers to the predicted climate development in a climate model on basis of a given emission scenario. A projection is thus a potential climate evolution, section 3.8.

Reanalysis refers to the technique of preprocessing observed data of climate variables to make it comparable to model data and to account for lacking data and not directly measured variables. Reanalysis is done with state of the art general circulation models which are updated with observed data after every timestep of modelling, section 3.6.1.

Scenario refers to a simplified possible future evolution of the climate system in a climate model or based on a simulation with several assumptions.

Spectrum of climate models was formerly known as hierarchy of climate models and describes the range of climate models from simple energy-balance models to complex GCMs, section 2.4.

Theory is an ambiguous term in referring to scientific theories, as for example quantum mechanics, on the one hand and to theoretic assumptions not necessarily belonging to a scientific theory on the other hand. Since there is no complete scientific theory of the climate system the term theory basically refers to theoretic assumptions, an exception is section 2.3.1.

Tuning of a climate model refers to the adjustment of parameters to achieve agreement with observations. This means weakly restricted parameters are adjusted in such a way that parameters restricted by observations match the observations, section 3.4.

Uncertainty is an expression to account for insufficient knowledge concerning the range of variables and mechanisms and processes in the climate system and in the models. Uncertainty is omnipresent in climate modelling as it has several sources, section 3.1 and disputed ways to deal with, section 6.1.2 and 7.2.1.

Bibliography

Annan, J. D. and J. C. Hargreaves (2011). On the generation and interpretation of probabilistic estimates of climate sensitivity. *Climatic Change 104*, 423–436.

Batchelor, G. K. (1967). *An Introduction to Fluid Dynamics.* Cambridge University Press.

Bender, F. A. M. (2008). A note on the effect of GCM tuning on climate sensitivity. *Environemental Research Letters 3*, 1–6.

Betz, G. (2006). *Prediction or Prophecy? The boundaries of economic foreknowledge.* Gabler Verlag, Wiesbaden.

Callendar, G. S. (1938). The artificial production of carbon dioxide and its influence on temperature. *Quarterly Journal of the Royal Meteorological Society 65*, 223–237.

Carrier, M. (2006). *Wissenschaftstheorie zur EinfÃ¼hrung.* Junius Verlag GmbH, Hamburg.

Cartwright, N. (1983). *How the Laws of Physics Lie.* Oxford University Press.

Cartwright, N. (1999). *The Dappled world: A Study of the Boundaries of Science.* Cambridge University Press.

Chalmers, A. F. (1973). On learning from our mistakes. *British Journal for the Philosophy of Science 24*, 164–173.

Chalmers, A. F. (1999). *What is this thing called science?* Open University Press. (3.ed.).

Claussen, M., L. A. Mysak, A. J. Weaver, M. Crucifix, T. Fichefet, M.-F. Loutre, S. L. Weber, J. Alcamo, V. A. Alexeev, A. Berger, R. Calov, A. Ganopolski, H. Goosse, G. Lohmann, F. Lunkeleit, I. I. Mokhov, V. Petoukhov, P. Stone, and Z. Wang (2002). Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. *Climate Dynamics 18*, 579–586.

co2now.org (2011). The earth's home. Homepage showing the current value of carbon dioxide concentration in the atmosphere http://co2now.org/.

Collins, M. (2007). Ensembles and probabilities: a new era in the prediction of climate change. *Philosophical transactions of the Royal Society A 365*, 1957–1970.

Crutzen, P. J. (2002). Geology of mankind. *nature 415*, 23.

Curd, M. and J. A. Cover (Eds.) (1998). *Philosophy of science. The central issues.* W. W. Norton & Company. Commentary Chapter 4.

Deutsche Forschungsgemeinschaft (DFG) (1998). *Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission: Selbstkontrolle in der Wissenschaft; Denkschrift = Proposals for safeguarding good scientific practice.*, Weinheim. Deutsche Forschungsgemeinschaft (DFG): Wiley-VCH.

