# Imprecise Probability Analysis for Integrated Assessment of Climate Change

DISSERTATION

zur Erlangung des akademischen Grades
“doctor rerum naturalium”
(Dr. rer. nat.)
in der Wissenschaftsdisziplin
“Theoretische Physik”

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

vorgelegt von
**Elmar Kriegler**

Potsdam, im Februar 2005

*When there is little information on which to base our conclusions, we cannot expect reasoning (no matter how clever or thorough) to reveal a most probable hypothesis or a uniquely reasonable course of action. There are limits to the power of reason.*

Peter Walley, 1991,
in *Statistical Reasoning with Imprecise Probabilities*, London, pg. 2

*In a predestinate world, decision would be illusory; in a world of perfect foreknowledge, empty; in a world without natural order, powerless. Our intuitive attitude to life implies non-illusory, non-empty, non-powerless decision... Since decision in this sense excludes both perfect foresight and anarchy in nature, it must be defined as choice in face of bounded uncertainty.*

George L.S. Shackle, 1961,
in *Decision, Order, and Time in Human Affairs*, Cambridge, pg. 43

# Abstract

We present an application of imprecise probability theory to the quantification of uncertainty in the integrated assessment of climate change. Our work is motivated by the fact that uncertainty about climate change is pervasive, and therefore requires a thorough treatment in the integrated assessment process. Classical probability theory faces some severe difficulties in this respect, since it cannot capture very poor states of information in a satisfactory manner. A more general framework is provided by imprecise probability theory, which offers a similarly firm evidential and behavioural foundation, while at the same time allowing to capture more diverse states of information. An imprecise probability describes the information in terms of lower and upper bounds on probability.

For the purpose of our imprecise probability analysis, we construct a diffusion ocean energy balance climate model that parameterises the global mean temperature response to secular trends in the radiative forcing in terms of climate sensitivity and effective vertical ocean heat diffusivity. We compare the model behaviour to the 20th century temperature record in order to derive a likelihood function for these two parameters and the forcing strength of anthropogenic sulphate aerosols. Results show a strong positive correlation between climate sensitivity and ocean heat diffusivity, and between climate sensitivity and absolute strength of the sulphate forcing. We have applied a series of statistical tests to the residual stochasticity, on the basis of which only values of climate sensitivity below $1.1\,\mathrm{K}$, and a sulphate aerosol cooling above $1.75\,\mathrm{W\,m^{-2}}$ in the year 1990 could be rejected at the 5% significance level.

We identify two suitable imprecise probability classes, probability boxes and $\varepsilon$-contamination models, for an efficient representation of the uncertainty about the climate model parameters and provide an algorithm to combine the information content of these two classes into a single imprecise probability representation described by a belief function. We construct a belief function for the prior parameter uncertainty from a set of probability estimates in the literature (climate sensitivity, sulphate aerosol cooling) and observational estimates of 20th century ocean heat uptake (ocean heat diffusivity). For the purpose of updating the prior with the likelihood function, we establish a methodological framework that allows us to perform the updating procedure efficiently on the entire event space and for two different updating rules: Dempster's rule of conditioning and the Generalised Bayes' rule. Dempster's rule yields a posterior belief function in good qualitative agreement with previous studies that tried to constrain climate sensitivity and sulphate aerosol cooling. It allocates small weight to high values of climate sensitivity (upper 95% quantile: $T_{2x} \approx 6.9\,\mathrm{K}$) and excludes low values of climate sensitivity $T_{2x} < 1.5\,\mathrm{K}$ with almost certainty. Moreover, the sulphate aerosol forcing in the year 1990 can be constrained to $[-1.53\,\mathrm{W\,m^{-2}}, -0.33\,\mathrm{W\,m^{-2}}]$ with 99% lower confidence. In contrast, we are not able to produce meaningful imprecise posterior probability bounds from the application of the Generalised Bayes' Rule. We can attribute this result mainly to our choice of representing the prior uncertainty by a belief function.

We project the Dempster-updated belief function for the climate model parameters onto estimates of future global mean temperature change under several emissions scenarios for the 21st century, and several long-term stabilisation policies. The upper end of the warming estimates is dominated by the possibility of high values of climate sensitivity, and exceeds corresponding estimates of the Intergovernmental Panel on Climate Change by more than 30%. At the low end, it is very unlikely that the warming in the 21st century will remain below 2 Kelvin in the absence of policy interventions. Within the limitations of our analysis we find that it requires a stringent stabilisation level of around 450 ppm $CO_2$ equivalent to obtain a non-negligible lower probability of limiting the warming to 2 Kelvin. We discuss several frameworks of decision-making under ambiguity and show that they can lead to a variety of, possibly imprecise, climate policy recommendations. We find, however, that poor states of information do not necessarily impede a useful policy advice.

We conclude that imprecise probabilities constitute indeed a promising candidate for the adequate treatment of uncertainty in the integrated assessment of climate change. We have constructed prior belief functions that allow much weaker assumptions on the prior state of information than a prior probability would require and, nevertheless, can be propagated through the entire assessment process. As a caveat, the updating issue needs further investigation. Belief functions constitute only a sensible choice for the prior uncertainty representation if more restrictive updating rules than the Generalised Bayes' Rule are available.

# Zusammenfassung

Diese Arbeit untersucht die Eignung der Theorie der unscharfen Wahrscheinlichkeiten für die Beschreibung der Unsicherheit in der integrierten Analyse des Klimawandels. Die wissenschaftliche Unsicherheit bezüglich vieler Aspekte des Klimawandels ist beträchtlich, so dass ihre angemessene Beschreibung von großer Wichtigkeit ist. Die klassische Wahrscheinlichkeitstheorie weist in diesem Zusammenhang einige Probleme auf, da sie Zustände sehr geringer Information nicht zufriedenstellend beschreiben kann. Die unscharfe Wahrscheinlichkeitstheorie bietet ein gleichermaßen fundiertes Theoriegebäude, welches jedoch eine größere Flexibilität bei der Beschreibung verschiedenartiger Informationszustände erlaubt. Unscharfe Wahrscheinlichkeiten erfassen solche Informationszustände durch die Spezifizierung von unteren und oberen Grenzen an zulässige Werte der Wahrscheinlichkeit.

Unsere Analyse des Klimawandels beruht auf einem Energiebilanzmodell mit diffusivem Ozean, welches die globale Temperaturantwort auf eine Änderung der Strahlungsbilanz in Abhängigkeit von zwei Parametern beschreibt: die Klimasensitivität, und die effektive vertikale Wärmediffusivität im Ozean. Wir vergleichen das Modellverhalten mit den Temperaturmessungen des 20. Jahrhunderts, um eine sogenannte Likelihood-Funktion für die Hypothesen zu diesen beiden Parametern sowie dem kühlenden Einfluss der Sulfataerosole zu ermitteln. Im Ergebnis zeigt sich eine stark positive Korrelation zwischen Klimasensitivität und Wärmediffusivität im Ozean, und Klimasensitivität und kühlendem Einfluss der Sulfataerosole. Wir haben eine Reihe von statistischen Tests angewandt, um den Raum möglicher Hypothesen einzuengen. Auf diese Weise lassen sich lediglich eine Strahlungswirkung der Sulfataerosole, die $-1.75\,\mathrm{W\,m^{-2}}$ im Jahr 1990 unterschritten hat, und eine Klimasensitivität unterhalb von $1.1\,\mathrm{K}$ mit einem Signifikanzniveau von 5% ausschließen.

Für die effiziente Beschreibung der Parameterunsicherheit ziehen wir zwei geeignete Modelltypen aus der unscharfen Wahrscheinlichkeitstheorie heran: Wahrscheinlichkeitsverteilungsbänder und sogenannte $\varepsilon$-Kontaminationsmodelle. Wir formulieren einen Algorithmus, der den Informationsgehalt beider Modelle vereint und durch eine sogenannte Belief-Funktion beschreibt. Mit Hilfe dieses Algorithmus konstruieren wir Belief-Funktionen für die A-priori-Parameterunsicherheit auf der Grundlage von divergierenden Wahrscheinlichkeitsschätzungen in der Literatur (Klimasensitivität, Strahlungswirkung des Sulfats) bzw. auf der Grundlage von auf Beobachtungsdaten beruhenden Schätzungen der Wärmeaufnahme des Ozeans in der zweiten Hälfte des 20. Jahrhunderts (Wärmediffusivität im Ozean). Wir leiten eine Methode her, um die A-priori-Belief-Funktion im Lichte der Likelihood-Funktion zu aktualisieren. Dabei ziehen wir zwei verschiedene Regeln zur Durchführung des Lernprozesses in Betracht: die Dempstersche Regel und die verallgemeinerte Bayessche Regel. Durch Anwendung der Dempsterschen Regel erhalten wir eine A-posteriori-Belief-Funktion, deren Informationsgehalt qualitativ mit den Ergebnissen bisheriger Studien übereinstimmt, die eine Einschränkung der Unsicherheit über die Klimasensitivität und die kühlende Wirkung der Sulfataerosole versucht haben. Einer hohen Klimasensitivität wird ein kleines, aber nicht verschwindendes Gewicht zugemessen (oberes 95% Quantil: $T_{2x} = 6.9\,\mathrm{K}$), während eine Klimasensitivität unterhalb von $1.5\,\mathrm{K}$ mit fast vollständiger Sicherheit ausgeschlossen werden kann. Außerdem kann die Strahlungswirkung der Sulfataerosole im Jahr 1990 auf den Bereich $[-1.53\,\mathrm{W\,m^{-2}}, -0.33\,\mathrm{W\,m^{-2}}]$ mit einem unterem Konfidenzgrad von 99% eingeengt werden. Im Gegensatz dazu finden wir bei Anwendung der verallgemeinerten Bayesschen Regel keine sinnvollen unteren und oberen Grenzen an die A-posteriori-Wahrscheinlichkeit. Wir haben festgestellt, dass dieses Resultat maßgeblich durch die Wahl einer Belief-Funktion zur Beschreibung der A-priori-Unsicherheit bedingt ist.

Die A-posteriori-Belief-Funktion für die Modellparameter, die wir aus der Anwendung der

Dempsterschen Regel erhalten haben, wird zur Abschätzung des zukünftigen Temperaturanstiegs eingesetzt. Wir betrachten verschiedene Emissionsszenarien für das 21. Jahrhundert sowie verschiedene Stabilisierungsziele für den Treibhausgasgehalt in der Atmosphäre. Der obere Bereich der geschätzten Erwärmung gegen Ende des 21. Jahrhunderts liegt ca. 30% oberhalb der Schätzung des Zwischenstaatlichen Ausschusses über Klimaänderungen (IPCC), was auf die Zulässigkeit sehr hoher Klimasensitivitäten im Rahmen der A-posteriori-Unsicherheit zurückzuführen ist. Bezüglich des unteren Bereichs läßt sich im Rahmen unserer Analyse sagen, dass eine Erwärmung unterhalb von 2 Kelvin im 21. Jahrhundert äußerst unwahrscheinlich ist. Darüberhinaus finden wir, dass sehr strikte Stabilisierungsziele im Bereich einer $CO_2$-Äquivalentkonzentration von ca. 450 ppm in der Atmosphäre notwendig sind, um nicht eine vernachlässigbar kleine untere Wahrscheinlichkeit für die Begrenzung der Erwärmung auf 2 Kelvin zu erhalten. Wir diskutieren verschiedene Kriterien für die Entscheidungsfindung unter unscharfer Wahrscheinlichkeit, und zeigen dass sie zu verschiedenen teilweise unscharfen Politikempfehlungen führen können. Nichtsdestotrotz stellen wir fest, dass eine klare Politikempfehlung auch bei Zuständen schwacher Information möglich sein kann.

Wir schließen, dass unscharfe Wahrscheinlichkeiten tatsächlich ein geeignetes Mittel zur Beschreibung der Unsicherheit in der integrierten Analyse des Klimawandels darstellen. Wir haben Algorithmen zur Generierung und Weiterverarbeitung von Belief-Funktionen etabliert, die eine deutlich größere A-priori-Unsicherheit beschreiben können, als durch eine A-priori-Wahrscheinlichkeit möglich wäre. Allerdings erfordert die Frage des Lernprozesses für unscharfe Wahrscheinlichkeiten eine weitergehende Untersuchung. Belief-Funktionen stellen nur dann eine vernünftige Wahl für die Beschreibung der A-priori-Unsicherheit dar, wenn striktere Regeln als die verallgemeinerte Bayessche Regel für den Lernprozess gerechtfertigt werden können.

# Preface

This work has been motivated by the objective to demonstrate the applicability of imprecise probability theory to the integrated assessment of climate change. To achieve this goal, we believe it to be important to demonstrate the applicability to each individual step in the assessment chain, i.e., 1. assessing an imprecise prior probability for a set of hypotheses about pertinent characteristics of the climate response (Chapter 3), 2. deriving a likelihood function for the set of hypotheses from a comparison with past and present climate observations (Chapter 2), 3. updating the imprecise prior probability with the likelihood information (Chapter 4), 4. projecting the resulting imprecise posterior probability onto estimates of future climate change (Chapter 5), and finally, 5. conducting a climate policy analysis under imprecise probability on the basis of these warming estimates (Chapter 5). An omission of any of these steps would have greatly diminished the value of this thesis. If we had omitted the first step, we would not have explicated how to generate the imprecise probability representation used later on. If we had omitted the second and third step, we would not have demonstrated how observational data can be included to constrain the uncertainty. If we had omitted the fourth and fifth step, we would not have delivered the results of our effort. Therefore, we believe that we had to include all of these steps into the thesis.

As a consequence, it was impossible to cover the wealth of material on a limited space of 100 pages. We want to assure the faculty that we have not breached this regulation light-heartedly. In contrast, we have taken every effort to limit the size of the presentation in the main body of the text as well as in the appendix. We have completely omitted all of our work that is not directly related to the application of imprecise probability theory to the integrated assessment of climate change (Kriegler and Bruckner, 2002, 2004; Edenhofer et al., 2005). Moreover, we have delegated our derivation of the climate model that is used throughout the analysis as well as a survey of imprecise probability theory that is necessary to understand our particular choice of imprecise probability analysis to the appendix. We have also taken care to avoid redundancies and excessive presentation. We have consulted a considerable number of proof-readers to double-check that this goal has been achieved in a satisfactory manner. As a result of our efforts, we could limit the main body of this thesis to 156 pages. Any further reduction would have detracted from the clarity of the presentation. Therefore, we are confident that the exceeding of the page limit will be met with acceptance.

# Contents

# Acknowledgment

This thesis could not have been accomplished without the support of many people. I am grateful to

- my supervisor, John Schellnhuber, who has always been a source of inspiration and encouragement to explore the untrodden paths,
- Ottmar Edenhofer and Hermann Held, who inspired my thinking, taught me a lot and in the end gave me the time and space to finish this thesis,
- Thomas Bruckner, who introduced me to the integrated assessment of climate change and since then has provided steady support and guidance for my work,
- Kirsten Zickfeld, Marian Leimbach and Ina Meyer, who have been great company since the days I joined the Potsdam Institute (PIK), where we have had numerous inspiring discussions,
- Gerhard Petschel-Held, who showed me a way to join the PIK,
- Jim Hall, who encouraged my interest in imprecise probability theory and provided guidance through its subtleties,
- Nico Bauer, Klaus Eisenack, Martin Füssel, Andrey Ganopolski, Thomas Kleinen, Malaak Kallache, Brigitte Knopf, Henning Rust, and Thomas Schneider-von Deimling for inspiring discussions and for proof-reading parts of this thesis,
- Kai Lessmann for urgently needed help with the layout of Matlab figures,
- all the other colleagues at PIK for providing a pleasant and inspiring atmosphere, in which interdisciplinary endeavours can flourish,
- and Natalia Andronova, Chris Forest, Reto Knutti, James Murphy, Thomas Schneider-von Deimling, Simon Gosling, and the UK MetOffice for the provision of data.

I gratefully acknowledge the support of the Deutsche Bundestiftung Umwelt (German Federal Foundation of the Environment), which enabled me to independently pursue my doctoral studies for three years.

Last but not least, I am deeply grateful to my family. My parents have supported and accompanied my years of study to this day. My almost two-year-old son Noah has (cleverly) not waited to explore the world until I finished this thesis. Finally, I want to express my deep gratitude to my wife Julia whose loving support and enduring patience has only enabled me to accomplish this work.

# Chapter 1

# Introduction

Anthropogenic environmental change is altering the flow of natural services as well as natural hazards to societies in many regions of the world. Meanwhile, the human impact on the environment continues to grow on a global scale, as the increasing land and water use and the rising concentrations of greenhouse gases in the atmosphere show. Decision makers are confronted with the difficult problem to address environmental change on the local, regional and global level and on time horizons of years to centuries. Among global environmental change, climate change is an issue of paramount concern. It has the potential to deliver severe consequences to societies around the globe ranging from extreme events like floods and droughts to abrupt changes of long-term climate regimes (Smith et al., 2001). At the same time, climate change constitutes an example of the intricate challenge posed to global environmental policy. As the spatial and temporal scales increase, the scientific uncertainty about the web of causal relationships grows, while the power of political institutions to take action becomes more and more limited.

In this situation, thorough assessments of the past and future impact of humankind on the climate and the past and future impact of the climate on humankind are needed to provide guidelines for action. Scientific investigations that are deliberately designed to support climate related economic, social and political decision making processes are called *integrated assessments of climate change* (e.g., Rotmans, 1990; Dowlatabadi and Morgan, 1995; Schellnhuber and Yohe, 1997). Ideally, such assessments have two key characteristics: their research objective is motivated by a real-world policy question; and their research findings provide a direct input to a *normative decision making process* with the objective to protect our well-being and environment in the face of climate change. As a minimum, integrated assessments will have to include models of the climate and socio-economic system and the interaction between them. Beyond this, there is the need for a consistent treatment of information about model structures, their validity in the light of historical data and their relevance with respect to the policy question under consideration. This can be a difficult task, since the assessment ranges across the border between the natural and social sciences.

A particularly important and still unresolved problem in the assessment process is the treatment of *epistemic uncertainty* like, e.g., lack of knowledge about causal relationships. Epistemic uncertainty is pervasive in the integrated assessment of climate change. How sensitive is the climate to an increase of the atmospheric carbon dioxide content? How strong is the cooling effect of anthropogenic aerosols? How will the ocean circulation react to a warming and freshening in the high latitudes? How large is the potential of renewable energy technologies to outcompete fossil fuels in the future? How fast and how resource intensive will the developing world develop in the 21st century?

We could continue the list of open questions for the remaining pages of the thesis. Integrated assessments of climate change will have to address many of them. This can be done by inserting the open questions as branching points in the analysis, by collecting the range of possible answers and by weighing their likelihood against each other as far as possible - in short, by measuring the uncertainty on the basis of the available evidence. However, the measure of uncertainty needs to be flexible enough to capture the widely different types of evidence from the natural sciences such as time series of stochastic climate variables, as well as from the social sciences such as stylised facts of economic growth. Moreover, the measure of uncertainty needs to be amenable to a decision analysis, since it is the very objective of an integrated assessment of climate change to support the real-word decision making process.

*Probability theory* provides a well-founded mathematical framework for quantifying epistemic uncertainty. The use of probability to deal with epistemic uncertainty rests in particular on a behavioural interpretation as *degree of belief* that is reflected in the disposition to act (de Finetti, 1974). In contrast, frequentist interpretations of probability are only of limited use if the uncertainty emerges from a lack of information rather than from stochasticity in observables. There exists also an impressive body of theoretical results and practical experience about the use of probabilities in decision making. Hence, probability theory exhibits several assets that are needed in the integrated assessment of climate change.

However, probability theory faces also severe difficulties in treating the diverse and often very poor information about the many aspects of climate change. In situations of near ignorance about a set of hypotheses, it is often recommended to assume that all possible hypotheses are equally probable. Obviously, such an expression of belief is something very different than ignorance. Since the equiprobability assumption generates the least informative probability model, additional information that is not backed by evidence has to be assumed deliberately in such situations. Indeed, probability assessments that do not rest firmly on an evidential basis can lead to controversial policy advice. Stakeholders may feel that the risk of recommended actions and the opportunity of discarded actions is undervalued. Others may reject the use of probabilities in light of the poor state of information. At the root of this controversy is a discomfort with the original probability assessment as accurate representation of the available information. In such situations probabilities cannot provide a satisfactory evidential foundation and, by this very reason, a convincing behavioural implication in a normative decision making context.

## 1.1 Motivation of the thesis

This thesis is motivated by the need for a framework of epistemic uncertainty with a clear evidential foundation and a clear behavioural implication that is well suited for the integrated assessment of climate change. Due to the poor state of information about many aspects of climate change, quantitative estimates of the associated uncertainty have been specified in the past mostly in terms of intervals or scenarios that delineate a range of plausible values for the quantity in question. The Third Assessment Report (TAR) of the Intergovernmental Panel on Climate Change (IPCC) serves as a good example to illustrate this point. However, such uncertainty estimates are not well-suited for a climate policy analysis. They are likely to produce large ranges of plausible policy outcomes, which will make it very difficult to distinguish the policies under consideration decisively.

Therefore, a debate ignited whether the IPCC should be more outspoken and provide probability assessments for, e.g., future emissions trajectories and climate sensitivity. While some argued that this is exactly what climate policy analysis needs (Schneider, 2001; Dessai and Hulme, 2003), others emphasised that this is not what a scientific inquiry can currently provide, in particular concerning socio-economic projections (Grübler and Nakicenovic, 2001). However, some climate scientists have been more optimistic about the prospect of probabilistic climate change forecasts (Allen and Ingram, 2002), and such forecasts have become indeed more widespread in the recent years (e.g., Stott and Kettleborough, 2002).

Probabilistic analyses of climate change are usually based on (i) probability estimates from experts or (ii) an ensemble of model runs with altered parameters or boundary conditions that are constrained by observations of past and present climate. In the former case, the subjectivity of the probability assessment is evident. Poor states of information usually disclose themselves by widely disagreeing expert opinions as was illustrated, e.g., by the expert elicitation of Morgan and Keith (1995) on climate sensitivity and other climate change related quantities. If the analyst insists on relying upon a single probability estimate for her assessment of climate change, she needs to choose some mechanism to aggregate the expert opinions (Genest and Zidek, 1986). Such a mechanism usually has to be based on a distribution of (probability) weights over the experts, whose existence and shape will be hard to justify (see, e.g., Keith 1996, Paté-Cornell 1997 for a discussion in the context of climate change). A more satisfactory solution would be to process all expert opinions on an equal footing. In this case, however, each emissions scenario or climate policy would have to be associated with an entire *set of probabilities*. Obviously, such a situation calls for a different policy analysis than can be provided by conventional decision frameworks such as expected utility theory (von Neumann and Morgenstern, 1944). Moreover, we have to ask the analyst if she really believes to have captured the state of information adequately by taking into account all elicited expert opinions. Given the large disagreement among the experts, she will hardly be able to dispute the fact that there can be many more probability estimates which are compatible with the state of information.