Eagle, A. (2005). Randomness is unpredictability. *British Journal for the Philosophy of Science 56*, 749–790.

Edwards, P. N. (2001). Representing the global atmosphere. In C. A. Miller and P. N. Edwards (Eds.), *Changing the Atmosphere.* The MIT press, Cambridge, Massachusetts.

Feyerabend, P. (2010). *Against method* (4 ed.). Verso, London, New York.

Gates, W. L. (1992). The atmospheric model intercomparison project. *Bulletin of the American Meteorological Society 80*, 29–55.

Ghil, M. (2001). Hilbert problems for the geosciences in the 21st century. *Nonlinear Processes in Geophysics 8*, 211–222.

Gramelsberger, G. (2006). Story telling with code - Archaeology of climate modelling. *TeamEthno-online 2*, 77–84.

Gramelsberger, G. (2010). *Computerexperimente.* transcript Verlag, Bielefeld.

Hack, J. J. (1992). Climate system simulation: basic numerical & computational concepts. In K. E. Trenberth (Ed.), *Climate System Modeling.* Cambridge university press.

Hacking, I. (1983). *Representing and intervening.* Cambridge University Press.

Hargreaves, J. C., J. D. Annan, N. R. Edwards, and R. Marsh (2004). An efficient climate forecasting method using an intermediate complexity Earth System Model and the ensemble Kalman filter. *Climate Dynamics 23*, 745–760.

Hume, D. (1993). *Eine Untersuchung über den menschlichen Verstand* (12 A ed.). Philosophische Bibliothek, Felix Meiner Verlag, Hamburg.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society 77*, 437–471.

Knutti, R., G. A. Meehl, M. R. Allen, and D. Stainforth (2006). Constraining climate sensitivity from the seasonal cycle in surface temperature. *Journal of Climate 19*, 4224–4233.

Kriegler, E., J. W. Hall, H. Held, R. Dawson, and H. J. Schellnhuber (2009). Imprecise probability assessment of tipping points in the climate system. *Proceedings of the National Academy of Science 106*, 5041–5046.

Kuhn, T. (1996). *The structure of scientific revolutions* (3 ed.). University of Chicago Press.

Kuhn, T. S. (1998). Logic of discovery or psychology research? In M. Curd and J. A. Cover (Eds.), *Philosophy of science. The central issues*, pp. 11–19. W. W. Norton & Company.

Lenton, T. M., H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber (2008a). Tipping elements in the earth's climate system. *Proceedings of the National Academy of Science 105*, 1786–1793.

Lenton, T. M., H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber (2008b). Tipping elements in the earth's climate system. *Proceedings of the National Academy of Science 105*, 1786–1793.

LeTreut, H., R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson, and M. Prather (2007). Historical overview of climate change. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Levermann, A., J. Schewe, V. Petoukov, and H. Held (2009). Basic mechanism for aprupt monsoon transition. *Proceedings of the National Academy of Science 106*, 20572–20577.

Lipton, P. (1998). Induction. In M. Curd and J. A. Cover (Eds.), *Philosophy of science. The central issues*, pp. 412–425. W. W. Norton & Company.

Lorenz, A., H. Held, E. Bauer, and T. Schneider von Deimling (2010). Constraining ocean diffusivity from the 8.2 ka event. *Climate Dynamics 34*, 719–734.

Ludwig, G. (1974). *Einführung in die Grundlagen der Theoretischen Physik*. Bertelsmann Universitätsverlag.

Lüthi, D., M. LeFloch, B. Bereiter, T. Blunier, J.-M. Barnola, U. Siegenthaler, D. Raynaud, J. Jouzel, H. Fischer, K. Kawamura, and T. F. Stocker (2008). High-resolution carbon dioxide concentration record 650,000-800,000 years before present. *Nature 453*, 379–382.