The example of subjective expert opinions highlights the difficulties that probabilistic uncertainty representations will have to face in the presence of poor information. The situation is similar, albeit less obvious for the second type of probabilistic climate change assessments that are based on an ensemble of model runs. In general, two ingredients are necessary to generate a probability distribution for the model ensemble from a comparison with observations. First, one need to specify a conditional probability that the model simulation reproduces the observations for each ensemble member. The collection of these conditional probabilities over the entire ensemble is called a *likelihood function*. Second, one needs to specify a probability distribution over the ensemble that expresses the degree of confidence in the individual ensemble members prior to comparing them with the data. The first, by no means trivial, requirement can be met by adapting the technique of optimal fingerprinting (Hasselmann, 1993; Allen and Tett, 1999). This method employs control run data from coupled atmosphere-ocean general circulation models (AOGCMs) to generate an estimate for the stochasticity in the climate, which is then fed into a linear regression of the ensemble of model response patterns against the observations. Hence, optimal fingerprinting allows to conduct a classical statistical analysis, on the basis of which ensemble members can be rejected at some confidence level. It can also be adapted to construct a likelihood function for the ensemble members (Berliner et al., 2000). However, in order to construct a probability distribution for the ensemble members, one needs to assume a second ingredient, the prior probability and update it with the likelihood function via Bayes' rule (Bayes, 1763). Hence, the subjective element of a probabilistic climate change assessment has re-entered the analysis in the shape of the prior probability.

Analysts that openly adopt a Bayesian framework usually consider several priors for generating probabilistic estimates of climate change or pertinent quantities like climate sensitivity (e.g., Forest et al., 2002). Again, their result is a set of probabilities with all its implications for climate policy analysis. Moreover, we can ask them the same question posed above as to why they did not consider all the uncountable prior probabilities that were compatible with the prior state of information. Other analysts that cling to the idea of objective probabilities usually make the implicit assumption that the prior probability is uniformly distributed, which enables them to identify the normalised likelihood function as posterior probability. This assumption is in accordance with both the concept of noninformative prior in objective Bayesianism (e.g., Kass and Wasserman, 1996) and the 'principle of indifference' in subjective Bayesianism, yet it can be challenged on both counts.

Turning first to the idea of an objective prior probability that ought to represent the 'absence of information' before any observation is received, we point to the fact that any probability distribution specifies a set of precise weights on the set of alternatives and, therefore, is markedly distinct from a state of complete ignorance. Due to this property, any attempt of identifying a specific probability distribution with complete ignorance will run into paradoxes. A nice example in the context of climate change is presented by

Frame et al. (2005). They consider two standard parameterisations for the sensitivity of the climate response to secular trends in the radiative forcing. This sensitivity can be either captured in terms of the climate feedback strength $\lambda$ or in terms of the climate sensitivity $T_{2x} \sim 1/\lambda$ to a doubling of the atmospheric $CO_2$ concentration relative to its preindustrial value (see Section 2.1.1). Due to the inverse proportional relationship between the two quantities, it is impossible to specify identical probabilities for both of them. Hence, if one identifies a single probability distribution with complete ignorance, then one can produce the paradoxical situation that there is complete ignorance about climate sensitivity, while at the same time there is some information about climate feedback strength and vice versa. More paradoxes can be constructed in a similar manner. For example, if you assumed complete ignorance about the atmospheric $CO_2$ concentration in the year 2100, you could not make the same assumption about the associated radiative forcing (logarithmic dependency). Or you might be completely ignorant about the effective ocean heat diffusivity, but not so about ocean heat uptake (square root dependency). Obviously, the attempt to associate a single probability with complete ignorance is doomed. On a much more informed basis than we can produce, Peter Walley states in his seminal book on the generalisation of probability theory (Walley, 1991, pg. 234):

> Our conclusion is that the quest for Bayesian noninformative priors is futile. So called 'noninformative priors' are not 'noninformative' (they have strong implications for behaviour),... The problem is not that Bayesians have yet to discover the 'truly' noninformative priors, but rather that no precise probability distribution can adequately represent ignorance.

It is fair to add that objective Bayesians have long retreated from the position that a noninformative prior represents a state of complete ignorance. It suffices, so they claim, that it 'lets the data speak', i.e., it does not bias the likelihood function in the process of constructing a posterior probability via Bayes' rule. However, we have to ask: 'Speak about what?' Ultimately, climate policy makers will be interested in a measure that combines climatic and socio-economic consequences of a particular policy under consideration. Hence, we would need to construct a likelihood function for the costs and benefits of climate policies. What kind of interdisciplinary data should be used, and what kind of interdisciplinary statistical method would be applicable to identify such a likelihood function? The author believes that such an attempt would be futile for a long time to come. The complexity of an integrated assessment of climate change will prevent the use of the objective Bayesian approach of simply 'letting the data speak'.

This brings us to 'the principle of indifference' which does not try to take recourse to the notion of 'objectivity', but instead, to the notion of 'rationality'. In the absence of any information, so it claims, it is rational to be indifferent between the alternatives. It can be shown that this behavioural disposition entails a uniform probability distribution on the set of alternatives. There have been more sophisticated approaches to motivate the existence of a probability distribution that describes the underlying subjective belief even in a state of complete ignorance. Conventional expected utility theories can be based on a set of 'rationality axioms' for individual preferences among a set of available actions, which guarantee the existence of such a subjective probability (Savage, 1954). However, this result hinges crucially on the restrictive nature of the rationality axioms, in particular on the comparability requirement for all acts and the so-called 'sure thing'-principle.

Ellsberg (1961) demonstrated in a famous experiment that these peculiar rationality requirements are at odds with real-life decision making. In an adjusted form the experiment consists in presenting an urn with altogether 90 red, yellow and black balls to a set of subjects. 30 balls are known to be red, while the other two colours are contained in unknown proportions. The subjects are offered a prize if they manage to obtain the colour of their choice with a random draw of a single ball from the urn. It turns out that the majority of subjects will pick the colour red, where they know the probability of winning the prize beforehand. They avoid the other two colours whose proportion of balls are surrounded by ambiguity. Obviously, the subjects refuse to make an equiproportion assumption for the black and yellow balls - in violation of the principle of indifference. Moreover, there exists no other pair of subjective probability weights for black and yellow that would make both colours less attractive than red. Hence, this instance of *ambiguity aversion* is incompatible with the exclusive use of subjective probabilities in the decision making process as motivated by Savage (1954).

Ellsberg's experiment shows that it can be misleading to discard the ambiguity in a decision situation by applying the principle of indifference or postulating a subjective probability in any other way. The author believes that this would be a particularly poor basis for conducting a viable climate policy analysis, because there exists indeed a large amount of ambiguity about the outcome of climate policies. It can be argued that concepts like the *precautionary principle* are motivated by the presence of ambiguity. The European Commission, for instance, states in a guideline on the applicability of the principle that a recourse to it presupposes *"a scientific evaluation of the risk which because of the insufficiency of the data, their inconclusive or imprecise nature, makes it impossible to determine with sufficient certainty the risk in question"* (European Commission, 2000, pg. 15). In the literature, various attempts have been made to formalise the precautionary principle by drawing on the difference between ambiguity and classical probabilistic uncertainty (e.g., Henry and Henry, 2002; Chevé and Congar, 2003).

Our discussion has highlighted the difficulties that an application of probability theory to the integrated assessment of climate change will inevitably encounter. The question naturally arises if the epistemic uncertainty about climate change can be processed by other means that avoid these difficulties at least to some extent? Since we have identified the major source for the inadequacy of the probability calculus in its inability to describe poor states of information, any other more promising framework would have to provide a satisfactory description of the state of complete ignorance. We have already seen above that a full treatment of subjectivity in probabilistic assessments gives easily rise to *imprecise probability* statements that encompass an entire set of plausible probability estimates. It is interesting to note that in response to the Ellsberg experiment axiomatic decision theories have been established that point in the same direction (Gilboa, 1987; Gilboa and Schmeidler, 1989; Schmeidler, 1989). These theories motivate the description of subjective beliefs in terms of convex sets of probabilities, which allows them to accommodate the phenomenon of ambiguity aversion as illustrated by Ellsberg (1961). At the same time, statisticians such as Peter Walley (1991) and philosophers like Isaac Levi (1980) have argued with fervor that a generalisation of probability to imprecise probability is necessary and sufficient for providing a satisfactory evidential and behavioural foundation for the treatment of epistemic uncertainty.

*Imprecise probabilities* come in several forms with different levels of generality (see Appendix C), but they all share the following characteristics: they yield a lower bound $\underline{P}(A)$ and upper bound $\overline{P}(A)$ on the probability that an event $A$ will occur. When lower and upper bound fall onto each other, the classical case of probability theory obtains. When the lower bound is strictly smaller than the upper bound, the state of information is said to be imprecise. This concept allows to capture the state of complete ignorance in a satisfactory manner (Walley, 1996a). It is modelled by the *vacuous lower and upper probability* $\underline{P}(A) = 0$ and $\overline{P}(A) = 1$ for all nonempty events $A$ that constitute a strict subset of the set $\Omega$ of all alternatives. It is easy to see that the vacuous probability bounds encompass the set of all probability distributions on $\Omega$. Since imprecise probability theory includes such a state of complete ignorance as well as the other extreme of complete information, it can model the entire continuum of information between the two extremes. It is important to note, however, that this does not allow to completely avoid subjective elements in an imprecise probability assessment along the lines of a Bayesian analysis. The reason is that the assumption of prior complete ignorance will also lead to posterior complete ignorance after updating the prior 'absence of information' with whatever observation is received (at least on the set of alternatives that cannot by excluded with certainty in light of the data) (Walley, 1991, Section 7.3.7). This indicates that likelihood information about some quantity in question can only convey a meaning if there is known more than nothing about this quantity beforehand. However, this condition is much weaker than requiring to know a precise prior probability beforehand. Imprecise probability theory allows to work with a set of (subjective) constraints that characterise a set of plausible probability distributions that are compatible with the prior state of information. These constraints can be very weak, and still might yield meaningful imprecise posterior probabilities in the updating process. Only the complete absence of any such constraints will render the inference process meaningless.

In summary, we believe that imprecise probability theory could be indeed a promising candidate for the consistent treatment of epistemic uncertainty in the integrated assessment of climate change. However, any promising uncertainty framework will be useless to the practitioner, if it cannot be applied to the complex problem of assessing climate change, e.g., due to a lack of mathematical tractability. Moreover, even if tractability can be established for some special type of imprecise probability model, the question remains whether the application will still show the particular strength of imprecise probability to describe the continuum of possible states of information. Ultimately, these questions can only be answered by giving it a try. This is the underlying motivation of the thesis presented here. We want to attempt a prototypical application of imprecise probability theory to the treatment of epistemic uncertainty in the integrated assessment of climate change. In doing so, we are seeking answers to the following three underlying questions:

1. Do mathematically tractable methods for processing imprecise probability information exist that suit the requirements of an integrated assessment of climate change?

2. Can the ability of imprecise probability to capture poor states of information be capitalised on for the quantification of uncertainties about climate change?

3. Can the impact of providing imprecise information to the climate policy process lead to a distinctly different decision support (e.g., more robust or more diffuse advice) than can be obtained from standard probability based decision analyses?

## 1.2 Outline of the thesis

To the best of our knowledge, we are undertaking the first attempt to investigate the potential of imprecise probability analysis in the integrated assessment of climate change. Therefore, we believe it to be important to cover the entire assessment chain from generating the imprecise prior probability representation for pertinent characteristics of the climate response, to updating it with likelihood information from a comparison with past and present climate observations, to projecting the imprecise posterior uncertainty onto estimates of future climate change, to finally conducting a climate policy analysis under imprecise probability. If imprecise probability theory is to be a viable uncertainty framework for the integrated assessment of climate change, it needs to be amenable to all these steps in a consistent manner.

The structure of the thesis reflects the stepwise proceeding through the assessment process. Chapter 2 contains the basis for our assessment of climate change by constructing the set of model hypothesis for the global mean climate response to human interference and by comparing it with the instrumental temperature record in the 20th century. For the former task, we need a simple physical-statistical climate model that parameterises the global mean temperature response to secular trends in the radiative forcing in an efficient manner. The model is briefly outlined in Section 2.1.1, while its construction is discussed in detail in Appendices A and B. To compare the model response with the instrumental temperature record in the 20th century, we need to specify an assumption about the radiative forcing during the industrial era (Section 2.2). In a next step, we define the likelihood that a particular model hypothesis can reproduce the observed 20th century temperature change on the basis of assumptions about the residual stochasticity between model simulation and data (Section 2.3 and 2.4). Finally, we calculate the resulting likelihood function on the uncertain model parameter space that is spanned by three key quantities influencing the climate response to secular trends in the radiative forcing: climate sensitivity, effective vertical ocean heat diffusivity, and the strength of the combined direct and indirect sulphate aerosol cooling effect (Section 2.5).

Chapter 3 turns to the problem of generating an imprecise prior probability for the model parameters. In doing so, we will use a variety of concepts from imprecise probability theory. To guide the reader through the rich hierarchy of these concepts we have provided a survey of imprecise probability theory in Appendix C. In Section 3.1 we identify a particular efficient representation that can be obtained for special cases of imprecise probability, and that has the potential to greatly improve mathematical tractability. We then discuss several imprecise probability models that might fit into this framework and show how they can be derived from expert opinions or diverse probability estimates in the literature (Section 3.2 and 3.3). In the remaining part of the chapter we construct prior belief functions, a special type of imprecise probability, for the individual model parameters (Section 3.4) and combine them to a joint belief function on the entire parameter space (Section 3.5).

Chapter 4 discusses the problem of updating the prior belief function with the likelihood function calculated in Chapter 2. After outlining the challenge of applying a generalised version of Bayes' rule in Section 4.1, we establish a methodology for the updating process in Section 4.2. The remaining section 4.3 discusses the imprecise posterior probability models that can be obtained under two different updating rules.

Chapter 5 tackles the remaining two steps of an integrated assessment of climate change. A methodology for projecting a posterior belief function for the climate model parameters onto future temperature change is discussed in Section 5.1. We apply the method under several scenarios for anthropogenic greenhouse gas emissions in the 21st century (Section 5.2), as well as several climate policies that aim to stabilise the atmospheric $CO_2$ concentration at various levels (Section 5.3). On the basis of these results, a stylised climate policy analysis under imprecise probability is conducted in Section 5.4.

Chapter 6 summarises our findings and arrives at a conclusion. To enhance readability, proofs of formal statements that are found throughout the text are delegated to Appendix D.

# Chapter 2

# An Energy Balance Climate Model and Its Likelihood

In order to estimate future climate change as well as to conduct a climate policy analysis, we need a model $M(\theta) : Q \rightarrow Y$ that calculates the mean response $y \in Y$ of the climate system to *secular trends* in the radiative forcing $q \in Q$ due to anthropogenic emissions of greenhouse gases. Naturally, such a model will contain uncertain parameters $\theta$ that influence the model projection of future climate change. For the purpose of demonstrating an imprecise probability analysis of this uncertainty, we have constructed a Diffusion Ocean Energy balance CLIMate model (DOECLIM). Energy balance models (EBMs) have been used in the past to derive the global or latitudinal surface temperature from the energy balance of the Earth system (Budyko, 1969; Sellers, 1969). They have found renewed interest for estimating the response of global surface temperature to a human-induced increase in atmospheric GHG concentrations (see, e.g., Hoffert et al., 1980; Harvey and Schneider, 1985; Schlesinger and Jiang, 1990; Wigley and Raper, 1992; Dutton, 1995; Murphy, 1995; Andronova and Schlesinger, 2000; Harvey and Huang, 2001), and are widely used in integrated assessments of climate impacts and policies (see, e.g., Nordhaus, 1994; Wigley et al., 1996; Harvey, 2004; Kriegler and Bruckner, 2004; Edenhofer et al., 2005). These energy balance models are mostly based on a linearised treatment of the climate response to small perturbations of the earth's energy balance. Section 2.1 discusses briefly the basic assumptions behind such linearised energy balance models, and then describes the structure of the model DOECLIM. At the end of the section, the model behaviour will be compared to simulations with more complex climate models. Due to the wealth of material that will be presented in the main body of this thesis, we have to delegate a detailed description of the model construction, the estimation of its parameters, and its numerical integration to Appendix A.

We then turn to the problem of estimating a *likelihood* that the model can reproduce the instrumental temperature record for a given choice of model parameters $\theta$. DOECLIM contains two free parameters: the *climate sensitivity* $T_{2x}$, and the *effective vertical diffusivity* $\kappa_v$ of heat in the ocean (see Section A.4). In addition, we include the cooling effect of anthropogenic sulphate aerosol into our set of model parameters, since it is the most important contributor to the uncertainty about the radiative forcing from anthropogenic agents (Ramaswamy, 2001). The cooling effect is parameterised by the combined *direct and indirect forcing strength of sulphate aerosols* $Q_{S90}$ in the year 1990. A similar choice of uncertain parameters has been made by recent studies that tried to quantify the uncertainty about climate sensitivity in terms of a probability measure (Andronova and Schlesinger, 2001; Knutti et al., 2002; Forest et al., 2002).

The model DOECLIM can be regarded as an *ensemble of hypotheses*

$$H = \{\, y := M(\theta) \,|\, \theta \in \Omega \} \tag{2.1}$$

about the climate response $y$ to radiative perturbations, where $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$ is the set of free model parameters, and $\Omega \subset \mathbb{R}^3$ the domain of possible assumptions about the parameters. Having made an actual observation $\hat{y}$, the likelihood function $\mathcal{L}(\theta; \hat{y})$ can be determined by comparing the hypotheses $M(\theta)$ with the observation. The likelihood is defined by the conditional probability of observing the event $\hat{y}$, if hypothesis $\theta \in \Omega$ is true. In case of an uncountable space of possible observations, as for climate data, the likelihood function equals the $\hat{y}$-value of the set of conditional probability densities $\rho(y|\theta)$ on $Y$, i.e. $\mathcal{L}(\theta; \hat{y}) := \rho(\hat{y}|\theta)$.

The likelihood function is an important concept in classical as well as Bayesian statistics. In classical statistics, it is frequently used to determine the parameter constellation $\theta^*$, which has the *maximum likelihood* to reproduce the observation. In Bayesian statistics, the likelihood function is used to update a *prior probability* to a *posterior probability* that includes the information gained from the observation $\hat{y}$. In the present analysis we will follow the Bayesian approach. However, since this requires us to calculate the likelihood function on the entire parameter space $\Omega$, we can conduct a classical statistical analysis en route.

Bayesian statistics constitutes a long-established paradigm (Berger, 1985) which has deep epistemological implications. The use of probabilistic information that is assembled *prior* to testing a hypothesis presupposes that such beliefs contribute eligible information for the evaluation of data from test experiments. In this sense, Bayesian statistics resembles an ‘expert mode’, where knowledge emerges from an interplay between beliefs and observations on hypotheses. In contrast, classical statistics tries to base knowledge purely on observations. We believe that the former model better suits the requirements of an integrated assessment of climate change, in which knowledge is collected for policy advice, and a set of repeated experiments is often not available. Bayesian analyses of climate change have been advocated by Hasselmann (1998), Leroy (1998) and Berliner et al. (2000). Compared to the large body of applications of techniques from classical statistics to climate research (von Storch and Zwiers, 1999), they have not yet received much attention. However, Bayesian applications are most likely to be found in climate change analyses that are directed to support some kind of climate policy decision making process.

The situation is illustrated by the efforts to detect and attribute an anthropogenic influence on 20th century climate. The state-of-the-art method consists in comparing observations with a superposition of (space-time) climate patterns from 3D climate models that were distilled from model runs with and without anthropogenic forcing agents (Hasselmann, 1993, 1997). The method has been called *optimal fingerprinting* because the patterns are chosen so that they maximise the signal-to-noise ratio for the detection of an anthropogenic influence. The estimation technique for the weighting factors of the patterns, and their variances, is firmly based on the classical technique of generalised linear regression (Allen and Tett, 1999; Allen and Stott, 2003). On this basis, several studies have shown that the hypothesis of a solely natural climate signal in the 20th century can be rejected with high confidence, and a large fraction of the warming is attributable to anthropogenic forcing agents (Hegerl et al., 1996; Tett et al., 1999; Stott, 2001; Tett, 2002). Due to the need to produce information for climate policy makers, Bayesian methods have been adopted that retain the classical core of optimal fingerprinting for estimating some kind of posterior probability for pertinent characteristics about climate change (e.g., Berliner et al., 2000).

In this analysis we will follow a different approach than is pursued in optimal fingerprinting and its Bayesian derivatives. The reliance on linear regression analysis is owed to the complex spatio-temporal structure of climate information that is not resolved by our energy balance model. Therefore, we have to restrict ourselves to the time series of global mean temperature in the 20th century. As an ancillary benefit from the reduction of complexity, we do not need to specify the ensemble of model hypotheses as a superposition of distinct patterns. Instead, we can use the continuous set of hypotheses $M(\theta)$ directly to derive a likelihood function $\mathcal{L}(\theta; \hat{y}) : \Omega \to \mathbb{R}^+$ from a comparison of model response with the instrumental temperature record. This approach has been used by Kaufmann and Stern (1997) and Smith et al. (2003) in the classical context of maximum likelihood estimation.

Section 2.2 lists the past contributions of anthropogenic and natural forcing agents that we include in our analysis. Section 2.3 discusses our assumptions about the *stochastic process* that underlies the residual between model output and data in the ideal case of a truthful representation of the climate response to the radiative forcing in the industrial era. Section 2.4 formalises these assumptions by defining a likelihood function for the climate model parameters in terms of the conditional probability density to observe the actual temperature record given some $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$. Section 2.5 discusses the likelihood function that emerges from the comparison of model simulations with the 20th century temperature record.

## 2.1 The diffusion ocean energy balance model DOECLIM

### 2.1.1 Energy balance of the earth system

In radiative equilibrium the infrared radiation from the earth's surface $F_S \uparrow$ is balanced by the portion of solar radiation $F_{\mathrm{Sol}} \downarrow$ that is absorbed by the earth system and the additional amount of energy $G$ that is distributed to the earth's surface due to the absorption of infrared radiation in the atmosphere. In a good approximation, the earth's surface radiates like a blackbody. According to the Stefan-Boltzmann law, it is $F_S \uparrow = \sigma T_S^4$ with $T_S$ the surface temperature and $\sigma = 5.67 \cdot 10^{-8}\ \mathrm{W\,m^{-2}\,K^{-4}}$ the Stefan-Boltzmann constant. Hence, the energy balance is described by

$$\sigma T_S^4 = F_{\mathrm{Sol}} \downarrow + G \,. \tag{2.2}$$

The presence of $G = \sigma T_S^4 - F_{\mathrm{Sol}} \downarrow > 0$ is called the *natural greenhouse effect*. Given a present day surface temperature $T_S \approx 287\,\mathrm{K}$ and an absorbed solar radiation of $F_{\mathrm{Sol}} \downarrow \approx 240\,\mathrm{W\,m^{-2}}$ on the annual average, we find $G \approx 150\,\mathrm{W\,m^{-2}}$. Thus, the greenhouse effect contributes considerably to the energy received by the earth's surface. A more detailed account of the greenhouse effect can be found in Appendix A.