MacMynowski, D. G. and E. Tziperman (2010). Testing and improving enso models by process using transfer functions. *Geophysical Research Letters 37*, L19701.

Mangeney, A. and F. Califano (1998). The shallow ice approximation for anisotropic ice: formulation and limits. *Journal of Geophysical Research 103*, 619–705.

Mastrandrea, M. D., C. B. Field, T. F. Stocker, O. Edenhofer, K. L. Ebi, D. J. Frame, H. Held, E. Kriegler, K. J. Mach, P. R. Matschoss, G.-K. Plattner, G. W. Yohe, and F. W. Zwiers (2010). *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. Intergovernmental Panel on Climate Change (IPCC). Available online at http://www.ipcc.ch.

McGuffie, K. and A. Henderson-Sellers (2001). Forty years of numerical climate modelling. *International Journal of Climatology 21*, 1067–1109.

Meehl, G. A., G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer (2000). The coupled model intercomparison project. *Bulletin of the American Meteorological Society 81*, 313–318.

Meehl, G. A., T. F. Stocker, W. D. Collins, P. Friedlingstein, A. T. Gaye, J. M. Gregory, A. Kitoh, R. Knutti, J. M. Murphy, A. Noda, S. C. B. Raper, I. G. Watterson, A. J. Weaver, and Z.-C. Zhao (2007). Global climate projections. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Meinshausen, M., N. Meinshausen, W. Hare, S. C. B. Raper, K. Frieler, R. Knutti, D. J. Frame, and M. R. Allen (2009). Greenhouse-gas emission targets for limiting global warming to 2C. *Nature 458*, 1158–1162.

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

Menke, C. (2009). *Zum methodologischen Wert von Vorhersagen*. mentis Verlag, Paderborn.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature 430*, 768–772.

Musgrave, A. (1974). Logical versus historical theories of confirmation. *British Journal for the Philosophy of Science 25*, 1–23.

Nakicenovic, N. and R. Swart (Eds.) (2000). *Special Report on Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change.* Cambridge University Press.

O'Connor, D. J. (1957). Determinism and predictability. *British Journal for the Philosophy of Science 7*, 310–315.

Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. *Science 263*, 641–646.

Osborn, T. J., S. C. B. Raper, and K. R. Briffa (2006). Simulated climate change during the last 1,000 years: comparing the ECHO-G general circulation model with the MAGICC simple climate model. *Climate Dynamics 27*, 185–197.

Parker, W. S. (2010a). Predicting weather and climate: Uncertainty, ensembles and probability. *Studies in History and Philosophy of Modern Physics 41*, 263–272.

Parker, W. S. (2010b). Whose propbabilities? predicting climate change with ensembles of models. *Philosophy of Science 77*, 985–997.

Petoukov, V., A. Ganopolski, V. Brovkin, M. Claussen, A. Eliseev, C. Kubatzki, and S. Rahmstorf (2000). CLIMBER2 a climate system model of intermediate complexity. Part I: model description and performance for present climate. *Climate Dynamics 16*, 1–17.

Popper, K. R. (1959). *The Logic of Scientific discovery.* Cambridge University Press.

Popper, K. R. (1998). Science: Conjectures and refutations. In M. Curd and J. A. Cover (Eds.), *Philosophy of science. The central issues*, pp. 3–10. W. W. Norton & Company.

Rahmstorf, S. (1994). Rapid climate transitions in a coupled ocean-atmosphere model. *Nature 372*, 82–86.

Rahmstorf, S. (2007). A semi-empirical approach to projecting future sea-level rise. *Science 315*, 368–370.

Rahmstorf, S. and H. J. Schellnhuber (2006). *Der Klimawandel.* Verlag C.H. Beck.

Randall, D. A. (2000). *General Circulation Model Development.* Academic Press.

Randall, D. A., R. A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Ssrinivasan, R. J. Stouffer, A. Sumi, and K. E. Taylor (2007). Climate models and their evaluation. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Rougier, J. (2007). Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change 81*, 247–264.