A perturbation of the energy balance in Equation 2.2 induces a heat flux $\dot{H}(t)$ at the earth's surface. This leads to an adjustment of surface temperature $\Delta T_S(t)$, which corresponds to a change in emitted infrared radiation, i.e.,

$$\Delta F_S \uparrow (t) = \sigma\ (T_{S,\mathrm{eq}} + \Delta T_S(t))^4 - \sigma T_{S,\mathrm{eq}}^4 \approx 4\sigma\ (T_{S,\mathrm{eq}})^3\ \Delta T_S(t)\,,$$

The linear approximation is valid for 'small' adjustments $\Delta T_S$ relative to $T_{S,\mathrm{eq}} \approx 287\,\mathrm{K}$. In this approximation, the heat flux $\dot{H}(t)$ is constituted by the net energy difference between the perturbation $\Delta E(t) := \Delta G(t) + \Delta F_{\mathrm{Sol}} \downarrow (t)$ of the incident energy and a radiative damping that is proportional to $\Delta T_S(t)$,

$$\dot{H}(t) = \Delta E(t) - r\Delta T_S(t)\,, \qquad \text{with} \quad r := 4\sigma\ (T_{S,\mathrm{eq}})^3\ .$$

In energy balance studies of climate change it is usually assumed that a small perturbation $\Delta E(t)$ of the incident energy can be separated into two terms:

- a *radiative forcing* $Q(t)$ that captures alterations of the global energy balance due to changes in the solar constant, atmospheric content of aerosols (due to volcanos and industrial activity) and GHG concentrations (due to industrial activity). There exist several ways to define the radiative forcing. At current, the so-called adjusted radiative forcing is employed in most energy balance studies of climate change. It is defined as the net heat flux at the top of the atmosphere (TOA) after the stratosphere has returned to equilibrium, but before tropospheric and surface temperatures were allowed to adjust. This is a sensible concept, because the stratosphere will adjust independently and much faster (within months) than the troposphere-surface system. The different concepts of radiative forcing are discussed in greater detail in Appendix A.

- a *temperature feedback* on the incident energy at the surface that scales linear with the surface temperature change $\Delta T_S$. The nature of the temperature feedback will be discussed below.

Under this assumption, the heat flux at the earth's surface can be modelled by a linear ordinary differential equation,

$$\dot{H}(t) = Q(t) - \lambda\,\Delta T_S(t)\ , \qquad \text{with } \lambda = r\ (1 - f) \tag{2.3}$$

where the *climate feedback parameter* $\lambda$ summarizes the net strength of the combined radiative damping $r\,\Delta T_S(t)$ and temperature feedback $rf\,\Delta T_S(t)$ on the incident energy. The temperature feedback is positive if $f > 0$. In this case, the radiative damping is weakened by feedback loops that enhance the greenhouse effect for rising temperatures. It has been investigated extensively whether it is justified to separate radiative forcing and temperature feedback and to approximate the latter by a linear relationship with $\Delta T_S(t)$. Appendix A will give a brief account of these investigations.

The human-induced accumulation of greenhouse gases (GHG) and aerosols in the atmosphere has lead to a radiative forcing on the order of $1\,\mathrm{W\,m^{-2}}$ in the last decade, which is projected to increase to $4 - 8\,\mathrm{W\,m^{-2}}$ in 2100, if no mitigation policies are put in place to reduce anthropogenic GHG emissions (Cubasch and Meehl, 2001). The resulting change in the natural greenhouse effect $G$ is called the *anthropogenic* or *enhanced greenhouse effect*. A crucial role in the assessment of the enhanced greenhouse effect is played by the feedback of surface temperature on the radiation balance, whose strength is captured by the climate feedback parameter $\lambda$. In the following, we will briefly discuss the dominant feedback mechanisms, that contribute to the magnitude of $\lambda$. A more detailed discussion can be found in Harvey (2000, Chapter 9).

Rising temperatures increase the atmospheric water vapour content, which further enhances the absorption of infrared radiation from the surface. This *water vapour feedback* is of paramount importance, since it is the single largest positive feedback on global warming. Clouds influence the greenhouse effect in several ways. They absorb the infrared radiation from warmer lower atmospheric layers and re-emit it with the colder temperature that prevails at their location. This supports the greenhouse effect, since part of the surface radiation is 'shielded' from an escape to space. The impact is the larger the higher the clouds. On the other hand, clouds reflect the incoming solar radiation and, therefore, also exert a cooling influence. It depends on the height of the clouds whether the two opposing effects lead to a net cooling or warming. Hence, the direction of the *cloud feedback* on the incident energy $\Delta E(t)$ hinges crucially on the change in cloud distribution in a warmer and wetter atmosphere. Although this change is still very uncertain, most models project a positive cloud feedback.

The strength of the temperature feedback depends also on changes in the surface albedo. If the surface temperature rises, the surface albedo will decrease due to a reduction in snow cover and sea ice extent. This would lead to a positive *albedo feedback* on the energy $\Delta E(t)$ incident at the surface. However, concurrent changes in cloud cover, atmospheric water vapour and lapse rate might partially offset the feedback. While the strength of the net feedback on the planetary albedo remains uncertain, most models project it to be positive. Despite all the uncertainties, the overall temperature feedback on the energy $\Delta E(t)$ incident at the surface is unanimously estimated to be positive due to the strong positive water vapour feedback ($f > 0$).

It has become evident that the temperature feedback is a complex and highly uncertain quantity. In the context of the anthropogenic greenhouse effect, it is common practice to quantify the temperature feedback in terms of the *climate sensitivity* $T_{2x}$. Climate sensitivity is defined as the increase in global mean surface temperature that would result from a doubling of the atmospheric $CO_2$ concentration relative to its preindustrial value. Climate sensitivity is closely linked to the climate feedback parameter $\lambda$ in Equation (2.3). If the radiative forcing due to a doubling of the atmospheric $CO_2$ concentration is denoted by $Q_{2x}$, we find for equilibrium conditions ($\dot{H} = 0$)

$$T_{2x} = \frac{Q_{2x}}{\lambda} \, . \tag{2.4}$$

### 2.1.2 Model description

While Equation (2.3) provides a good approximation to the net global mean heat flux that warms or cools the surface, a detailed comparison with the historical temperature record will require a more resolved energy balance model. Since the transient climate response is dominated by the heat uptake of the world's ocean, a minimum requirement will be to consider the ocean and landmass separately. Many EBMs used for the integrated assessment of climate change go a step further, and also separate the southern and northern hemisphere. Two-hemisphere models are useful to capture the spatially inhomogeneous forcing from anthropogenic aerosols and tropospheric ozone, which are concentrated over the northern hemisphere landmass. However, they come at the cost of additional parameters to describe the inter-hemispheric heat fluxes.

The focus of our analysis lies on a prototypical assessment of model uncertainty with imprecise probabilities. We wish to keep the model as simple as possible in order to limit the number of uncertain parameters. For our purposes, the gain from a separation of hemispheres is outweighed by the associated costs of additional parameter uncertainty. Hence, we restrict ourselves to model the response of global mean land surface air temperature $T_L$, and global mean sea surface temperature $T_S$ to a radiative forcing $Q_L$ over land, and $Q_S$ over the ocean, respectively. As motivated in detail in Appendix A, the model equations of our diffusion ocean energy balance model DOECLIM are given by

$$C_{AL}\,\dot{T}_L = Q_L - \lambda_L\,T_L - \frac{k}{f_L}\,(T_L - b_{SI}\,T_S) \tag{2.5}$$

$$C_{AS}\,\dot{T}_S = Q_S - \lambda_S\,T_S - \frac{k}{1-f_L}\,(b_{SI}\,T_S - T_L) - F_O \, . \tag{2.6}$$

$T_S$ and $T_L$ describe *temperature anomalies* that are measured relative to the idealised radiative equilibrium before the perturbations $Q_L$ and $Q_S$ were applied. It is important to note that the heat fluxes on the left-hand and right-hand side of Equation (2.5) and (2.6) are modelled per unit area (see Table 2.1 for the units).

$C_{AL}$ denotes the combined heat capacity of upper land layer and atmosphere over land, while $C_{AS}$ captures the much larger heat capacity of the ocean mixed layer combined with the (neglectable) heat capacity of the atmosphere over the sea. The climate feedback parameters $\lambda_L$ and $\lambda_S$ can take on different values over land and over the ocean. Since we are considering two neighbouring boxes, we need to include the anomalous heat transfer between the boxes in the model. In an often used approximation, the heat flux anomaly is assumed to be proportional to the temperature anomaly gradient $T_L - b_{SI} T_S$ between the boxes. The scaling factor $b_{SI}$ for sea surface temperature (SST) derives from the fact that the heat exchange between ocean mixed layer and land is mediated through the atmosphere, and that the marine air warms faster than the sea surface due to the retreating sea ice cover (see Appendix A for details). Since we model the heat fluxes per unit area, we need to weigh the heat transfer between land and sea by the land fraction $f_L$ of the earth's surface.

The remaining term $F_O$ in Equation (2.6) denotes the heat flux from the mixed layer into the interior ocean. We have modeled the interior ocean as a 1-D pure diffusion ocean with a uniform vertical heat diffusivity $\kappa_v$ throughout the water column (see Appendix B for a motivation of this choice). Under this assumption, the anomalous heat transfer in the interior ocean can be modelled by the 1-D heat diffusion equation with Dirichlet boundary condition at the interface to the mixed layer (depth $z = 0$), i.e., $T_O(z = 0, t) = T_S(t)$), and Neumann boundary condition at the ocean bottom $z = z_B$, i.e., $\frac{\partial}{\partial z} T_O(z_B, t) = 0$ (see Appendices A and B for further details). We have solved this heat diffusion problem analytically in Appendix B. The analytical solution allows us to calculate the heat flux into the interior ocean

$$F_O(t) = -c_V\, \kappa_v\, \frac{\partial}{\partial z}\, T_O(z,t)|_{z=0}$$

as a function of the mixed layer temperature $T_S$, which then can be inserted into Equation (2.6) to close the energy balance model without having to model the entire ocean column explicitly. We find

$$F_O(t) = f_{SO}\, c_V \sqrt{\frac{\kappa_v}{\pi}} \int_0^t \frac{\dot{T}_S(t')}{\sqrt{t-t'}} \left(1 + 2 \sum_{n=1}^{+\infty} (-1)^n\, e^{-\frac{n^2 z_B^2}{\kappa_v (t-t')}}\right) dt' \ . \tag{2.7}$$

The infinite series converges very fast, so that it will be sufficient for our application to consider just the zeroth order term describing the behavior of an infinitely deep ocean and a few next order bottom correction terms (see Appendix B, in particular Fig. B.4). In Equation (2.7), the heat flux $F_O$ (in Watts per unit area) has been scaled by a parameter $f_{SO} < 1$, which captures the reduction of ocean area at the bottom of the mixed layer relative to the ocean surface area. This accounts for the portion of shallow coastal water, where the heat cannot diffuse into the interior ocean.

Due to the inclusion of an 1-D ocean model, the dynamics of the mixed layer temperature is described by an integro-differential equation. This complicates the model integration, but numerical solutions can still be obtained straightforwardly (see Appendix A.5). A similar approach for including a 1-D ocean model in an energy balance

| Topographical parameters | | Value |
|---|---|---|
| Fractional land area | $f_L$ | 0.29 |
| Ocean fractional area below 60 m depth | $f_{SO}$ | 0.95 |
| Interior ocean depth | $z_B$ | 4000 m |
| **Heat capacities** | | **Value** |
| Effective troposphere-land heat capacity | $C_{AL}$ | $0.52\,\mathrm{Wyr\,m^{-2}K^{-1}}$ |
| Effective troposphere-ocean mixed layer heat capacity | $C_{AS}$ | $7.80\,\mathrm{Wyr\,m^{-2}K^{-1}}$ |
| Specific heat capacity of seawater | $c_V$ | $0.13\,\mathrm{Wyr\,m^{-3}K^{-1}}$ |
| **Heat flux parameterisation** | | **Value** |
| Land-sea heat exchange coefficient | $k$ | $k = b_k - a_k\,\lambda_L$ |
| Heat exchange coefficient parameters | $a_k$ | 0.31 |
| | $b_k$ | $1.59\,\mathrm{W\,m^{-2}K^{-1}}$ |
| Marine surface air warming enhancement | $b_{SI}$ | 1.3 |
| **Climate feedback parameterisation** | | **Value** |
| Climate feedback parameter over land | $\lambda_L(T_{2x}, Q_{2x}, R_\lambda, b_{SI}, a_k, b_k, f_L)$ | |
| Climate feedback parameter over sea | $\lambda_S(T_{2x}, Q_{2x}, R_\lambda, b_{SI}, a_k, b_k, f_L)$ | |
| Climate sensitivity over land enhancement | $R_\lambda$ | 1.43 |
| Radiative forcing for atmospheric $CO_2$ doubling | $Q_{2x}$ | $3.7\ \mathrm{W\,m^{-2}}$ |
| **Free model parameters** | | **Unit** |
| Global climate sensitivity | $T_{2x}$ | K |
| Effective vertical heat diffusivity | $\kappa_v$ | $\mathrm{m^2\,s^{-1}}$ |

Table 2.1: The parameters of the 1-D diffusion ocean energy balance model (2.5)-(2.6). The climate feedback parameterisation is given by Equation (2.8).

model of anthropogenic climate change was pursued by Wigley and Schlesinger (1985). They approximated the analytical solution of the heat diffusion problem in the idealised case of an infinitely deep ocean ($z_B \to \infty$) to derive an ordinary differential equation for the mixed layer temperature $T_S$. The drawback of their approximation was that they had to introduce an artificial parameter which depended on the particular forcing scenario under consideration. In contrast, we consider the more realistic case of an ocean with finite depth, and do not approximate the resulting integro-differential equation by an ordinary differential equation, since we are seeking a general solution for arbitrary forcing trajectories.

In summary, our diffusion ocean energy balance climate model DOECLIM consists of the (integro)-differential equations (2.5) and (2.6), where the expression for the heat flux into the interior ocean is given by Equation (2.7). The model parameters of DOECLIM are listed in Table 2.1. We have detailed the estimation of the parameter values in Appendix A. The climate feedback parameters $\lambda_S$ and $\lambda_L$ are directly related to the climate sensitivity over land, $T_{L,2x}$, and over the sea, $T_{S,2x} = b_{SI}\,T_{S,2x}$. The functional dependence between these quantities can be identified from the equilibrium solution of Model (2.5)-(2.6), i.e., $\dot{T}_S = \dot{T}_L = 0$, and $F_O = 0$, for a doubling of the atmospheric $CO_2$ concentration relative to its preindustrial value. We find

$$\lambda_L = \frac{Q_{2x}}{T_{L,2x}} - \frac{k}{f_L}\,\frac{T_{L,2x} - b_{SI}\,T_{S,2x}}{T_{L,2x}}\,, \qquad \lambda_S = \frac{Q_{2x}}{T_{S,2x}} + \frac{k}{1-f_L}\,\frac{T_{L,2x} - b_{SI}\,T_{S,2x}}{T_{S,2x}}\,. \tag{2.8}$$

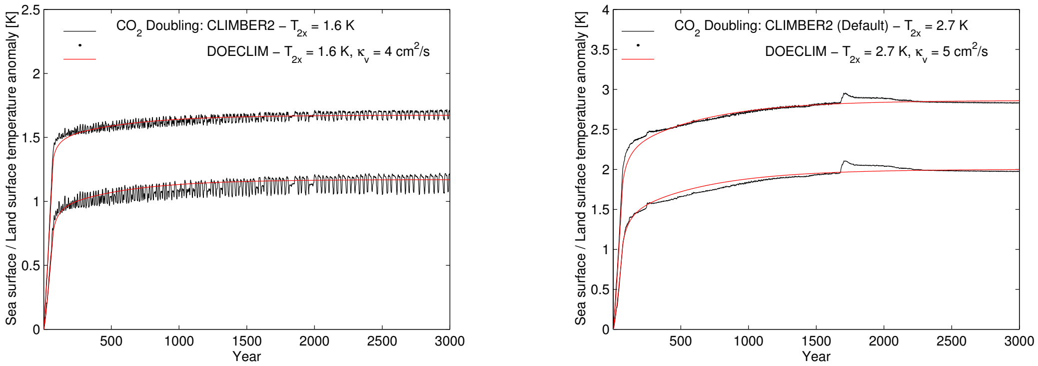
We have investigated the relationship between the climate sensitivity $T_{L,2x}$ of the surface air over land and the climate sensitivity $T_{S,2x}$ of the sea surface with data from an ensemble of CLIMBER-2 $CO_2$ doubling experiments (Schneider von Deimling et al., 2005), and found an almost strictly linear relationship $T_{L,2x} = R_\lambda \, T_{S,2x}$ with $R_\lambda = 1.43 \pm 0.02$ (see Appendix A for details). The larger climate sensitivity over land reflects the fact that temperatures at high latitudes will respond more strongly to an increase in atmospheric carbon dioxide concentrations than temperatures at low latitudes. Since the land fraction is highest at mid to high latitudes, this suggests a higher equilibrium temperature increase averaged over land than averaged over the sea (see also Murphy 1995, Raper and Cubasch 1996).

The proportionality between climate sensitivity over land and sea enables us to express these to parameters in terms of just one tunable quantity that is widely used to capture the uncertainty about the climate response to anthropogenic interference: the global climate sensitivity $T_{2x} = f_L \, T_{L,2x} + (1 - f_L) \, b_{SI} \, T_{S,2x}$ defined in Equation (2.4). As a consequence, we can reduce the number of free parameters in DOECLIM to two key factors influencing the climate response to radiative perturbations: climate sensitivity $T_{2x}$, and the effective vertical ocean heat diffusivity $\kappa_v$.

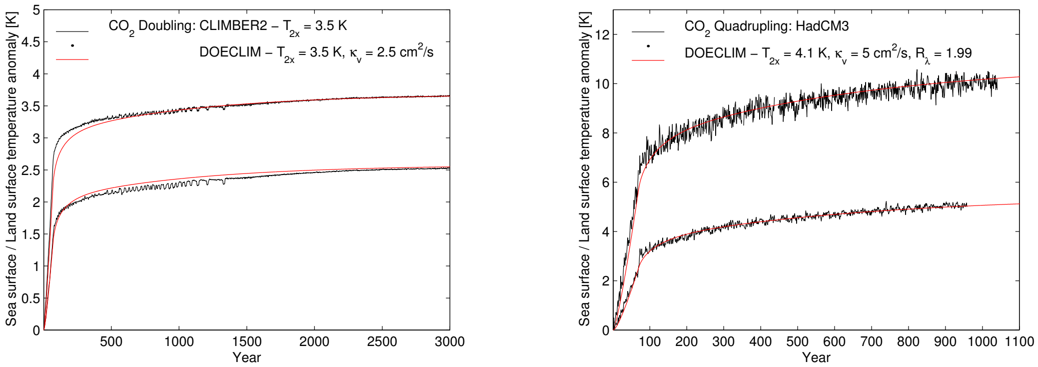
### 2.1.3 Comparison with coupled atmosphere-ocean climate models

The model DOECLIM can be seen as a hybrid physical-statistical model. The model is partly physical, because it has been derived from energy balance considerations and the fundamental heat diffusion equation. It is also partly statistical, because it does not derive pertinent climate system properties from physical principles, but either determines them from a comparison with (model) data (in the case of $C_{AL}, C_{AS}, k, b_{SI}, R_\lambda$), or includes them as a degree of freedom (in the case of $T_{2x}, \kappa_v$) to emulate observations and the behaviour of more complex models.

We have compared the response of DOECLIM to a doubling of the atmospheric $CO_2$ concentration from 280 ppm to 560 ppm in 70 years with a set of analogous experiments conducted with the climate model of intermediate complexity CLIMBER-2 for different model parameterizations (Schneider von Deimling et al., 2005). Fig. 2.1.a-c shows the sea surface temperature anomalies (lower curves) and land surface air temperature anomalies (upper curves) simulated by CLIMBER-2 for its standard version exhibiting a climate sensitivity of $T_{2x} = 2.68\,\mathrm{K}$ (Panel b), and two altered model parameterisations with a low climate sensitivity of $T_{2x} = 1.57\,\mathrm{K}$ (Panel a), and a high climate sensitivity of $T_{2x} = 3.45\,\mathrm{K}$ (Panel c), respectively. The temperature response simulated by DOECLIM was fitted to these CLIMBER-2 experiments by setting the climate sensitivity $T_{2x}$ to the CLIMBER-2 value, and adjusting the effective vertical diffusivity $\kappa_v$ of heat in the ocean in order to match the transient temperature response. Since the marine air warming enhancement, $b_{SI}$, and the enhancement of climate sensitivity over land, $R_\lambda$, were adopted from the ensemble of CLIMBER-2 experiments, it is no surprise that DOECLIM accurately reproduces the long term behaviour of sea surface and land surface air temperature simulated by CLIMBER-2. It can be seen,



a. CLIMBER-2: low climate sensitivity b. CLIMBER-2: standard parameter values



c. CLIMBER-2: high climate sensitivity d. HadCM3 - $CO_2$ quadrupling experiment

Figure 2.1: Comparison of $CO_2$ doubling experiments (doubling until year 70) with the diffusion ocean energy balance model integrated by numerical approximation (A.27) and the intermediate complexity climate model CLIMBER-2 (Panels a, b, c). Also shown is a comparison for a $CO_2$ quadrupling experiment (quadrupling until year 70) with the Hadley Center's general circulation model HadCM3 (Panel d). Lower curves depict sea surface temperature anomalies, upper curves land surface temperature anomalies. HadCM3 data were provided by courtesy of Simon Gosling from the UK MetOffice.

however, that DOECLIM is also able to reproduce the transient temperature behaviour in a satisfactory manner, if appropriate values for the vertical diffusivity $\kappa_v$ are chosen.

As an independent check, we have compared the behaviour of DOECLIM with a $CO_2$ quadrupling experiment conducted with the Hadley centre's current atmosphere-ocean general circulation model HadCM3 (Gordon et al., 2000). In this experiment, the equilibrated Hadley centre model was run for 1040 years with a quadrupling of the atmospheric $CO_2$ concentration from 280 ppm to 1120 ppm in the first 70 years. The sea surface and land surface air temperature response in HadCM3 are shown in Fig. 2.1.d. It can be seen that the land surface air warming exceeds the sea surface warming by a much bigger factor than in the CLIMBER-2 doubling experiments. Since the marine air warming enhancement is slightly smaller in HadCM3 than in CLIMBER-2 ($b_{SI} \approx 1.22$

instead of 1.3), the difference persists between land surface air and marine surface air. $CO_2$ doubling experiments with HadCM2, the predecessor of HadCM3, exhibited an enhancement of climate sensitivity over land that is very close to the CLIMBER-2 value of $R_\lambda = 1.43$ (Raper et al., 2001). The reason for the much larger land warming enhancement in HadCM3 remains unclear.

For the purpose of this analysis, it is sufficient to test DOECLIM against the HadCM3 data by adjusting the climate sensitivity over land enhancement to the larger value displayed by HadCM3. We find a value of $R_\lambda = 1.99$ by comparing the mean anomaly of land surface air temperature in the last 100 years of the HadCM3 experiment to the mean SST anomaly in the same period. The global climate sensitivity in DOECLIM was set to $T_{2x} = 4.1\,\mathrm{K}$, which is the effective climate sensitivity of HadCM3 at the end of the simulation period identified by Gregory et al. (2004). Due to the computational costs of the full coupled model, and the long time scales of the ocean to equilibrate, the true climate sensitivity of HadCM3 is unknown. Given these choices for $R_\lambda$ and $T_{2x}$, we have adjusted the vertical heat diffusivity $\kappa_v$ so that DOECLIM reproduces the temperature response of HadCM3 to a $CO_2$ quadrupling scenario. Fig. 2.1 shows that DOECLIM can approximate the HadCM3 behaviour convincingly.

We conclude that the diffusion ocean energy balance model DOECLIM, as captured by Equations (2.5) and (2.6) and numerically integrated by Equation (A.27), provides a flexible and reasonable description of the global mean temperature response to radiative perturbations on the order of magnitude of the anthropogenic interference with the climate system. In the following we will use the model DOECLIM to simulate the temperature behaviour in the industrial era, and compare it with the instrumental temperature record in the 20th century. This experiment requires knowledge of the historical radiative forcing trajectory, which will be established in the next section.

## 2.2 Radiative forcing during the industrial era

The radiative forcing trajectory during the industrial era is determined by the anthropogenic increase in atmospheric concentrations of forcing agents like well-mixed greenhouse gases, tropospheric ozone, and aerosols, as well as by natural contributions from solar and volcanic activity. There exists great uncertainty about the individual contribution of some forcing agents to a perturbation of the earth's radiation balance. It results from sparse data about the atmospheric load of these agents and from complex chemical as well as radiative interactions in the atmosphere. Since there are so many facets to the historical radiative forcing, an adequate consideration of the large body of literature would go beyond the scope of this thesis. Knowledge of the historical radiative forcing trajectory does not directly influence estimates of the climate response to scenarios of future anthropogenic interference. However, it will have an indirect influence since this knowledge is needed to constrain important properties of the climate system like climate sensitivity with observations of the 20th century climate. In this sense, the historical radiative forcing trajectory can be regarded as a *nuisance* *parameter* $\theta_N$. Nuisance parameters are characterised by the fact that they have no direct concern for the analysis, but uncertainty about them impedes the derivation of a likelihood function $\mathcal{L}(\theta;\hat{y})$ for a set of 'primary' parameters $\theta$ of interest.

By varying the nuisance parameters $\theta_N$ within a plausible range, one will receive a set of plausible likelihood functions $\mathcal{L}_{\theta_N}(\theta;\hat{y})$ for the 'primary' parameters $\theta$. In this way, the presence of nuisance parameters can lead naturally to *imprecise likelihood functions*, if a full treatment of the uncertainty about $\theta_N$ in probabilistic terms is to be avoided, e.g., due to a lack of prior information or due to computational limitations. Imprecise probability theory provides methods to deal with such objects, in particular in the context of the robust Bayesian approach (Berger, 1985). We regard the imprecise treatment of nuisance parameters as an important extension of our investigation into the usefulness of imprecise probability concepts for the integrated assessment of climate change. In this thesis, however, we make a first step and limit ourselves to imprecision in prior probabilities that are needed for a Bayesian analysis (see Chapter 3). Therefore, we will determine a precise likelihood function $\mathcal{L}_{\hat{\theta}_N}(\theta;\hat{y})$ by either making fixed assumptions about the historical forcing contribution of radiative agents, or by adjusting the remaining nuisance parameters $\theta_N$ in a way so to provide a best match between model output and historical temperature record (see Section 2.5).

### 2.2.1 Well-mixed greenhouse gases and tropospheric ozone

The well-mixed greenhouse gases include *carbon dioxide* ($CO_2$), *methane* ($CH_4$), *nitrous oxide* ($N_2O$), and *halocarbons* from industrial processes (HCFCs, CFCs, HFCs, PFCs, $SF_6$). Well-mixed greenhouse gases have long life times in the atmosphere so that the atmospheric circulation can distribute them fairly homogeneously around the globe. Their atmospheric concentration can be deduced from point measurements at locations that are distant from centres of agriculture, industrial production or fossil fuel combustion. The longest instrumental record of the atmospheric $CO_2$ concentration comes from air samples that have been collected since 1958 at the Mauna Loa Observatory, Hawaii (Keeling and Whorf, 2004). The instrumental record can be extended into the past for 600,000 years by analysing air bubbles that are trapped in ice cores from Greenland and Antarctica (e.g., EPICA community members, 2004). Hence, the historical concentrations of well-mixed greenhouse gases are well-known. The corresponding radiative forcings can be deduced from radiative transfer calculations, and simple parameterisations in terms of the concentrations are available in most cases (Ramaswamy, 2001, Table 6.2). In our analysis, we have used the radiative forcing estimates from the energy balance model MAGICC (Wigley and Raper, 1992; Raper et al., 1996), which was employed in the Third Assessment Report of the IPCC (Cubasch and Meehl, 2001), and includes a simple representation of the atmospheric chemistry (Wigley and Raper, 2002, Table 3). Figure 2.2.a summarises the radiative forcing contribution from well-mixed greenhouse gases. The halocarbon forcing of $0.34\,\mathrm{W\,m^{-2}}$ in the year 2000 (Ramaswamy, 2001, Table 6.11) has been reduced by the net negative forcing impact from the depletion of stratospheric ozone due to chlorfluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFC) controlled under the Montreal Protocol (Ramaswamy, 2001, Chapter 6.4).

*Tropospheric ozone* ($O_3$) differs from the well-mixed greenhouse gases in many respects. It is not directly emitted by humankind, but formed by photochemical reactions of precursor substances like carbon monoxide (CO), volatile organic compounds (VOC), and nitrogen oxides ($NO_x$) in the lower troposphere. These substances are emitted mainly from the transport sector and biomass burning, so that the increase in tropospheric ozone in the last century clearly is of anthropogenic origin. Unlike well mixed greenhouse gases, tropospheric ozone is depleted very fast in the lower troposphere by photochemical reactions with cleansing agents like hydroxyl (OH). Therefore, the tropospheric ozone load is distributed very inhomogenously around the globe. Concentrations downwind of polluted metropolitan areas can be two to three orders of magnitude higher than over remote tropical oceans. Due to the inhomogenous loading and the complex photochemistry for the creation and depletion of ozone, it is very difficult to estimate the ozone contribution to the radiative forcing on a global scale, let alone its development during the industrial era (Ramaswamy, 2001, Table 6.11). In this situation, we have adopted the scenario employed for the model MAGICC in the Third Assessment Report of the IPCC (Cubasch and Meehl, 2001).

Since the forcing is spatially inhomogenous, we need to specify it separately for the oceans and for the landmass. Since tropospheric ozone of anthropogenic origin is concentrated around metropolitan areas, we expect its abundance to be higher over land than over the oceans, and higher in the northern than in the southern hemisphere. We assume that the land-ocean ratio of the corresponding radiative forcing is of comparable magnitude as the NH:SH ratio, which has been estimated to range between 1.4 and 2 from studies with various atmospheric chemistry and radiative transfer models (Harvey, 2000, Table 7.3). Hence, we roughly estimate that the tropospheric ozone forcing over land exceeds the forcing over the oceans by a factor $R_{O_3} = 1.5$. Fig. 2.2.a depicts the forcing contribution of tropospheric ozone. It can be seen that tropospheric ozone is the third most important GHG of anthropogenic origin after carbon dioxide and methane.

### 2.2.2 Aerosols

Aerosols scatter and absorb solar and infrared radiation, thereby perturbing the energy budget of the earth system and exerting a direct radiative forcing. They also influence the radiation balance indirectly by altering cloud formation processes (1st indirect effect) and decreasing the precipitation efficiency of warm clouds (2nd indirect effect). Optically active aerosol particles of anthropogenic origin can be broadly categorised into *organic and black carbon* from the combustion of fossil fuels and biomass, and *sulphate aerosols* ($NH_4HSO_4$) produced by chemical reactions in the atmosphere with $SO_2$ from fossil fuel burning as precursor substance. Aerosols have very short lifetimes in the atmosphere ranging from hours to days for sulphate, and up to several weeks for carbonaceous aerosols of submicron size. Therefore, their forcing contribution is fairly localised around the centres of emissions. Due to the localised nature of the forcing and the variety of optical properties exhibited by aerosols, it is difficult to estimate their direct effect on the global radiation balance. The strength of their 1st and 2nd indirect forcing effect is even less understood (Penner, 2001).

Sulphate is believed to be the most important aerosol species of anthropogenic origin that had an impact on the global radiation balance in the past. It backscatters solar radiation from space and enhances cloud formation, so that its presence reduces the energy received by the earth's surface. Therefore, sulphate aerosols exert a negative radiative forcing. Given the large uncertainty about its direct and indirect radiative forcing, we will parameterize the range of assumptions about its forcing contribution, and include it in the calculation of the likelihood function. For this purpose, we adopt the simple parameterisation by Harvey et al. (1997) that has been used frequently in the literature:

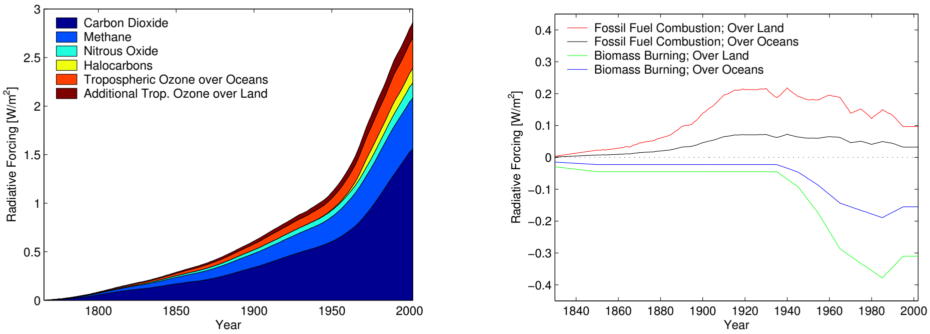
$$Q_{\mathrm{Sul}}(t) \;=\; Q_{\mathrm{S90}} \left( \lambda \frac{E_{\mathrm{SO_2}}(t)}{E_{\mathrm{SO_2}}(1990)} + (1-\lambda)\, \frac{\ln\,(1+\frac{E_{\mathrm{SO_2}}(t)}{E_{\mathrm{nat}}})}{\ln\,(1+\frac{E_{\mathrm{SO_2}}(1990)}{E_{\mathrm{Nat}}})} \right), \tag{2.9}$$

$$\text{with} \qquad \lambda = 0.33\,, \quad E_{\mathrm{SO_2}}(1990) = 71.5\,\mathrm{MtS}\,, \quad E_{\mathrm{Nat}} = 42\,\mathrm{MtS}\,.$$

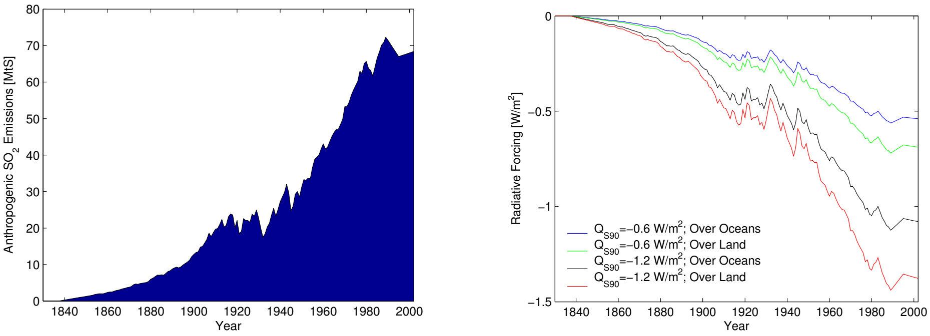
Obviously, the parameter $Q_{\mathrm{S90}}$ captures the total radiative forcing contribution of sulphate aerosols in the year 1990. The first term of the sum describes the direct forcing effect of sulphate aerosols. It can be assumed to vary linearly with the sulphate loading of the atmosphere and, hence close to linear with the total anthropogenic sulphur emissions $E_{SO_2}$. The indirect forcing is captured in the second term. Its absolute value is highly uncertain, but should saturate the more the natural sulphur emissions from volcanos and biogenic sources like marine plankton are exceeded by anthropogenic emissions. The saturation process is modelled by a logarithmic dependency on the ratio of anthropogenic and natural sulphur emissions. The latter are typically assumed to amount to an average value of $E_{\mathrm{Nat}} = 42$ Megatons of sulphur (MtS) (Harvey, 2000, Table 7.9).

The anthropogenic $\mathrm{SO_2}$ emissions are relatively well-known, since they are mostly related to fossil fuel burning. For this analysis we have adopted the emissions data for the period 1850-1990 from Lefohn et al. (1999, publicly available at www.asl-associates.com/sulfur.htm). We extended the data to earlier years by ramping $\mathrm{SO_2}$ from zero to 1.2 megatons sulphur in the period 1839-1849. For the time after 1990, we have adopted estimates from Smith et al. (2001) for the years 1995 and 2000, and interpolated the data to the remaining years in the period 1991-2002. The resulting trajectory of anthropogenic sulphur emissions in the industrial era is shown in Fig. 2.2.c. Lefohn et al. (1999) estimate a value of $E_{SO_2}(1990) = 71.5$ MtS for the year 1990.

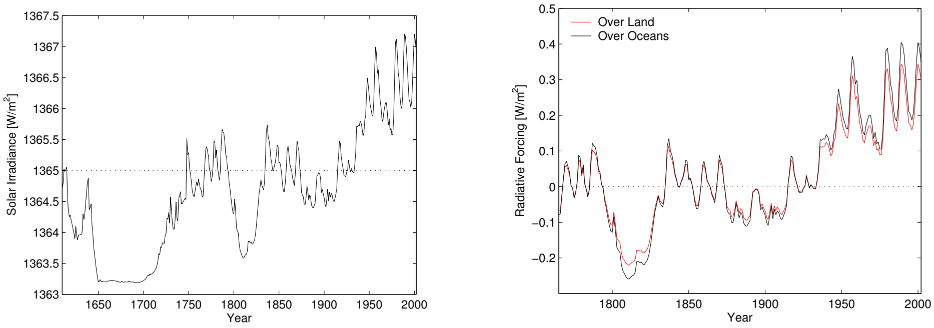
The combined direct and indirect radiative forcing from sulphate aerosols in the year 1990 and the relative strength of direct and indirect effect, captured by $\lambda \in (0,1)$, are uncertain. Estimates for the 1990s range between $-0.2\,\mathrm{W\,m^{-2}}$ and $-0.8\,\mathrm{W\,m^{-2}}$ for the direct effect, with a mean estimate of $-0.4\,\mathrm{W\,m^{-2}}$, and between $0\,\mathrm{W\,m^{-2}}$ and $-1.5\,\mathrm{W\,m^{-2}}$ for the 1st indirect effect (Ramaswamy, 2001, Table 6.11). The 2nd indirect



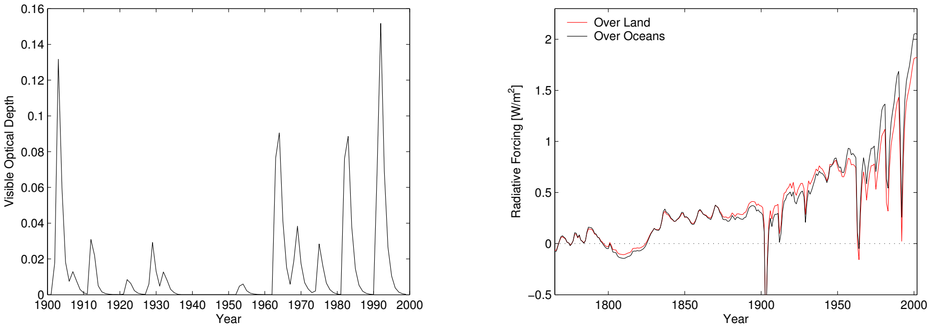
a. Anthropogenic greenhouse gases b. Carbonaceous aerosol forcing



c. Anthropogenic SO$_2$ emissions d. Sulphate aerosol forcing



e. Solar irradiance f. Solar forcing



g. Volcanic aerosols h. Total radiative forcing

Figure 2.2: Radiative forcing during the period 1750-2002 (Industrial era).

effect could not be quantified at all by Ramaswamy (2001) and, therefore, is ignored in most assessments of past climate change. Under this assumption, $Q_{\mathrm{S90}}$ might vary between $-0.2$ and $-2.3\,\mathrm{W\,m^{-2}}$, which is the range that will be considered for the calculation of the likelihood function. Although the ratio of direct and indirect effect is also highly uncertain, we choose a fixed value of $\lambda = 0.33$. This is value is obtained by assuming mean values of $-0.4\,\mathrm{W\,m^{-2}}$ for the direct effect and $-0.8\,\mathrm{W\,m^{-2}}$ for the indirect effect.

Since DOECLIM resolves ocean and landmass, we need to separate the radiative forcing contribution of anthropogenic sulphate aerosols over land and over sea. The bulk of the atmospheric $SO_2$ load is concentrated over Central Europe, the Eastern United States, and Southeast Asia, so that the direct forcing effect will be much larger over land. We make the rough estimate that it exceeds the direct sulphate forcing over the sea by a factor $R_{\mathrm{Sul}} \approx 2$ (Harvey, 2000, Figure 7.10, Plate 9). The situation is completely different for the indirect effect of sulphate aerosols. It is estimated to be of similar strength or even stronger over the ocean than over land. This is due to some combination of greater percentage cover of low-level stratus clouds and greater susceptibility of cloud optical properties to sulphate aerosols over the ocean (Harvey, 2000, Chapter 7.4). In this analysis, we assume equal strength of the indirect radiative forcing from sulphate aerosols over land and over the ocean. Fig. 2.2.d depicts the total radiative forcing from sulphate aerosols over land and ocean for two different assumptions about the combined direct and indirect global mean forcing strength $Q_{\mathrm{S90}}$ in the year 1990.

The radiative forcing from organic and black carbon of anthropogenic origin is even less understood than the forcing contribution of sulphate aerosols. Among others, this is due to the lack of reliable estimates about anthropogenic emissions and the resulting atmospheric loads of these substances. Moreover, black carbon is an absorber of solar radiation, which leads to an inhomogeneous heating of the atmospheric column influencing cloud cover (see Appendix A). If black carbon is mixed with sulphate aerosols, the resulting net radiative forcing depends heavily on the assumed mixing scheme (Harvey, 2000, Chapter 7.4). Some models even show that black carbon can neutralise the cooling effect of sulphate aerosols (Jacobson, 2001). Therefore, the radiative impact of black carbon might have been underestimated to date (e.g., Hansen et al., 2002). However, we do not wish to go into this field of active research here. Instead, we simply adopt the forcing scenarios for carbonaceous aerosols from Myhre et al. (2001). They are depicted in Fig. 2.2.b. Due to the absorptive properties of black carbon, the carbonaceous aerosols from fossil fuel combustion contribute a net positive forcing. The trajectory reflects the increase in fossil fuel use until 1950. Thereafter, the reduction of soot emissions from fossil power plants results in a steadily diminishing role of black carbon vs. organic carbon. The scenario for biomass burning aerosols reflects the rapid increase in deforestation activity after 1950 with its peak in the 80s.

Myhre et al. (2001) specified the historical forcing trajectory for the period 1850-1995. To extend the scenarios to the period 1750-2002, we have ramped them from

1826 (fossil fuel combustion) and 1790 (biomass burning), respectively, to their 1850 values, and kept the 1995 values constant thereafter (a total of $0.05\,\mathrm{W\,m^{-2}}$ for black and organic carbon from fossil fuel combustion, and $-0.20\,\mathrm{W\,m^{-2}}$ for biomass burning aerosols). The latter value agrees with the mean estimate of Ramaswamy (2001) for the forcing from biomass burning aerosols around the year 2000 (uncertainty $-0.07\,\mathrm{W\,m^{-2}}$ to $-0.6\,\mathrm{W\,m^{-2}}$). Concerning carbonaceous aerosols from fossil fuel combustion, Ramaswamy (2001) considered a net negative as well as net positive forcing plausible ($+0.1\,\mathrm{W\,m^{-2}}$ to $+0.4\,\mathrm{W\,m^{-2}}$ for black carbon, and $-0.3\,\mathrm{W\,m^{-2}}$ to $0\,\mathrm{W\,m^{-2}}$ for organic carbon). Fossil fuel aerosols are highly concentrated over land (Ramaswamy, 2001, Figure 6.7f). We roughly estimate that their radiative forcing over land exceeds the forcing over the oceans by a factor of $R_{\mathrm{FF}} \approx 3$ (Harvey, 2000, Figure 7.10). The situation is slightly different for biomass burning aerosols. While the plume from burning biomass in the Amazonian area is concentrated over land, the African and Indonesian plumes spread over land as well as over the oceans (Ramaswamy, 2001, Figure 6.7e). Therefore, we roughly estimate a somewhat lower factor of $R_{\mathrm{BB}} \approx 2$ by which the forcing from biomass burning aerosols over land exceeds the forcing over the oceans.