Russell, B. (1999). *The problems of philosophy.* Dover Publications, inc.

Salmon, W. C. (1998). Rational prediction. In M. Curd and J. A. Cover (Eds.), *Philosophy of science. The central issues*, pp. 433–444. W. W. Norton & Company.

Schellnhuber, H. J. and J. Kropp (1998). Geocybernetics: Controlling a complex dynamical system under uncertainty. *Naturwissenschaften 85*, 411–425.

Schneider, S. H. (1992). Introduction to climate modeling. In K. E. Trenberth (Ed.), *Climate System Modeling.* Cambridge university press.

Schneider, S. H. and R. E. Dickinson (1974). Climate modeling. *Reviews of Geophysics and Space Physics 25*, 447–493.

Schneider von Deimling, T., H. Held, A. Ganopolski, and S. Rahmstorf (2006). Climate sensitivity estimated from ensemble simulations of glacial climate. *Climate Dynamics 27*, 149–163.

Shukla, J., T. DelSole, M. Fennessy, J. Kinter, and D. Paolino (2006). Climate model fidelity and projections of climate change. *Geophysical Research Letters 33*, L07702.

Smith, L. A. (2002). What might we learn from climate forecasts? *Proceedings of the National Academy of Science 99*, 2487–2492.

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

Stainforth, D. A., M. R. Allen, E. R. Tredger, and L. A. Smith (2007). Confidence, uncertainty and decision-support relevance in climate predictions. *Philosophical transactions of the Royal Society A 365*, 2145–2161.

Tebaldi, C. and R. Knutti (2007). The use of the multi-model ensemble in probailistic climate projections. *Philosophical transactions of the Royal Society A 365*, 2053–2075.

Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns (2005). Quantidying Uncertainty in Projections of Regional Climate Change: A Bayesian Approach to the Analysis of Multimodel Ensembles. *Journal of Climate 18*, 1524–1541.

Tetens, H. (2003). Naturalismus und die Grenzen unseres Wissens. In W. R. Köhler and H.-D. Mutschler (Eds.), *Ist der Geist berechenbar?* Wissenschaftliche Buchgesellschaft.

von Storch, H., E. Zorita, J. M. Jones, Y. Dimitriev, F. Gonzalez-Rouco, and S. F. B. Tett (2004). Reconstructing past climate from noisy data. *Science 306*, 679–682.

Watson, A. J. and J. E. Lovelock (1983). Biological homeostasis of the global environment: the parable of daisyworld. *Tellus 35 B*, 286–289.

Weaver, A. J., M. Eby, E. C. Wiebe, C. M. Bitz, P. B. Duffy, T. L. Ewen, A. F. Fanning, M. M. Holland, A. MacFadyen, H. D. Matthews, K. J. Meissner, O. Saenko, A. Schmittner, H. Wang, and M. Yoshimori (2001). The UVic Earth System Climate Model: Model Description, Climatology, and Applications to Past, Present and Future Climates. *Atmosphere-Ocean 39*, 361–428.

Worrall, J. (1978). The ways in which the methodology of scientific research programms improves on popper's methodology. In *Progress and rationality in science*, Dordrecht. Boston Studies in the Philosophy of Science 58; Synthese Library 125.

Zickfeld, K., A. Levermann, H. M. Granger, S. Rahmstorf, T. Kuhlbrodt, and D. W. Keith (2007). Expert judgements on the response of the atlantic meridional overturning circulation to climate change. *Climatic Change 82*, 235–265.

Oxford, den 23. Februar 2012

Versicherung die vorliegende Dissertation selbstständig verfasst zu haben

Hiermit versichere ich die vorliegende Dissertation, Modelling the earth's climate - an epistemic perspective, selbstständig verfasst zu haben und alle verwendeten Hilfsmittel und Hilfen in der Arbeit angegeben zu haben.

Friederike Otto