### 2.2.3 Solar and volcanic forcing

Changes in *solar activity* and *volcanic eruptions* constitute the major sources of natural perturbations of the earth's energy balance on secular time scales. The sun fluctuates between states of low and high magnetic activity with a period of approximately 11 years. During periods of high activity, the irradiance is $0.5\,\mathrm{W\,m^{-2}}$ to $1\,\mathrm{W\,m^{-2}}$ larger than during periods of low activity. Satellite measurements have shown that the solar irradiance fluctuates currently between $1366\,\mathrm{W\,m^{-2}}$ and $1367\,\mathrm{W\,m^{-2}}$ (Fröhlich, 2000). Long-term trends of solar irradiance on secular time scales are more speculative. They have been invoked on the basis of evidence from solar-like stars, and variations in the interplanetary magnetic field. This evidence suggested a long-term increase of solar insolation from the Maunder Minimum to present day conditions by approx. $3\,\mathrm{W\,m^{-2}}$, which has been included in most reconstructions of solar irradiance since the 17/18th century (e.g., Hoyt and Schatten, 1993; Lean et al., 1995; Fligge and Solanki, 2000; Lean, 2000). In recent years, however, the existence of a long-term trend in solar irradiance has been questioned. New data from solar-like stars have not supported the old findings, and the increase in interplanetary magnetic activity does not seem to have influenced solar irradiance significantly (Lean et al., 2002; Foukal et al., 2004). Therefore, it is likely that the available reconstructions exaggerate the increase in solar forcing during the 20th century. In the absence of corrected estimates, however, we employ the widely used reconstruction by Lean et al. (1995) in our analysis that was updated to 1997 with the results of Fröhlich and Lean (1998) (data publicly available at www.sparc.sunysb.edu/html/clim_forc.html). In the light of the new developments, this can be seen as a very optimistic assumption about the influence of the sun on the climate. The time series of solar irradiance since 1600 is shown in Fig. 2.2.e. We have extrapolated the solar cycle for the period 1998-2002 and decreased all irradiance values by $1\,\mathrm{W\,m^{-2}}$ to adjust them to the absolute values presented by Lean (2000). The solar forcing in the industrial era can be directly calculated from knowledge of the solar irradiance $S$. If the average holocene irradiance is set to $S_0 = 1365\,\mathrm{W\,m^{-2}}$, the solar forcing $\Delta\,Q_L$ over land and $\Delta\,Q_O$ over the oceans, respectively, is given by

$$\Delta\,Q_L(t) = \frac{\overline{DQ_L}}{4}\,(S(t) - S_0)\,(1 - \overline{\alpha_L})\;, \qquad \Delta\,Q_O(t) = \frac{\overline{DQ_O}}{4}\,(S(t) - S_0)\,(1 - \overline{\alpha_O})\;,$$

where $\overline{DQ_L} = 0.96$ and $\overline{DQ_O} = 1.02$ are the fractions of global mean solar insolation received by land and oceans, respectively, and $\overline{\alpha_L} = 0.35$ and $\overline{\alpha_O} = 0.28$ is the albedo over land and oceans, respectively (see Section A.4.1). The resulting solar forcing scenarios are shown in Fig. 2.2.f.

Volcanic eruptions can influence the climate by injecting large amounts of chemical active gases and aerosol particles into the stratosphere (Robock, 2000). In several weeks, a volcanic cloud forms by $SO_2$ conversion to sulphate aerosols, which is advected rapidly around the globe. The stratospheric sulphate cloud alters the radiation balance mainly by backscattering solar radiation to space, thus exerting a strong cooling effect. Since the sulphate aerosols remain 1-3 years in the stratosphere, the radiative forcing after large volcano eruptions has the form of a distinct negative spike of several years. The 20th century has experienced five major eruptions that injected a large amount of particles into the stratosphere: two early eruptions (Santa Maria, Guatemala, 1902; Novarupta, Alaska, 1912), followed by a period of relative calm, and three eruptions interspersed in the second half of the century (Agung, Bali in Indonesia, 1963; El Chichón, Mexiko, 1982; Pinatubo, Phillipines, 1991). Since the climate forcing from these eruptions was very strong, with typical estimates for, e.g., Pinatubo around $3\,\mathrm{W\,m^{-2}}$ (Hansen et al., 2002), they are imprinted in the instrumental temperature record. Therefore, it is necessary to account for these events in our analysis. We use the reconstruction of volcanic aerosols by Ammann et al. (2003) (data publicly available at www.ngdc.noaa.gov/paleo/pubs/ammann2003/ammann2003.html). Fig. 2.2.g shows the aerosol optical depth in the mid-visible wave length for the period 1900-1999[1]. The major eruptions in the 20th century clearly stand out.

The spatial distribution of aerosol optical depth will depend on the location of eruption. Ammann et al. (2003) provide the zonal mean optical depth on a 2.8° latitudinal grid, which does not suffice to separate the data for land and oceans. In this situation, we will use the global mean aerosol optical depth for both land and ocean. This assumption will not be completely unreasonable, since volcanoes are situated mostly on the rims of continental plates, and the volcanic cloud circles the globe rapidly.

Aerosol optical depth $\tau(\lambda)$ is specified as the fraction of light with wavelength $\lambda$ that is scattered when passing through the aerosol layer. It provides an excellent proxy for the radiative forcing $Q_{\mathrm{vol}}$ that is exerted by the volcanic cloud. In a good approximation, $Q_{\mathrm{vol}} = \beta_{\mathrm{Vol}}\,\tau$ (e.g., Andronova et al., 1999). We do not use direct

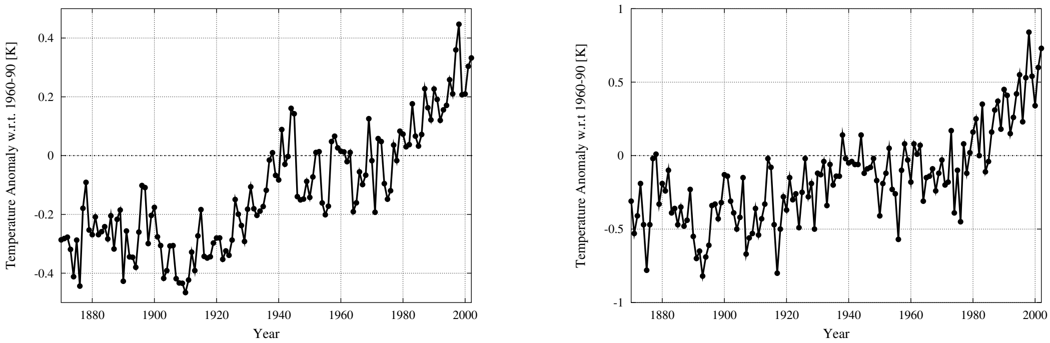
---

[1] Ammann et al. (2003) reconstruct aerosol optical depth for the period 1890-99. We have omitted their data for 1890-1899 to avoid a sharp cut-off in 1890 in the aftermath of preceding eruptions.

estimates of $Q_{\mathrm{vol}}$ for two reasons. First, there exists considerable uncertainty about the radiative forcing from volcanic aerosol after stratospheric adjustments that adds to the uncertainty about aerosol optical depth $\tau$. For the well-monitored Pinatubo eruption, for instance, estimates range from $Q_{\mathrm{vol}} = -15\,\mathrm{W\,m^{-2}}\,\tau$ (Ramachandran et al., 2000), and $Q_{\mathrm{vol}} = -21\,\mathrm{W\,m^{-2}}\,\tau$ (Hansen et al., 2002), to $Q_{\mathrm{vol}} = -25.4\,\mathrm{W\,m^{-2}}\,\tau$ (Andronova et al., 1999). Second, we want to employ a simple energy balance model which only resolves ocean and landmass. Since the volcanic forcing is strongly localised horizontally over the earth's surface, and vertically in the atmospheric column, our model might produce an inadequate response to a global mean forcing that was generated from a 3-D general circulation model. There are indications that highly aggregated energy balance models might overestimate the temperature response to global mean volcanic forcing (Andronova and Schlesinger, 2000). Hansen et al. (1997) have shown that the sensitivity of surface temperature to a forcing in the stratosphere is lower than it is to a rather uniform heating of the troposphere induced by an increase in $\mathrm{CO_2}$ concentration.

Therefore, we introduce the conversion factor $\beta_{\mathrm{Vol}}$ as a *nuisance parameter* to the analysis. When estimating the likelihood $\mathcal{L}(\theta; \hat{y}; \beta_{\mathrm{Vol}})$, we allow $\beta_{\mathrm{Vol}}$ to be adjusted freely, so that the maximum likelihood for a given $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$ is achieved (see Section 2.5). Figure 2.2.h summarises our best estimate for the total radiative forcing during the industrial era. The forcing trajectory was generated for values of $Q_{\mathrm{S90}} = -0.97\,\mathrm{W\,m^{-2}}$ and $\beta_{\mathrm{Vol}} = -9.21\,\mathrm{W\,m^{-2}}$ which provided a good fit to the 20th century temperature record (see Fig. 2.5). The low value for $\beta_{\mathrm{Vol}}$ indicates that the volcanic forcing had to be scaled down considerably in order to simulate a realistic surface temperature response.

It can be seen that the total forcing is a complex mixture of a strongly increasing anthropogenic contribution modulated by solar forcing and interspersed by strong radiative cooling events after volcano eruptions. Note that we have only taken into account volcano eruptions in the 20th century. The 19th century has seen some of the largest eruptions in history, in particular the eruptions of Tambora (Sumbawa in Indonesia, 1815) and Krakatau (Indonesia, 1883). However, the volcanic forcing in the early part of the instrumental record, i.e., in the period 1870-1899, is very uncertain. Estimates for the Krakatau eruption are generally high, but its climatic fingerprint is virtually absent from the SST time series (see Fig. 2.3.a). Harvey and Kaufmann (2002) noted that this mismatch lead to the endorsement of low climate sensitivities from a comparison of their energy balance model with the instrumental temperature record. To avoid the possibility of spurious results due to an overestimation of the Krakatau aerosol, we restrict ourselves to the better-known forcing and temperature estimates for the 20th century. In contrast to Harvey and Kaufmann (2002), an overestimation of volcanic forcing would lead to a bias to higher climate sensitivities in our analysis (see Section 2.5).



a. Sea surface temperature (SST)

b. Land surface air temperature (LAT)

Figure 2.3: Global mean temperature record for the period 1870-2002. Note the different scalings of the $y$-axis.

## 2.3 The residual between model response and data

The response of our energy balance model to the radiative forcing in the industrial era can be compared with observations of global mean sea surface temperature (SST) and land surface air temperature (LAT). Figure 2.3 shows the time series of globally aggregated temperature measurements since 1870. The global mean LAT series is based on the comprehensive CRU data set of instrumental temperature observations on land (Jones and Moberg, 2003). The SST series has been derived from the HadSST data set (Jones et al., 2001). Both time series were provided by courtesy of the UK MetOffice. Measurement errors are estimated to be around 0.05 K for the first half, and 0.025 K for the second half of the 21st century (Folland et al., 2001). We neglect measurement errors here, since they are superseded by fluctuations due to natural temperature variability. Estimates of the standard deviation of global and annual mean surface temperatures from detrended observational data and control runs with atmosphere-ocean general circulation model lie in the range $0.09 - 0.14\,\mathrm{K}$ (Collins et al., 2001; Harvey and Wigley, 2003). The variance of natural temperature variability over land is approximately three times larger than over the ocean, so that standard deviations of annual mean global LAT and SST can be estimated to lie in the range $0.12\,\mathrm{K} - 0.20\,\mathrm{K}$ and $0.07\,\mathrm{K} - 0.11\,\mathrm{K}$, respectively.

As is apparent from Fig. 2.3, it is not possible to directly observe the undiluted secular response of the climate system to the radiative forcing in the industrial era. The annual global mean temperature responds not only to the external forcing, but also to heat fluxes from fast atmospheric processes as well as to interannual variability of the ocean-atmosphere system. Therefore, we need to state carefully our assumptions about the residual between the observed temperature record and the model response to the radiative forcing. On the basis of these assumptions, we will have to specify the functional form of the likelihood function $\mathcal{L}(\theta; \hat{y}; \beta_{\mathrm{Vol}})$.

Hasselmann (1976) proposed to account for the influence of fast atmospheric processes on the slow components of the climate system by adding them as *white noise* to the deterministic dynamics on longer times scales. Weather phenomena usually exhibit time scales of days to weeks, while the ocean operates on decadal and longer time scales. Hasselmann (1976) showed that slow climate components like the ocean act as a low pass filter that dampens the fast frequencies, and retains the slow frequencies in the weather noise. Hence, the weather-driven variability in SSTs constitutes a *red noise* process, which concentrates more energy in the slow frequencies. In contrast, white noise includes all frequencies indiscriminately (Honerkamp, 1994). If the dynamic equations underlying the slow climate component are fairly linear, the weather forcing of the sea surface is converted into an *Autoregressive Process (AR)* of low order, typically AR(1), which is the simplest stochastic process exhibiting a red power spectrum. An AR(1) process emerges when the realisation $X_t$ of a random variable $X$ depends on its predecessor $X_{t-1}$, and a white noise process $\eta$, i.e.,

$$X_t = \alpha\, X_{t-1} + \eta_t\,, \tag{2.10}$$

where $0 \leq \alpha < 1$ is the *propagator* of the AR(1) process, and $\eta = N(0, \sigma)$ a normal IID random variable with zero mean and variance $\sigma^2$ (von Storch and Zwiers, 1999).

In many cases, AR(1) processes are successful to explain the residual variability in slow climate components after removal of a decadal or secular trend. Collins et al. (2001), for instance, analyse a 1000 year control run of HadCM3, and find the interannual variability of global mean surface air and sea surface temperatures to be statistically indistinguishable from an AR(1) process with the exception of the *El Niño Southern Oscillation (ENSO)* (see below). Therefore, we will assume in the following that the weather driven variability in the annual SST time series constitutes an AR(1) process. We will check, however, whether the residual between modelled trend and instrumental SST record is better explained by higher order AR processes. There exists an ongoing discussion about the presence of a power-law persistence in observed temperature time series (Koscielny-Bunde et al., 1998; Maraun et al., 2004), but an inclusion of this hypothesis is beyond the scope of our analysis.

The AR(1) assumption extends our set of model hypotheses $M(\theta; \beta_{\mathrm{Vol}})$ to a set $M(\theta; \beta_{\mathrm{Vol}}) + X(\alpha)$ that also includes the interannual variability of SST due to weather forcing. Since our AR(1) assumption introduces the propagator $\alpha$ as new parameter, we have added another nuisance parameter to the analysis. It captures our uncertainty about the stochastic variability of SST that emerges from heat exchange with fast atmospheric processes. In analogy to our treatment of the volcanic forcing uncertainty, we will allow $\alpha$ to be adjusted freely, so that the maximum likelihood for a given $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$ is achieved (see Section 2.5).

The variability in the HadCM3 ocean as well as in the detrended instrumental temperature record shows that not all of it can be explained in terms of an AR(1) process. Notably the El Niño Southern Oscillation in the tropical Pacific can be identified as a pattern of interannual variability that is neither directly related to stochastic weather forcing nor determined by radiative perturbations at the top of the atmosphere (Collins et al., 2001). ENSO consists of two components: a dramatic warming of the normally cold waters of the eastern equatorial Pacific Ocean (El Niño) and large east-west shifts of mass in the tropical atmosphere over the Pacific (Southern Oscillation) (Peixoto and Oort, 1992, Chapter 16.3). During normal conditions, convection of air parcels occur over the warm surface waters of the Indonesian area, which sink back to the surface over the eastern Pacific (Walker circulation). The resulting pressure difference induces an easterly trend wind over the equatorial Pacific, which deepens the thermocline in the Western Pacific. During El Niño conditions the surface waters near the South American coast warm up, which leads to a reduction of the east-west pressure difference and a weakening of the easterly trade winds. This further supports the warming of the eastern Pacific, since warm water sloshes east. During Anti-ENSO (La Niña) conditions the easterly trade winds are stronger than normal, and the temperature difference between the cold eastern and warm western Pacific waters is larger.

A good indicator of the ENSO phenomenon is the pressure difference between the Eastern and the Western Pacific. This is captured by the *Southern Oscillation Index (SOI)*, which constitutes the difference between sea level pressure anomalies in Tahiti (17S, 149W, French Polynesia) and Darwin (12S, 131E, Australia) (e.g., Ropelewski and Jones, 1987; Allan et al., 1991, data publicly available at www.cru.uea.ac.uk/cru/data/-soi.htm). Negative SOI values, i.e. weaker than normal east (Tahiti) - west (Darwin) pressure gradients, indicate ENSO conditions, while positive SOI values indicate Anti-ENSO (La Niña) conditions. Hence, we can use the SOI as a record of ENSO variability in the 20th century. In analogy to the fingerprinting methodology (Hasselmann, 1993; Allen and Tett, 1999), we will account for the possibility that ENSO explains some of the variability in the instrumental SST and land temperature record by extending our set of hypotheses as follows (cmp. Equation 2.1):

$$\tilde{H} = \left\{ \tilde{M}(\theta;\theta_N) = \left( \begin{array}{c} T_L(\theta;\beta_{\text{Vol}}) + r_L + \beta_{L,\text{SOI}}\, I_{\text{SOI}} \\ T_S(\theta;\beta_{\text{Vol}}) + X(\alpha) + \beta_{S,\text{SOI}}\, I_{\text{SOI}} \end{array} \right) \mid (\theta,\theta_N) \in \Omega \times \Omega_N \right\}, \tag{2.11}$$

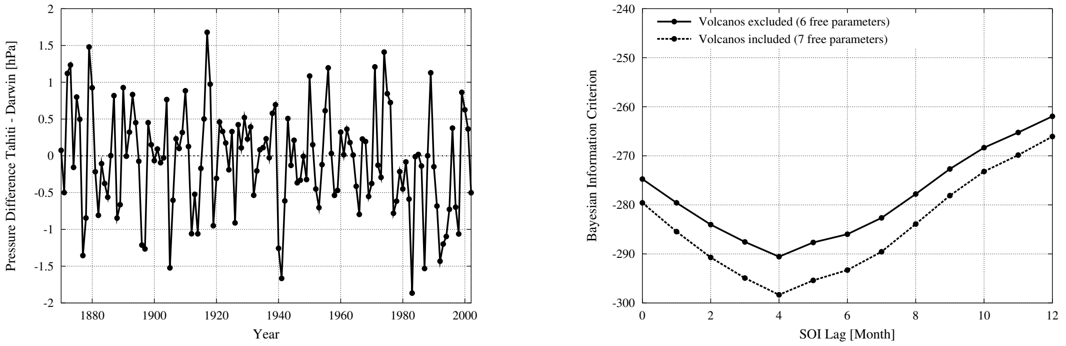
where $I_{\text{SOI}}$ denotes the ENSO 'fingerprint' in time, i.e., the SOI, which is simply added to the model with a free scaling vector $\beta_{\text{SOI}}$. Since ENSO affects ocean as well as land areas, $\beta_{\text{SOI}} = (\beta_{L,\text{SOI}}, \beta_{S,\text{SOI}})$ contains two components for land and sea surface temperature, respectively. The treatment of the residual variability $r_L$ of land temperature will be discussed in Section 2.4. A similar ansatz for the inclusion of ENSO was pursued by Kaufmann and Stern (1997) and Smith et al. (2003). In the framework of our analysis, the scaling vector $\beta_{\text{SOI}}$ introduces two additional nuisance parameters, which can be adjusted to fit the instrumental SST and LAT record. The set of nuisance parameters is summarised by $\theta_N = (\beta_{\text{Vol}}, \alpha, \beta_{L,\text{SOI}}, \beta_{S,\text{SOI}})$.

Fig. 2.4.a shows the annual fluctuations of the SOI since 1870. We have aggregated the monthly SOI data to $x$-month lagged annual averages by averaging over the last $x$ months of the preceding, and the first $12 - x$ months of the present year. The time lag for the annual aggregation was chosen in a way that provides the largest explanatory

power for the observed temperature in the 20th century. To identify the ‘optimal lag’, we have calculated the maximum likelihood that the set of hypotheses $M(\theta;\theta_N)$ explains the instrumental temperature record in the period 1900-2002 by maximising the Log Likelihood (2.19) across the parameters space $\theta, \theta_N$ for different $x$ months lagged SOI indices. The results are shown in Fig. 2.4.b for lags between 0 and 12 months. The use of the *Bayesian Information Criterion (BIC)* (see Equation 2.20) as the goodness of fit measure is substantiated in Section 2.4. For comparisons between fits with the same number of fit parameters, it simply constitutes a negative linear transformation of the maximum Log Likelihood. Hence, the lower BIC the better the fit. Fig. 2.4.b shows that lags between 3 and 6 months provide the best fits to the data with a distinct minimum of BIC at a lag of 4 months. The qualitative behaviour of BIC is not affected if we switch off the forcing from volcanoes ($\beta_{\mathrm{Vol}} = 0\,\mathrm{W\,m^{-2}}$). Our result agrees well with Smith et al. (2003), who identified optimal lags between 4 and 7 months for the annual SOI average from a comparison of hemispheric temperature data with the model MAGICC. In the following, we will use the 4-month lagged SOI index.

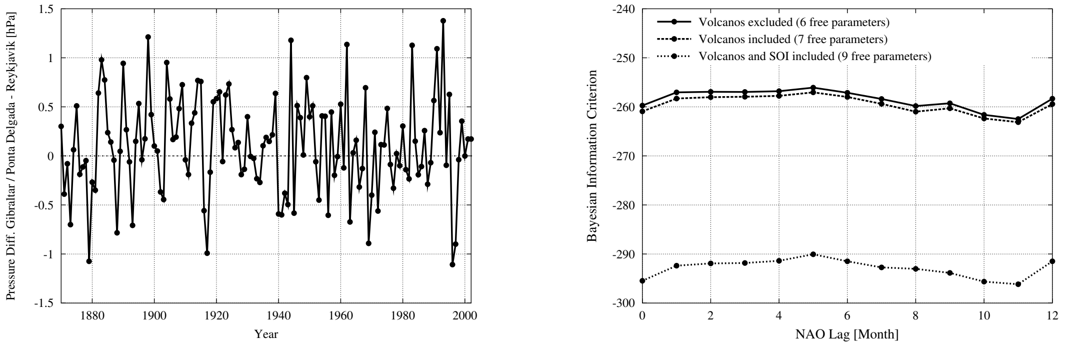
The inclusion of the SOI improves the fit to the global temperature record considerably. Hence, a portion of the observed variability shows the fingerprint of the El Niño Southern Oscillation in the 20th century. This is not a surprising result, since the modulation of global mean SST by ENSO has been noted for some time (Jones, 1989). For our best fit to the data, we find values of $\beta_{S,\mathrm{SOI}} = -0.063$ for SST, and $\beta_{L,\mathrm{SOI}} = -0.148$ for LAT. The difference between land and ocean reflects the fact that the temperature variability over land is larger. The scaling factor needs to be negative since the east-west pressure gradient is weaker than normal in warm El Niño years.

Besides ENSO, we have considered other modes of interannual climate variability. The North Atlantic Oscillation (NAO) is associated with cyclical fluctuations of air pressure between Iceland and the Azores. It explains a large fraction of the variability of atmospheric disturbances in the North Atlantic region, and consequently changes in temperature and moisture distribution and the intensity, number and track of storms (Hurrell et al., 2003). A frequently used NAO index comprises the difference in sea level pressure anomalies between Ponta Delgada (Azores) and Rejkyavik (Iceland). Fig. 2.4.c shows an extended version of the index which was supplemented by pressure differences between Gibraltar and Rejkyavik for the winter half of the year (Hurrell, 1995, data publicly available at www.cru.uea.ac.uk/cru/data/nao.htm). In the same manner as for ENSO, we checked the power of the index to explain the variability in the global SST and LAT record for different lags in the annual aggregation. As depicted in Fig 2.4.d, the index helps little to explain the temperature variability, no matter what lag is assumed for the annual aggregation. The minimum at a lag of 11 months is due to an improved match of variability in land temperatures, but this correlation may be spurious. The overall result is not unexpected, since NAO owes its existence not to atmosphere-ocean interactions, but to dynamics intrinsic to the troposphere (Thompson et al., 2003). Furthermore, it covers only a small region of the world’s land and ocean surface. If NAO is included together with ENSO, its explanatory power improves, but



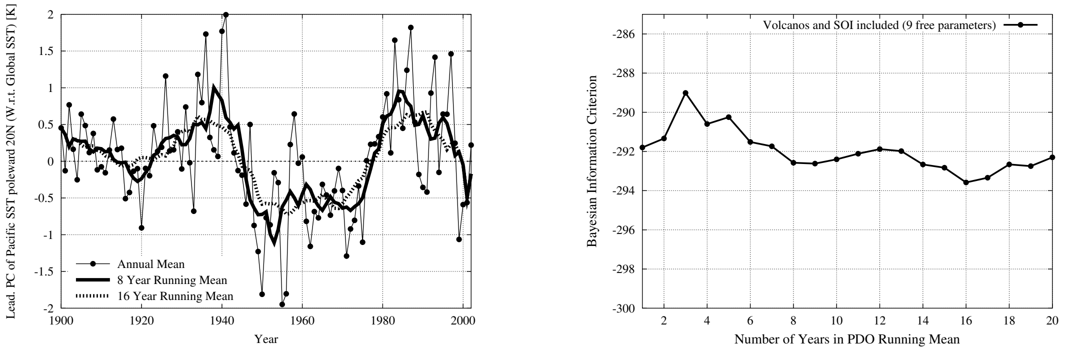
a. Southern Oscillation Index (4-month lagged)

b. SOI: BIC for lags 0-12 months



c. NAO Index (11-month lagged)

d. NAO: BIC for lags 0-12 months



e. PDO Index

f. PDO: BIC for 1-20-year running mean

Figure 2.4: Indices of decadal climate variability over the oceans. Shown are the "$x-$month lagged" averages (last $x$ months of the preceding year plus first $12-x$ month of the present year) of ENSO and NAO (Panels a, c), and the $x$-year running mean of the PDO (centered around the running year; Panel e). Right panels show the goodness of fit in terms of the Bayesian Information Criterion (BIC) that can be achieved when the indices with various lags (SOI, NAO), respectively running means (PDO), are added to the model response.

the gain is still outweighed by the introduction of two more nuisance parameters (SOI: BIC = -298.4, NAO+SOI: BIC =-296.2). Therefore, we will not include NAO in our analysis. Since the analysis of NAO confirms the expectation that purely atmospheric oscillations contribute little to interannual variability in global mean temperatures, we do not consider other modes of such variability. This concerns in particular the large scale annular modes in the northern and southern hemisphere, known as the Arctic Oscillation (AO) and Antarctic Oscillation (AAO).

Long-lived patterns of SST variability have been identified predominantly in the Pacific ocean: an ENSO-like decadal variability in the tropical Pacific (Zhang et al., 1997; Knutson and T. R. Manabe, 1998), and an interdecadal variability of SSTs in the North Pacific (Latif and Barnett, 1996; Zhang et al., 1997), which was called Pacific Decadal Oscillation (PDO) by Mantua et al. (1997). It has remained speculative to date whether the PDO constitutes a genuine pattern of atmosphere-ocean oscillation apart from ENSO (Mantua and Steven, 2002). Collins et al. (2001) have found similar patterns of SST variability in HadCM3, but could not identify a cyclic mode of variability. Nevertheless, we have investigated an index of PDO variability for the period 1900-1993 that was provided by Zhang et al. (1997), and whose extended version to 2003 is available publicly at jisao.washington.edu/pdo/PDO.latest. It comprises the leading principal component from an EOF analysis of monthly ‘residual’ North Pacific SST anomalies, poleward of 20N. Residuals are here defined as the difference between observed anomalies and the monthly mean global average SST anomaly. Fig. 2.4.e shows the time series of the annual PDO index, along with 8-year and 16-year centered running means. Incorporating the PDO index in our set of hypotheses improves the fit to the global SST record notably, but much less than an incorporation of the SOI can achieve (SOI: BIC = -298.4, PDO: BIC = -282.4). Since the PDO index is strongly correlated with ENSO, much of this improvement can be attributed to ENSO variability. The question remains, whether the specific interdecadal fingerprint of the PDO, which is not present in the ENSO signal, helps to explain SST variability on these time scales. Therefore, we have checked the explanatory power of adding the smoothed PDO index for running means of 2 to 20 years together with ENSO to the set of hypotheses. Fig. 2.4.f shows that the improvement is too small to justify an inclusion of the PDO index in our analysis (PDO+SOI: BIC = -293.6). The result might be surprising since the PDO index seems to reflect the cooling and warming periods in the 20th century. A close inspection reveals, however, that the Pacific cooling in the 40s and the Pacific warming in the late 70s precedes global trends by 4-5 years.

## 2.4 Definition of a likelihood for the model parameters

The extended set of hypotheses $\tilde{H} = \{\, M(\theta, \theta_N) \,|\, (\theta, \theta_N) \in \Omega \times \Omega_N \,\}$ about the climate response to the radiative forcing and the residual temperature variability can be directly compared with the temperature record in the 20th century. We recall that $\theta$ summarises the climate and forcing model parameters of interest, and $\theta_N$ the nuisance parameters. In the following, we establish a likelihood function $\mathcal{L}(\theta; \theta_N; \hat{T}_S)$ that the combined SST and LAT record would have been observed for a given assumption about $(\theta, \theta_N)$. The joint likelihood function consists of two components: one for SST, and the other for the difference $T_{L-S} = T_L - T_S$ between land air and sea surface temperature.

**SST:** The likelihood for observing the SST record $\hat{T}_S$ for a hypothesis $\tilde{H}(\theta, \theta_N)$ can be calculated from the residual $r_S = \hat{T}_S - T_S(\theta; \beta_{\text{Vol}}) - \beta_{S,\text{SOI}} I_{\text{SOI}}$ between the data and the deterministic part of our model including the SOI. If our hypothesis was true, then the residual would have to be a realization of the AR(1) process $X(\alpha)$ (see Equation 2.11). The probability density of obtaining a time series $y$ with length $n$ as a realization of an AR(1) process $X(\alpha)$ is well-known from the literature, i.e.,

$$\rho(y|\alpha) = \left( \sqrt{2\pi}^n \sqrt{|\Sigma(\alpha, \sigma_S)|} \right)^{-1} e^{-\frac{1}{2} y^T \Sigma^{-1}(\alpha, \sigma_S) y} ,$$

with $\Sigma(\alpha, \sigma_S)$ the $n \times n$ covariance matrix of the n-step AR(1) process $X(\alpha)$ (von Storch and Zwiers, 1999, Chapter 11). Thus, the likelihood function is given by

$$\mathcal{L}_S(\theta; \theta_{S,N}; \hat{T}_S) = \frac{e^{-\frac{1}{2} r_S(\theta, \beta_{\text{Vol}}, \beta_{S,\text{SOI}})^T \Sigma^{-1}(\alpha, \sigma_S) r_S(\theta, \beta_{\text{Vol}}, \beta_{S,\text{SOI}})}}{\sqrt{2\pi}^n \sqrt{|\Sigma(\alpha, \sigma_S)|}} , \tag{2.12}$$

where we have summarised the active nuisance parameters by $\theta_{S,N} = (\beta_{\text{Vol}}, \alpha, \beta_{S,\text{SOI}})$. The functional form of $\mathcal{L}_S(\theta; \theta_{S,N}, \hat{y}_S)$ can be simplified by noting

$$r_S^T \Sigma^{-1}(\alpha, \sigma_S) r_S = \frac{1}{\sigma_S^2} \sum_{i=1}^{n} (r_{S,i} - \alpha r_{S,i-1})^2 , \tag{2.13}$$

$$|\Sigma(\alpha, \sigma_S)| = \frac{\sigma_S^{2n}}{1 - \alpha^2} , \tag{2.14}$$

where $r_S = (r_{S,1}, ..., r_{S,n})$ is the time series of the residual (von Storch and Zwiers, 1999, 12.2.4). Since we compare the model output with the SST record in the period 1900-2002, the time series contains 103 data points between $t_1 = 1900$ and $t_n = 2002$. If the predecessor residual $r_{S,0}$ was unknown, the term $(r_{S,1} - \alpha r_{S,0})^2$ in Equation (2.13) would need to be replaced by $r_{S,1}^2 (1 - \alpha^2)$. In our case, however, the predecessor residual in the year 1899 is known. We note that the exponent of the likelihood function as written in Equation (2.13) constitutes the square sum of $n$ IID random variables $\sim N(0, 1)$, and is distributed like $\chi^2$. To further simplify matters, the logarithm of the likelihood function is usually used instead of the likelihood itself. When inserting Equations (2.13) and (2.14) into expression (2.12), the *Log Likelihood (LLH)* is given by

$$LLH_S = -\frac{1}{2} \sum_{i=1}^{n} \frac{(r_{S,i} - \alpha r_{S,i-1})^2}{\sigma_S^2} - \frac{n}{2} \ln\left(2\pi \sigma_S^2\right) + \frac{1}{2} \ln\left(1 - \alpha^2\right) .$$

In contrast to $\alpha$, we do not treat the variance $\sigma_S^2$ of the Gaussian white noise process $\eta$ (see Equation 2.10) as a nuisance parameter. Under the AR(1) assumption, $\sigma_S^2$ is determined by the variance $Var(r_S)$ of the residual, and the propagator $\alpha$ via $\sigma_S^2 =$

$Var(r_S)\,(1-\alpha^2)$. Hence, the Log Likelihood can be rewritten as

$$LLH_S = -\frac{1}{2}\sum_{i=1}^{n}\frac{\left(r_{S,i}-\alpha r_{S,i-1}\right)^2}{Var(r_S)(1-\alpha^2)} - \frac{n}{2}\ln\left(2\pi\, Var(r_S)\right) - \frac{n-1}{2}\ln\left(1-\alpha^2\right) . \tag{2.15}$$

An accurate value for $Var(r_S)$ is difficult to extract from control runs with AOGCMs because it describes the variance of internal variability in global and annual mean SST *after* substraction of the ENSO signal as described by the Southern Oscillation Index. All state-of-the-art AOGCMs produce some kind of ENSO, but the model signals often deviate in aspects like periodicity, amplitude and spatial extent from the observations (Harvey and Wigley, 2003). In this situation, we estimate $Var(r_S)$ directly from the observations after detrending the instrumental SST record from the forced climate response and the ENSO signal. The detrending is done by fitting our energy balance model together with the SOI index to the data, so that the Log Likelihood (2.15) is maximised. In each step of the optimisation process, $Var(r_S)$ is estimated from the variance $\hat{V}ar_S(r) = \frac{1}{n-1}\sum_{i=1}^{n}\left(r_{S,i}-\bar{r}_S\right)^2$ of the residual between data and model plus ENSO signal. In this way, we find a maximum likelihood estimate of $\hat{\sigma}_S := \hat{\sigma}(r_S) = 0.078\,\mathrm{K}$ for the standard deviation of the detrended SST variability. If ENSO is included in the SST variability, the standard deviation increases to 0.090 K. This value is in the centre of the range of estimates that can be deduced from AOGCM control runs (see Section 2.3). If the observations are detrended only with a linear trend, we find a higher standard deviation of 0.10 K for SST variability.

The question remains if the detrended SST variability conforms indeed with our assumption of an AR(1) process. We test the hypothesis in three steps. First, the residual is decorrelated by calculating the time series $\eta_{S,i} = r_{S,i} - \hat{\alpha} r_{S,i-1}$ of the supposedly Gaussian white noise process (see Equation 2.10), where $\hat{\alpha} = 0.45$ is the maximum likelihood estimate for the propagator. Secondly, we test the residual $\eta_S$ for white noise with a *Portmanteau statistics*, which evaluates the variability of autocorrelation in the residual over various lags. If the residual was white, the autocorrelation $\rho_\tau$ over the first $\tau \leq k$ lags would fluctuate around zero, and its square sum should be distributed like $\chi^2(k-1)$ with $k-1$ degrees of freedom (von Storch and Zwiers, 1999, Chapter 12.2). Since the sample size of $n = 103$ is small, we can only use lags up to $k \leq 5-7$ for the calculation of the square sum. To account for the bias in the estimator of the autocorrelation function, i.e., $\hat{\rho}_\tau = \frac{1}{n}\frac{1}{\hat{V}ar(\eta)}\sum_{i=\tau}^{n}\left(\eta_{i-\tau}-\bar{\eta}\right)\left(\eta_i-\bar{\eta}\right)$ (von Storch and Zwiers, 1999, Chapter 12.1), we use the Young-Box-Pierce Portmanteau Test which inflates the empirical autocorrelation $\hat{\rho}_\tau$ by a factor $n/(n-\tau)$. The test yields a $p$-value of 0.49 (lags up to $k = 5$) which means that the Null hypothesis of $\eta_S$ being the realisation of a white noise process cannot be rejected at any reasonable significance level. To check the power of the test we apply it to the correlated residual $r_S$, and find a p-value of $10^{-4}$. Hence, the test correctly rejects the white noise hypothesis for the residual between observations and model plus ENSO signal. Finally, we need to test the hypothesis that $\eta_S$ is normally distributed. The standard normality tests compare the empirical distribution function of the realisation $\eta_S$ with the normal distribution function $N(\hat{\mu}, \hat{\sigma})$, where mean and variance are estimated from the sample. We have applied a set of established normality tests to the realisation of $\eta_S$ (Kolmogorov-Smirnov-Lilliefors, Cramer-von Mises, Anderson-Darling, Pearson chi-square, Shapiro-Wilk, Shapiro-Francia), and found $p$-values between 0.09 and 0.30. On this basis, the Null hypothesis of $\eta_S$ being normally distributed cannot be rejected at the 5% significance level by any of these tests. To check their power, we have applied them to the correlated residual $r_S$ between model output and data. 4 out of 6 tests correctly rejected the Null hypothesis of a normally distributed residual $r_S$ at a 5% significance level. We conclude that the residual variability in observations after removal of the model and ENSO signal exhibits no significant evidence against the assumption of an AR(1) process.

**LAT:** The variability in global and annual mean land temperatures will be highly correlated with the variability in global and annual mean sea surface temperatures. Therefore, it is not possible to aggregate likelihoods for the reproduction of LAT and SST observations under an independence assumption. This is highlighted by investigating the residuals between data and model plus ENSO signal for our best fit depicted in Fig. 2.5. We check for correlations between the residual $r_L$ for LAT and the residual $r_S$ for SST by means of an asymptotically optimal test that was recently presented by Hallin and Saidi (2003, 2005) for time series of multivariate ARMA processes. They showed that their test is more powerful and exhibits more accurate rejection frequencies for small time series with $n \approx 100$ data points, than the classical Portmanteau test of cross-correlations by Haugh (1976). When applying the test to the residual time series for LAT and SST that were obtained from our best fit to the data depicted in Fig. 2.5, we find a $p$-value of 0.004. Hence, the test is powerful enough to clearly reject the Null Hypothesis of no correlation between LAT and SST residuals.

Since the variability in the SST and LAT record is correlated, we need to find some piece of additional information in the LAT record, that is fairly independent of the SST record. A good candidate is the difference in warming and cooling trends over land and over the ocean, $\Delta T = \hat{T}_L - \hat{T}_S$. When applying the test of Hallin and Saidi (2003) to the residuals for SST and $\Delta T$ in the period 1900-2002, we find a p-value of 0.14 for the Null Hypothesis of no correlation between the two residuals (under the restrictive assumption that the $\Delta T$ residual is white noise). A problem arises from the presence of radiative forcing spikes after volcano eruptions. Since the thermal inertia of the ocean is much larger than that of the landmass, these spikes are more visible in the land temperature record than in the SST record. Since the energy balance model DOECLIM has no spatial resolution, it overestimates the amplitude of the resulting spikes in $\Delta T$, i.e., the cooling over land relative to the cooling of the sea surface is too large. In addition, we are using a globally uniform volcanic forcing, which is an inadequate choice when it comes to assess the difference in cooling over ocean and land. As a result of this model inadequacy, inclusion of the residual on $\Delta T$ along with the residual on SST in the likelihood function leads to artificially low choices for the nuisance parameter $\beta_{\text{Vol}}$ ($\beta_{\text{Vol}} \approx -5.4\,\text{W}\,\text{m}^{-2}$ instead of $\beta_{\text{Vol}} \approx -11.2\,\text{W}\,\text{m}^{-2}$ for a fit to SST alone). In order limit the bias from the inadequate representation of volcanic spikes in $\Delta T$, we remove periods of high volcanic activity from the $\Delta T$ time series to be included in the likelihood function. Only periods of five years or longer were retained that are not interspersed with years of volcanic aerosol optical depth that surpassed an heuristically chosen threshold value ($\tau \leq 0.007$). This leaves us with the periods 1914-1921, 1924-1928, 1934-1952, 1955-1962, 1977-1981, 1986-1990, and 1996-2002, where no globally significant volcanic activity took place (see Figure 2.2.e). If the model plus ENSO signal is fitted to the $\Delta T$ values in these 57 years together with the SST values for the period 1900-2002, the optimisation algorithm adopts a volcanic forcing conversion factor of $\beta_{\text{Vol}} \approx -9.4\,\text{W}\,\text{m}^{-2}$. Hence, the overestimation of volcanic spikes in the modelled $\Delta T$ signal was indeed responsible for the artificially low estimates of $\beta_{\text{Vol}}$. The removal of periods with volcanic activity also increases the p-value from 0.14 to 0.55 that can be obtained from the test of Hallin and Saidi (2003) for a correlation between the residuals in SST and $\Delta T$ in the remaining 57 years.

Encouraged by these preliminary checks, we have decided to incorporate the information provided by the observed temperature differences $\hat{T}_{L-S} = \hat{T}_L - \hat{T}_S$ in the remaining 57 years into the likelihood function. Let $r_{L-S}(\theta, \beta_{\text{Vol}}, \beta_{L-S,\text{SOI}}) = \hat{T}_{L-S} - (T_L(\theta; \beta_{\text{Vol}}) - T_S(\theta; \beta_{\text{Vol}})) - (\beta_{L,\text{SOI}} - \beta_{S,\text{SOI}})\ I_{\text{SOI}}$ be the residual between the land-sea temperature differences in the instrumental record and in the model plus ENSO signal. Since the variability over land is essentially a white noise process (Collins et al., 2001), any autocorrelation in the LAT series will stem from heat exchange with the ocean areas. We expect this autocorrelation to be largely removed in the land-sea temperature difference $\Delta\hat{T}$. Hence, we specify the likelihood function for reproducing $\hat{T}_{L-S}$ with our model plus ENSO hypothesis by assuming that the residual $r_{L-S}$ constitutes a Gaussian white noise process, i.e.,

$$\mathcal{L}_{L-S}(\theta; \theta_{L-S,N}; T_{L-S}) = \frac{1}{\sqrt{2\pi}^{m}\ \sigma_{L-S}^{m}}\, e^{-\frac{1}{2\,\sigma_{L-S}^2}\ \sum_{i=1}^{m}\ r_{L-S,i}^2(\theta, \beta_{\text{Vol}}, \beta_{L-S,\text{SOI}})}\ , \tag{2.16}$$

where $m = 57$ is the number of years included in the likelihood estimation, and the active nuisance parameters are summarised by $\theta_{L-S,N} = (\beta_{\text{Vol}}, \beta_{L-S,\text{SOI}} := \beta_{L,\text{SOI}} - \beta_{S,\text{SOI}})$.

In analogy to the standard deviation of the residual SST variability, we estimate $\sigma_{L-S}$ directly from observations after detrending them with the model plus ENSO signal. This is done by maximising the Log Likelihood,

$$LLH_{L-S} = -\frac{1}{2} \sum_{i=1}^{m} \frac{r_{L-S,i}^2}{\sigma_{L-S}^2} - \frac{m}{2} \ln\left(2\pi\, \sigma_{L-S}^2\right), \tag{2.17}$$

where $\sigma_{L-S}^2$ is estimated from the variance $\hat{V}ar(r_{L-S})$ of the residual in each step of the optimisation procedure. In this way, we find a best estimate of $\hat{\sigma}_{L-S} = 0.126\,\text{K}$ for the standard deviation of the variability in the residual $r_{L-S}$. If we do not remove the ENSO signal from the $\hat{T}_{L-S}$ record when calculating the residual, the standard deviation increases to 0.139 K. An analysis of the global and annual mean temperature difference $T_{L-S}$ in a 1000-year HadCM3 control run yields a little lower estimate of 0.132 K. We conclude that the estimate from our detrending procedure lies in a reasonable range supported by the HadCM3 control run. For completeness, we test the residual $r_{L-S}$ for our assumption of Gaussian white noise. The Young-Box-Pierce Portmanteau test for white noise yields a p-value of 0.49 (lags up to $k = 3$), and the standard normality tests find p-values above 0.5 with the exception of the Kolmogorov-Smirnov-Lilliefors test (p-value = 0.25). Hence, the tests provide no significant evidence that our assumption of Gaussian white noise is false.

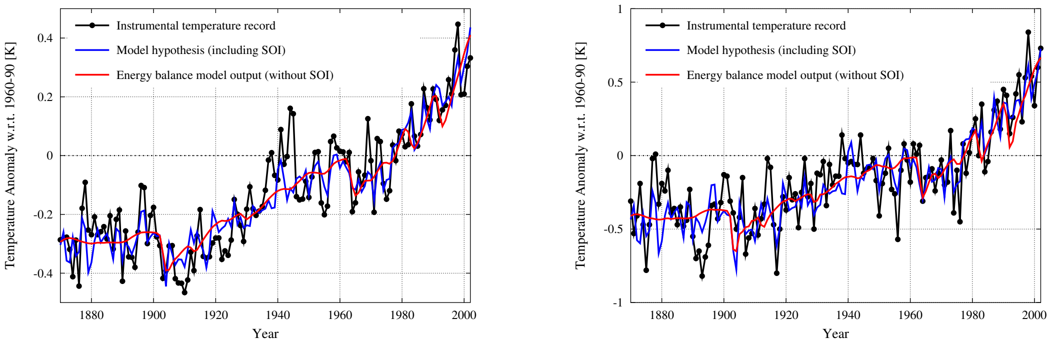
We aggregate the two likelihood functions for SST and the land-sea temperature difference by assuming independence of the corresponding residual time series between data and model plus ENSO signal, i.e.,

$$\mathcal{L}(\,\theta;\theta_N;(\hat{T}_S,\hat{T}_{L-S})\,) = \mathcal{L}_S(\theta;\theta_{S,N};\hat{T}_S)\cdot\mathcal{L}_{L-S}(\theta;\theta_{L-S,N};\hat{T}_{L-S})\,. \qquad (2.18)$$

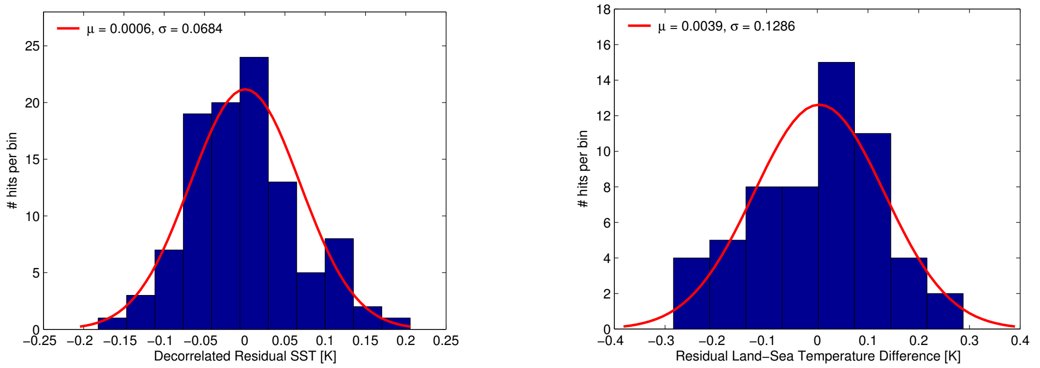
For the calculation of the likelihood function, we compute the joint Log Likelihood

$$LLH = -\frac{1}{2}\sum_{i=1}^{n}\frac{\left(r_{S,i}-\alpha r_{S,i-1}\right)^2}{\hat{\sigma}_S^2\left(1-\alpha^2\right)} - \frac{n-1}{2}\ln\left(1-\alpha^2\right) - \frac{1}{2}\sum_{j=1}^{m}\frac{r_{L-S,j}^2}{\hat{\sigma}_{L-S}^2} + N\,, \qquad (2.19)$$

which includes altogether $n + m = 160$ data points. The constant $N = -\frac{n}{2}\ln\left(2\pi\,\hat{\sigma}_S^2\right) - \frac{m}{2}\ln\left(2\pi\,\hat{\sigma}_{L-S}^2\right)$ summarises the fixed contribution from the estimated variance of natural variability in the residuals. As discussed above, we assume values of $\hat{\sigma}_S = 0.078\,\mathrm{K}$ and $\hat{\sigma}_{L-S} = 0.126\,\mathrm{K}$. The Log Likelihood is a function of the uncertain climate model and forcing parameters $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$, and the nuisance parameters $\theta_N = (\beta_{\mathrm{Vol}}, \alpha, \beta_{L,\mathrm{SOI}}, \beta_{S,\mathrm{SOI}})$. The maximum likelihood estimate for these parameters can be found by maximising Expression (2.19) under the dynamic constraints (2.5) and (2.6) provided by the energy balance model DOECLIM, and the linear addition of the scaled SOI index (see Equation 2.11). We tackled the optimisation problem by implementing the model in the optimisation package GAMS (Brooke et al., 1992), and solving it with the non-linear problem solver CONOPT2 (Drud, 1992). GAMS views the difference equations (A.25) and (A.26) for the numerical integration of DOECLIM as $2\,l$ static constraints ($l = 253$ years in the period 1750-2002). In a first step, CONOPT2 searches for a feasible solution satisfying the constraints, which is equivalent to a forward integration of DOECLIM for some parameter constellation. In a second step, the free parameters $\theta, \theta_N$ are varied with a gradient-based optimisation procedure to search for a local maximum of expression (2.19), while assuring dynamic consistency by observing the $2l$ constraints. It is important to note that the simple structure of DOECLIM allows us to search the uncountable space of possible hypotheses in an efficient way. This is complementary to the case of complex climate models, which only



a. Sea surface temperature b. Land surface air temperature



c. Decorrelated Residual SST (Data - Model) d. Residual $\Delta T = T_L - T_S$ (Data - Model)

Figure 2.5: Comparison of the model simulation with the instrumental temperature record for the period 1870-2002. The model was fitted to the SST record in the period 1900-2002 and to the difference in land and sea surface temperatures in the periods 1914-1921, 1924-1928, 1934-1952, 1955-1962, 1977-1981, 1986-1990, and 1996-2002.

allow to compare a linear superposition of a finite number of response patterns with the observations.

The optimisation procedure finds a maximum of the likelihood function at the point $T^*_{2x} = 3.26\,\mathrm{K}$, $\kappa^*_v = 0.55\,\mathrm{cm^2\,s^{-1}}$, $Q^*_{\mathrm{S90}} = -0.97\,\mathrm{W\,m^{-2}}$, $\beta^*_{\mathrm{Vol}} = -9.21\,\mathrm{W\,m^{-2}}$, $\alpha^* = 0.45$, $\beta^*_{S,\mathrm{SOI}} = -0.063$, and $\beta^*_{L,\mathrm{SOI}} = -0.147$. We have varied the initial point of the optimisation to check for multiple local maxima, and found no indication for their existence. Our scan of the entire parameter space $\Omega$ discussed below reveals a quasi-concave shape of the likelihood function which further supports this finding. However, we warn to put too much emphasis on the maximum likelihood estimate. There exists a large range of other parameter constellations with similarly high likelihood values. The full structure of the likelihood function will be discussed in Section 2.5.

Figure 2.5.a-b shows the resulting best fit of the model and ENSO signal to the instrumental temperature record. The response of DOECLIM to the radiative forcing in the industrial era is depicted in red, and the model hypothesis after the addition of the ENSO signal is shown in blue. The model plus ENSO signal reproduces the

instrumental temperature record well. This is also true for the period 1870-1899, which was not included in the fit. Significant deviations of the hypothesized SST trajectory exist for the periods 1907-11, and 1922-25, where the observations show a stronger cooling than simulated by the model, and for the exceptionally warm years 1937-38, and 1943-45 (The exceptionally warm year 1941 was a strong El Niño year, and therefore can be explained). The deviations between modelled SST and observed SST in the early 20th century are not found for land temperatures. Therefore, we speculate that they might arise from a combination of interannual ocean variability that is not accounted for in the model, and a bias to cooler SSTs in the observations due to sparse data coverage in these periods. However, the exceptionally warm years 1937-38, and 1943-45 are present in the SST and LAT record, and cannot be explained by the model. We also note that the model fits the data considerably better in the second half of the 20th century than in the first half. We attribute this to the combination of two effects. First, the temperature record in the last decades was dominated by the strong increase in anthropogenic forcing which the model can reproduce well. In the earlier years, a greater role was played by natural variability and forcing, which are more difficult to account for adequately. Secondly, the systematic errors in the instrumental temperature record are certainly lower for the second half of the 20th century. This is also true for the Southern Oscillation Index, whose quality has been questioned for Tahitian sea level pressure measurements before 1935 (Ropelewski and Jones, 1987).

Histograms of the residuals between instrumental temperature record and the maximum likelihood fit of model plus ENSO signal are shown in Figure 2.5.c-d together with fits of a normal distribution (SST: Time decorrelated $\eta_S$ in the period 1900-2002, $T_{L-S}$: $r_{L-S}$ in the years 1914-1921, 1924-1928, 1934-1952, 1955-1962, 1977-1981, 1986-1990, and 1996-2002). We test the residuals for our Null hypothesis of independent Gaussian white noise by the series of tests established above. The Young-Box-Pierce Portmanteau test for white noise yields p-values of 0.61 ($\eta_S$, up to $k = 5$ lags) and 0.45 ($r_{L-S}$, up to $k = 3$ lags). The standard normality tests find p-values for $\eta_S$ in the range 0.12 to 0.36, and p-values for $r_{L-S}$ above 0.52. The test of Hallin and Saidi (2003) for correlation between $\eta_S$ and $r_{L-S}$ yields a p-value of 0.54. We conclude again that there exists no evidence for rejecting our Null hypothesis at any reasonable significance level.

Nevertheless, we have checked whether our likelihood improves considerably, if we extend the hypothesis about the residual between observations and model simulation in several directions. As already discussed in Section 2.3, we have considered the inclusion of two other modes of interannual to decadal climate variability, i.e., NAO and PDO. We have also checked whether the remaining residual after removal of the ENSO signal is better explained by a higher order AR process rather than an AR(1) process. Since the different hypotheses about the residual include different numbers of free parameters, we need to apply a criterion that allows us to compare the value of the maximum Log Likelihood (2.19) under these circumstances. The *Akaike Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC) are common choices (von Storch and Zwiers, 1999, Chapter 12.2). They both penalise the introduction of new parameters

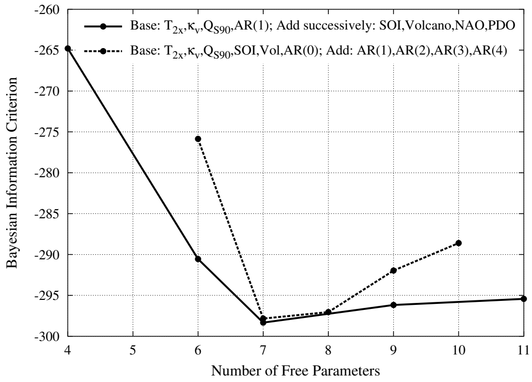


Figure 2.6: Bayesian information criterion for several degrees of freedom in the fit of the model output to the 20th century temperature record. In one case, an AR(1) residual was assumed, and then additional degrees of freedoms (SOI, Volcano forcing, NAO, PDO) successively added. In another case, we included SOI and volcano forcing, and then increased the order of the stochastic process assumed for the residual (White noise to AR(4)). The dashed curve is offset by a small amount for better visibility. Points at $x = 7$ are identical.

to avoid overfitting. In our analysis, we will use BIC which is defined by

$$BIC = -2 \ln \mathcal{L} + p \ln n \, , \tag{2.20}$$

where $p$ is the number of free parameters in the likelihood function, and $n$ the number of data points. In contrast to AIC, BIC scales the penalty with the number of data points, and therefore selects less complex models (in terms of $p$) than AIC even for small samples with $n > 7$. This is an important feature for our application, because we are seeking for new parameters that improve the performance of our hypothesis throughout the period under consideration, i.e., at all data points. Since such improvements are summed up in the Log Likelihood, we want a threshold for the inclusion of a new parameter to depend on the sample size.

The comparison of different model hypotheses is shown in Figure 2.6. Since the BIC comprises the negative Log Likelihood together with a positive penalty for the parameters, the model hypothesis with the lowest BIC is the winner. We have conducted two experiments. In the first experiment, we assumed an AR(1) process for the residual variability between model output and observations, and then increased the complexity of the model hypothesis. We started with a model that excluded the contribution of volcanos to the radiative forcing as well as the temperature signal of ENSO (free parameters: $T_{2x}, \kappa_v, Q_{\mathrm{S90}}, \alpha$) and then switched on consecutively the ENSO signal (two new parameters: $\beta_{S,\mathrm{SOI}}, \beta_{L,\mathrm{SOI}}$), volcanic forcing (one new parameter: $\beta_{\mathrm{Vol}}$), the NAO signal (two new parameters: $\beta_{S,\mathrm{NAO}}, \beta_{L,\mathrm{NAO}}$) and the PDO signal (two new parameters: $\beta_{S,\mathrm{PDO}}, \beta_{L,\mathrm{PDO}}$). In this experiment as well as in the second, the variances $Var(X)$ of the residual SST variability, and $\sigma^2_{L-S}$ of the residual land-sea temperature difference, were estimated directly from the residual between observations and model hypothesis. It can be seen that the inclusion of the ENSO signal improved the BIC the most, followed by the inclusion of volcanic forcing. In contrast, the addition of NAO and PDO is not supported by the BIC. This repeats our discussion in Section 2.3.

The second experiment consisted in including volcanic forcing and the ENSO signal throughout, and increasing the order of the AR hypothesis for the residual from white noise (AR(0), six free parameters: $T_{2x}, \kappa_v, Q_{\text{S90}}, \beta_{\text{Vol}}, \beta_{S,\text{SOI}}, \beta_{L,\text{SOI}}$) to AR(4) (addition of new parameters $\alpha, \alpha_2, \alpha_3, \alpha_4$ one at a time). A change in the AR hypothesis requires to modify the Log Likelihood (2.19). The case of white noise is obtained by simply setting $\alpha = 0$. For $AR(k)$, we need to replace the terms in the square sum for the SST residuals by $r_{S,i} - \sum_{j=1}^{k} \alpha_j \, r_{S,i-j}$. In our case, no additional complications arise from the truncation of the time series, since the predecessor residuals $r_{S,0}$ to $r_{S,-3}$ in the years 1896-99 are known. In addition, we need to adjust the estimator $\hat{V}ar(X)/(1-\alpha^2)$ for the variance of the white noise process $\eta_S$ in the AR(k) process by $\hat{V}ar(X)/(1 - \sum_{j=1}^{k} \rho_j \, \alpha_j)$, where $\rho_j$ is the autocorrelation of the AR(k) process at lag $j$. The autocorrelation function $\rho$ is completely determined by the $k$ propagators, and can be derived analytically from the *Yule-Walker equations* (von Storch and Zwiers, 1999, Chap. 11.1). We have implemented the dependence of $\rho_j$ as additional constraints to the optimisation in order to ensure that a change in the free parameters $\alpha_j$ is accompanied by the corresponding change in the autocorrelation function. Hence, the complexity of the optimisation problem increases with the order of the AR hypothesis. Figure 2.6 shows that the assumption of a white residual for sea surface temperature performs badly compared with the AR(1) hypothesis. In contrast, the assumption of higher order AR-processes is not supported by the BIC. The BIC recommends a model that includes volcanic forcing, the ENSO signal, and an AR(1) hypothesis for the residual SST variability, but no other extension that we have considered here. Hence, we will continue to use the set of model hypotheses (2.11).

## 2.5 The likelihood function for the climate model

Since we want to use the likelihood function $\mathcal{L}(\,\theta; \theta_N; (\hat{T}_S, \hat{T}_{L-S})\,)$ in the framework of a Bayesian analysis, we would ideally evaluate it on the entire space of assumptions $\theta, \theta_N$. However, the dimensionality of the set of parameters $\theta, \theta_N$ precludes such an evaluation, as it will be the case in most other analyses of climate change. Fortunately, we have separated the parameters into a set $\theta = (T_{2x}, \kappa_v, Q_{\text{S90}})$ of interest, and a set $\theta_N$ of nuisance parameters that are solely needed for the determination of the likelihood. As pointed out before, we will restrict the calculation of the likelihood function to the space $\Omega \subset \mathbb{R}^3$ spanned by $\theta$, and use the power of optimisation routines to search the remaining space $\Omega_N$ of nuisance parameters for the combination that maximises the likelihood at a given point $\theta$. Hence, we construct the upper envelope $\overline{\mathcal{L}}(\theta; \hat{T})$ of all intersects $\mathcal{L}_{\theta_N}(\theta; \hat{T})$ of the likelihood function at fixed points $\theta_N$. Ideally, we should take all these intersects into account, not just their upper envelope. This would give rise to an *imprecise likelihood function* on the space $\Omega$, which would need to be treated with robust Bayesian methods and imprecise probability analysis, respectively. As pointed out before, we consider this a potentially important area of application of imprecise probability methods in climate change research. In this thesis, however, we need to make a first step, and limit ourselves to imprecision in the prior probabilities that are needed for a Bayesian analysis (see Section 3).

However, we briefly want to assess the implication of using the upper envelope $\overline{\mathcal{L}}(\theta;\hat{T})$ of the set of possible likelihood functions, instead of the likelihood function $\mathcal{L}_{\theta_N}(\theta;\hat{T})$ for some fixed value $\theta_N$. Typically, the choice of $\theta_N$ would be the maximum likelihood estimate $\theta_N^*$, which in our case is given by $\beta^*_{\text{Vol}} = -9.21\,\text{W}\,\text{m}^{-2}$, $\alpha^* = 0.45$, $\beta^*_{S,\text{SOI}} = -0.063$, and $\beta^*_{L,\text{SOI}} = -0.147$ (see Section 2.4). We compare the two choices by evaluating the likelihood of the maximum likelihood estimate $\theta^*$, i.e., in our case $T^*_{2x} = 3.26\,\text{K}$, $\kappa^*_v = 0.55\,\text{cm}^2\,\text{s}^{-1}$, $Q^*_{\text{S90}} = -0.97\,\text{W}\,\text{m}^{-2}$, relative to the likelihood of any other assumption about $\theta$. Clearly, it is

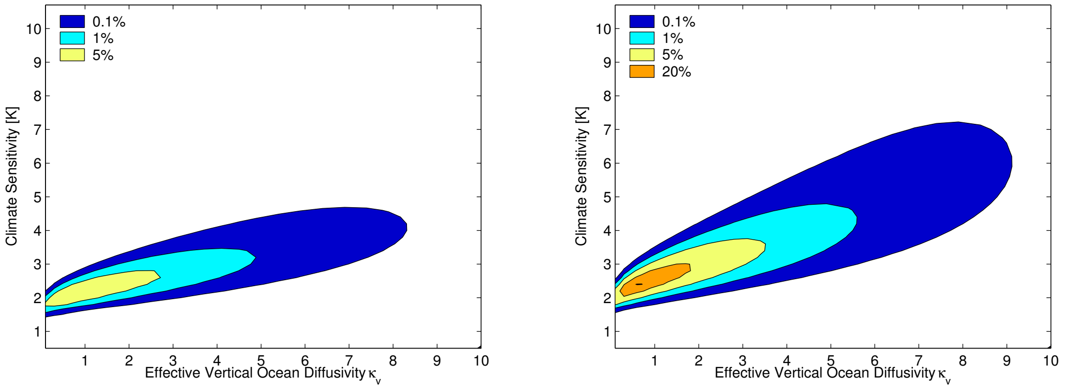
$$\frac{\mathcal{L}_{\theta_N^*}(\theta^*;\hat{T})}{\mathcal{L}_{\theta_N^*}(\theta;\hat{T})} \quad = \quad \frac{\overline{\mathcal{L}}(\theta^*;\hat{T})}{\mathcal{L}_{\theta_N^*}(\theta;\hat{T})} \quad \geq \quad \frac{\overline{\mathcal{L}}(\theta^*;\hat{T})}{\overline{\mathcal{L}}(\theta;\hat{T})}\,.$$

Hence, the choice of the upper envelope decreases the weight of evidence for the maximum likelihood estimate $\theta^*$ relative to the weight of evidence for other assumptions $\theta$. Since the likelihood ratio influences the posterior probability density ratio directly by

$$\frac{\rho(\theta^*|\hat{T})}{\rho(\theta|\hat{T})} \quad = \quad \frac{\mathcal{L}(\theta^*;\hat{T})}{\mathcal{L}(\theta;\hat{T})}\cdot\frac{\rho(\theta^*)}{\rho(\theta)}\,,$$

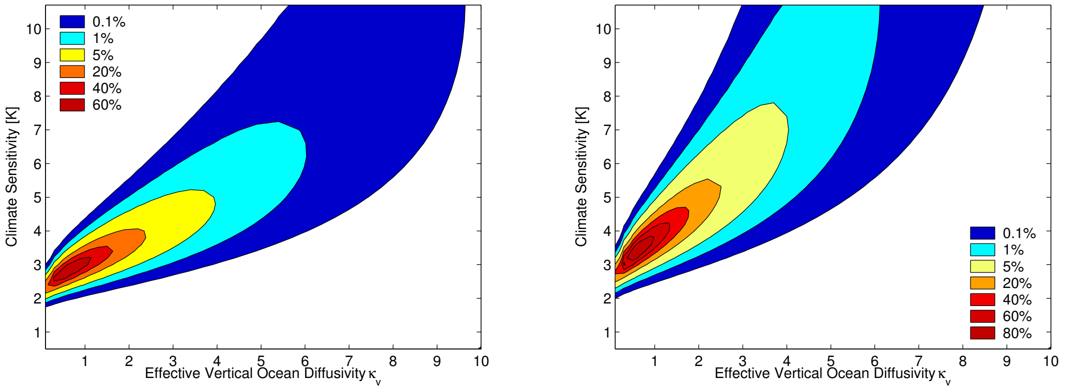
the adoption of the upper envelope $\overline{\mathcal{L}}(\theta;\hat{T})$ can be seen as the most conservative choice for accumulating probability mass around the maximum likelihood.

We have evaluated the likelihood $\overline{\mathcal{L}}(\theta;\hat{T})$ at $35 \times 30 \times 13 = 13650$ points of the parameter space $\Omega = \Omega_{T_{2x}} \times \Omega_{\kappa_v} \times \Omega_{Q_{\text{S90}}}$, and used a cubic spline to interpolate in between these points. Since each evaluation involved an optimisation of the nuisance parameters $\theta_N$, we have fully capitalised on the computational efficiency of our energy balance model. Fig. 2.7 shows six intersects of the three-dimensional likelihood function for fixed values of the sulphate aerosol forcing. The likelihood values are displayed in units of the maximum likelihood. It can be seen that the area of $(T_{2x}, \kappa_v)$ pairs with non-negligible likelihood shrinks considerably towards low values $|Q_{\text{S90}}| < 0.4\,\text{W}\,\text{m}^{-2}$ and high values $|Q_{\text{S90}}| > 1.4\,\text{W}\,\text{m}^{-2}$ for the cooling effect of sulphate aerosols. The shape of the likelihood function shows a positive correlation between climate sensitivity and ocean heat diffusivity, which is particularly strong at intermediate values of $Q_{\text{S90}}$. The rotation of the 'likelihood dumbbell' from an almost horizontal $\kappa_v$-extension to an almost vertical $T_{2x}$-extension for an increase from $|Q_{\text{S90}}| = 0.4\,\text{W}\,\text{m}^{-2}$ to $|Q_{\text{S90}}| = 1.4\,\text{W}\,\text{m}^{-2}$ indicates also a strong positive correlation between climate sensitivity and the strength of the sulphate aerosol cooling. As a consequence of these correlations between $T_{2x}$, $\kappa_v$, and $Q_{\text{S90}}$ in the likelihood function, the point of maximum likelihood reacts very sensitive to changing assumptions about the nuisance parameters $\theta_N$. If



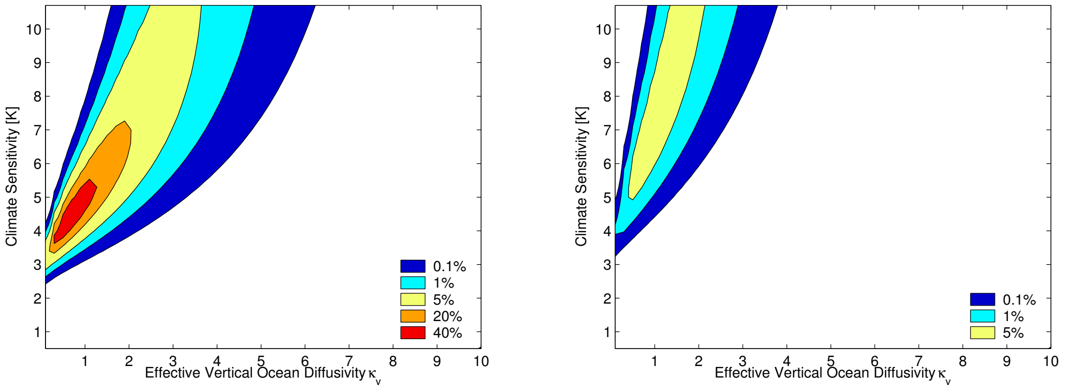
c. $Q_{\mathrm{S90}} = -0.4\,\mathrm{W\,m^{-2}}$

d. $Q_{\mathrm{S90}} = -0.6\,\mathrm{W\,m^{-2}}$



e. $Q_{\mathrm{S90}} = -0.8\,\mathrm{W\,m^{-2}}$

f. $Q_{\mathrm{S90}} = -1.0\,\mathrm{W\,m^{-2}}$



g. $Q_{\mathrm{S90}} = -1.2\,\mathrm{W\,m^{-2}}$

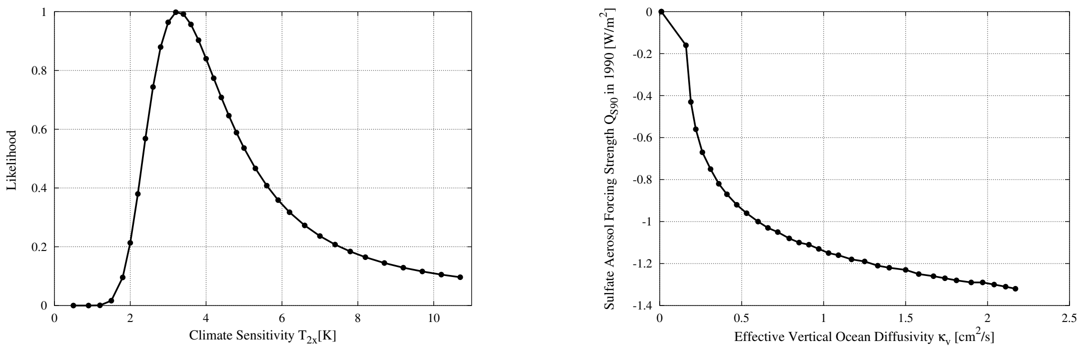
h. $Q_{\mathrm{S90}} = -1.4\,\mathrm{W\,m^{-2}}$

Figure 2.7: Likelihood function for the parameters $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$ generated from the comparison of the model simulations with the 20th century temperature record. Values are given in percentage of the maximum likelihood at the point $T_{2x} = 3.26\,\mathrm{K}$, $\kappa_v = 0.55\,\mathrm{cm^2\,s^{-1}}$, $Q_{\mathrm{S90}} = -0.97\,\mathrm{W\,m^{-2}}$.

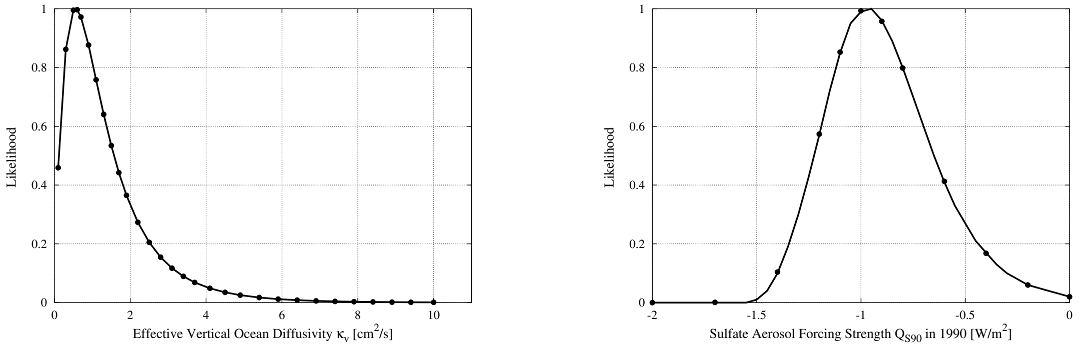
we fixed the volcanic forcing strength, for instance, at a higher value than $\beta^*_{\mathrm{Vol}} = -9.21\,\mathrm{W\,m^{-2}}$, the optimisation routine would increase the ocean heat diffusivity in order to smooth out the emerging volcanic spikes in the residual between modelled and observed temperatures. This would also result in an increase of climate sensitivity at the point of maximum likelihood. It is interesting to note that an opposite effect was identified by Harvey and Kaufmann (2002). Since these authors did not include ocean heat uptake as an additional degree of freedom into the maximisation of the likelihood, a smoothing of volcanic spikes could only be accomplished by a decrease in climate sensitivity.

To further explore the structure of the likelihood function $\overline{\mathcal{L}}(\theta; \hat{T})$, we investigate its projection onto the dimensions of the individual model parameters. Fig. 2.8 shows the 'ridges' of the likelihood function, i.e., the maximum likelihood values that can be obtained for fixed values of $T_{2x}$ (Panel a), $\kappa_v$ (Panel c), and $Q_{\mathrm{S90}}$ (Panel d), when adjusting the remaining parameters appropriately (shown for climate sensitivity in Panel b). It can be seen that the likelihood drops sharply towards low climate sensitivities $T_{2x} \leq 1.5\,\mathrm{K}$ (maximum likelihood: $e^{-4.1}$) and strongly negative sulphate aerosol forcing $Q_{\mathrm{S90}} \leq -1.7\,\mathrm{W\,m^{-2}}$ (maximum likelihood: $e^{-6.9}$). In contrast, very high climate sensitivities $T_{2x} \geq 10\,\mathrm{K}$ (maximum likelihood $e^{-2.2}$) have a non-negligible likelihood, when combined with a sulphate forcing around $-1.3\,\mathrm{W\,m^{-2}}$, and an effective ocean heat diffusivity around $2\,\mathrm{cm^2\,s^{-1}}$ (see Fig. 2.8.b). The low end of the sulphate cooling is better constrained (maximum likelihood for $Q_{\mathrm{S90}} = 0\,\mathrm{W\,m^{-2}}$: $e^{-3.9}$). These results are in qualitative agreement with the findings of previous studies that tried to constrain climate sensitivity and the radiative forcing from aerosols by means of a comparison between model simulations and 20th century temperature data (Andronova and Schlesinger, 2001; Knutti et al., 2002; Forest et al., 2002). They all failed to exclude very high values of climate sensitivity, but could produce stringent constraints on the net cooling effect of aerosols. For ocean heat diffusivity $\kappa_v$, we find that the area of high likelihoods is concentrated at $\kappa_v < 2\,\mathrm{cm^2\,s^{-1}}$. Values above $\kappa_v > 5\,\mathrm{cm^2\,s^{-1}}$ obtain their maximum likelihood for unrealistically high climate sensitivities $T_{2x} > 10\,\mathrm{K}$. If we had constrained climate sensitivity to values $T_{2x} \leq 10\,\mathrm{K}$ then the likelihood would drop faster beyond $\kappa_v = 5\,\mathrm{cm^2\,s^{-1}}$. Hence, the likelihood function constrains $\kappa_v$ at the high end, but not at the low end. Fig. 2.8.b underlines that there exists a strong positive correlation between climate sensitivity and the other two parameters $\kappa_v$ and $Q_{\mathrm{S90}}$. The optimisation procedure utilises the compensating effect of the sulphate aerosols up to $T_{2x} \approx 4 - 5\,\mathrm{K}$, beyond which it relies more and more on the compensating effect of a larger heat uptake of the ocean. This indicates that a restriction of ocean heat diffusivity from above will be necessary to exclude very large values of climate sensitivity (cf. Forest et al., 2002).

We conclude the chapter by comparing the likelihood information with the acceptance regions of the parameter space $\Omega$ that can be obtained from a series of statistical tests on the decorrelated residuals $\eta_S(\theta)$ between modelled and measured SST values, and on the residual $r_{L-S}(\theta)$ between modelled and measured land-sea temperature dif-



a. Maximum likelihood - climate sensitivity b. Corresponding optimal pair $\kappa_v$, $Q_{\mathrm{S90}}$

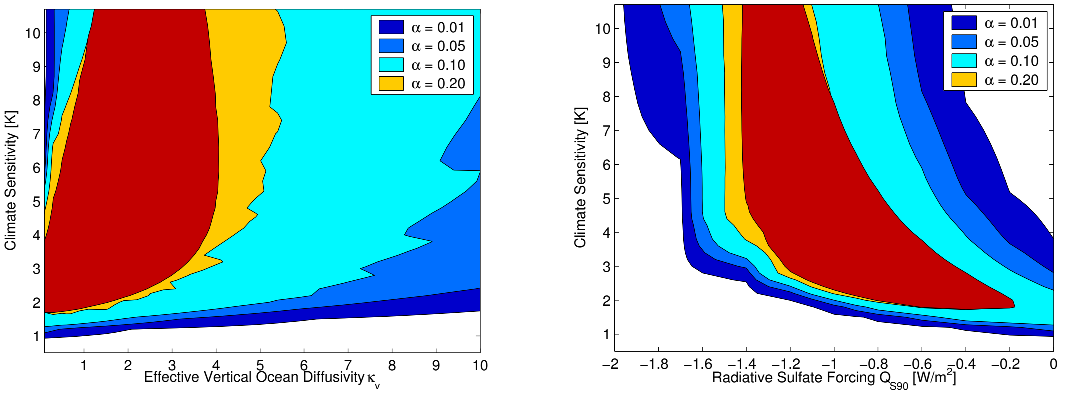


c. Maximum likelihood - ocean heat diffusivity d. Maximum likelihood - sulphate aerosol forcing

Figure 2.8: Maximum likelihood for fixed values of the parameters $T_{2x}$, $\kappa_v$, and $Q_{\mathrm{S90}}$ with respect to the absolute maximum of the likelihood function at $T_{2x} = 3.26\,\mathrm{K}$, $\kappa_v = 0.55\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, $Q_{\mathrm{S90}} = -0.97\,\mathrm{W\,m}^{-2}$ (Panels a, c, d). Panel b shows the adjustment of $\kappa_v$, and $Q_{\mathrm{S90}}$ that tracks the maximum likelihood for fixed values of $T_{2x}$.

ference. Since both residuals are assumed to constitute IID normal random variables (see Equations 2.12 and 2.16), there individual as well as combined square sums needs to be distributed like $\chi^2$. Therefore, we test the square sum $\sum_{i=1}^{n=103} \eta_{S,i}(\theta)^2/\hat{\sigma}_S^2$ for $\chi^2$ with $n-3$ DoF (free parameters $\alpha, \beta_{\mathrm{Vol}}, \beta_{S,\mathrm{SOI}}$), the square sum $\sum_{j=1}^{m=57} r_{L-S,j}(\theta)^2/\hat{\sigma}_{L-S}^2$ for $\chi^2$ with $m-3$ DoF (free parameters $\alpha, \beta_{\mathrm{Vol}}, \beta_{L-S,\mathrm{SOI}}$), and the sum of the two square sums for $\chi^2$ with $n+m-4$ DoF. In addition, we apply the Young-Box-Pierce Portmanteau Test to $\eta_S(\theta)$ (lags up to $k=5$) and $r_{L-S}(\theta)$ (lags up to $k=3$) to test both residuals for white noise. On the basis of these five tests, we can define the $\alpha$-acceptance region $A(\alpha) \subset \Omega$ as the collection of all parameter constellations $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$, for which the corresponding residuals could not be rejected at significance level $\alpha$ by any of the five tests.

Fig. 2.9 shows the projections of the resulting acceptance regions for various significance levels $\alpha$ onto the $\kappa_v$-$T_{2x}$ plane (Panel a) and $Q_{\mathrm{S90}}$-$T_{2x}$ plane (Panel b). It is surprising that only a small portion of the $(T_{2x}, \kappa_v)$ and $(T_{2x}, Q_{\mathrm{S90}})$ pairs can be rejected on the basis of all five tests. We can only exclude climate sensitivities below 1.1 K and a sulphate aerosol cooling effect $|Q_{\mathrm{S90}}| \geq 1.75\,\mathrm{W\,m}^{-2}$ in the year 1990 at a 5% signif-



c. $\kappa_v$-$T_{2x}$ plane

d. $Q_{\mathrm{S90}}$-$T_{2x}$ plane

Figure 2.9: Acceptance regions for $(T_{2x}, \kappa_v)$ pairs (Panel a), and $(T_{2x}, Q_{\mathrm{S90}})$ pairs (Panel b) at various significance levels $\alpha$. For the definition of the acceptance regions see text. Also shown is the regions of $(T_{2x}, \kappa_v)$ and $(T_{2x}, Q_{\mathrm{S90}})$ pairs, respectively, whose likelihood is above 5% of the maximum likelihood (dark red area). Non-convexities in the regions are due to the resolution of the grid, on which the likelihood function was evaluated.

icance level. The result shows how the multi-dimensionality of the parameter space impedes the confinement of the uncertainty about the individual parameters. For most values of $T_{2x}$, $\kappa_v$, and $Q_{\mathrm{S90}}$, one can find a constellation of the other two parameters that improves the simulation of the 20th century temperature record to a degree which suffices to pass the statistical tests. It is also interesting to note that the 5%-cut of the likelihood function (in units of the maximum likelihood) is more restrictive than the 5%-acceptance region, and more similar to the 20%-acceptance region of the combined five tests (see Fig 2.9, dark red area). However, the likelihood function as such conveys no probabilistic meaning, since it collects the $\hat{T}$-values of an uncountable number of conditional probabilities $\rho(\hat{T}|\theta)$ (see Equation 2.12). This highlights the importance of a Bayesian analysis that employs the likelihood information to update an (imprecise) prior probability.

# Chapter 3

# Generating Imprecise Probabilities for Climate Change Assessments

Uncertainty is often separated into two categories: *epistemic uncertainty* and *aleatory uncertainty* (cf., e.g., Walley, 1991, Chap. 1). Aleatory uncertainty arises from variability in observable quantities, which originates from randomness or chaotic behaviour. It is ubiquitous in the climate system as exemplified, e.g., by the natural variability of climatic variables. The concept of aleatory uncertainty fits in nicely with the *frequentist interpretation* of probability. In this context, a probability of $P(A) = 0.5$ for observing an event $A$ means that $A$ will be observed half of the time in an infinite sequence of identical and independent experiments. The natural sciences in the tradition of physics have thought about uncertainty almost exclusively in terms of aleatory uncertainty. This corresponds to the tradition of requiring a number of repeated, identical experiments for testing a hypothesis. But what to do if no repeatable, identical experiments are available to determine the uncertainty about an unknown quantity? In almost every assessment of climate change there exist such quantities that evade a frequentist assessment. Climate sensitivity serves as a paramount example, which we will use throughout this chapter. The uncertainty about climate sensitivity is not aleatory, but epistemic in nature.

Epistemic uncertainty arises from a lack of information about a quantity in question. Aleatory uncertainty may contribute to this lack of information, but in general there are many more sources of epistemic uncertainty. The assessment of climate change, for instance, is confronted with a lack of knowledge about causal relationships that is much larger than the inherent variability of the climate system would imply. The social sciences have always had a type of epistemic uncertainty in mind, where the lack of information is generated by ignorance rather than the existence of random variables. Consequently, they have used a concept of probability which is completely different from the frequentist concept based on aleatory uncertainty. They proposed a *behavioural approach* which identifies probabilities with degrees of belief that are reflected in behaviour (Ramsey, 1931; de Finetti, 1937). In this approach, probabilities are not revealed by observation, but by a disposition to act. Betting situations have served as a paradigm for the behavioural interpretation. If belief can be separated from risk aversion in an idealized betting situation, your degree of belief (probability) about the occurrence of an event $A$ is defined by the maximum amount of utility $0 \leq u \leq 1$ you are willing to trade for a gamble which pays you one unit of utility if $A$ occurs and nothing else.

The interpretation of probability as degree of belief expressed by the actions of an individual gives rise to subjective probabilities. Beliefs will differ among individuals. We have tried to point out in the introduction that probabilities based purely on behavioural dispositions face severe difficulties when applied to the important class of *normative social decision problems*. Clearly, climate policy making belongs to this class as exemplified by the guiding principle to protect our welfare and environment in the face of climate change. To address this challenge, it is not sufficient to base a policy analysis on subjective degrees of belief. Rather, it should be based on the available evidence that society has gathered up to this point in time.

Hence, what is needed is a framework of epistemic uncertainty with a clear evidential foundation and a clear behavioural implication. We have argued in the introduction that the evidential basis about many aspects of climate change is too weak to be described by a single probability measure. As a remedy, statisticians such as Peter Walley (1991) and philosophers like Isaac Levi (1980) have proposed a generalisation of the classical concept of probability to imprecise probability. The main theme of this thesis is motivated by the question in how far the theory of imprecise probability can meet this promise in the integrated assessment of climate change.

The term '*Imprecise Probability*' summarises a variety of mathematical representations, e.g. convex sets of probabilities (Good, 1962; Levi, 1980), lower previsions (Walley, 1991), Choquet capacities (Choquet, 1953), and interval probabilities (Kuznetsov, 1991; Weichselberger, 2000). They have become an object of active research during the last two decades (for introductions see Cozman 1999a and the website of the Society of Imprecise Probabilities and Their Applications at www.sipta.org). Imprecise probabilities of whatever form share the following characteristics: they yield a lower and upper bound on the probability that an event will occur and a lower and upper bound on the expected value of a gamble and random variable, respectively. When lower and upper bound fall onto each other, the classical case of probability theory obtains[1]. When the lower bound is strictly smaller than the upper bound, we say that the belief is imprecise.

In this chapter we will focus on the evidential foundation of imprecise probability models that are chosen *prior* to the consideration of data. As pointed out in the introduction, the assumption of an (imprecise) prior probability is necessary and sufficient to

---

[1] Here and in the following, "classical probability theory" means the axiomatic theories of countably additive and finitely additive probability measures as presented by Kolmogorov (1933) and de Finetti (1937), respectively.

convert information about the likelihood of alternative hypotheses from a comparison with observations (see Chapter 2) into an (imprecise) posterior probability for these hypotheses. In what follows, we will make use of various concepts and relationships from the theory of imprecise probability. Due to the wealth of material that we have to cover on a limited space, we cannot introduce these concepts in the main body of the thesis. However, we have provided a survey of imprecise probability theory in Appendix C. If the reader is unfamiliar with imprecise probabilities, we recommend to consult this survey before proceeding with the remaining part of the chapter. In Section 3.1, we investigate suitable mathematical representations of imprecise probabilities that can match the computational requirements of an integrated assessment of climate change. Section 3.2 discusses what type of imprecise probability model is generated from what type of evidence. As an illustration, we will amalgamate a set of recently published probability estimates for climate sensitivity to various types of imprecise probability. Section 3.3 describes the belief function representation of the particular imprecise probability model that will be used in this analysis. On the basis of this representation, marginal belief functions for climate sensitivity $T_{2x}$, anthropogenic sulphate aerosol forcing $Q_{\mathrm{S90}}$ in 1990, and effective vertical ocean heat diffusivity $\kappa_v$ are constructed in Section 3.4. Section 3.5 combines the information to generate a joint belief function on the entire parameter space $\Omega = \Omega(T_{2x}) \times \Omega(\kappa_v) \times \Omega(Q_{\mathrm{S90}})$.

## 3.1 Tractability of imprecise probability models

Integrated assessments of climate change are usually based on dynamic models, whose complexity ranges from simple box models to fully coupled three-dimensional earth system models. To date, probabilistic analyses of climate change are mainly conducted with energy balance models (e.g., Andronova and Schlesinger, 2001) that have integration times of seconds to minutes, and two- or three-dimensional climate models of intermediate complexity (e.g., Forest et al., 2002) with integration times on the order of hours. Only recently, atmosphere general circulation models coupled to a mixed layer ocean were employed to generate probability estimates for climate sensitivity (Murphy et al., 2004; Stainforth et al., 2005). However, they require the massive use of parallel computing power to produce a sample of even a moderate number of model realisations. Therefore, energy balance models and models of intermediate complexity will continue to play a crucial role in probabilistic assessments of climate change.

Applications of imprecise probabilities will have to cope with the complexity of these models, if they want to be useful for the integrated assessment of climate change. Given the prototypical character of our analysis, we will focus on a low level of complexity as exhibited by one-dimensional energy balance climate models. Despite being very simple climate models, they are fairly complex for imprecise probability standards. In addition to their dynamic nature, they include in general several real-valued uncertain parameters. The associated uncertainty space will be described by an uncountable *universal set* $\Omega$ of *elementary events* (in frequentist language) or *possible states of the world* (in behavioural language). Even if the uncertainty space is simplified by considering a finite partition of $\Omega$, the finite approximation can easily contain thousands or more elements. The challenge is to collect concepts from imprecise probability theory that are mathematical tractable enough to be applied to dynamic problems with such large uncertainty spaces $\Omega$, and still flexible enough to capture different types of uncertainty about climate change.

The survey in Appendix C shows that the body of imprecise probability theory is very rich, since it collects a variety of uncertainty models under one umbrella. These models range from the general concepts of *convex sets of probabilities* (see Definition C.2) and *coherent lower previsions* (see Definition C.4) to special classes like *belief functions* (see Definition C.13). For our application of imprecise probabilities in the integrated assessment of climate change, we basically have two options. The first option is to stick to the general concepts and try to establish methods for solving problems of *natural extension* (see Definition C.5) in climate change and climate policy analysis. In the context of reliability analysis, such an approach has been pushed, e.g. by Utkin and Gurov (1999), Kozine and Filimonov (2000), and Utkin and Kozine (2002). It has the advantage of working with a flexible representation that can incorporate a very wide class of assessments about epistemic uncertainty. The drawback will be that the natural extension is very difficult to compute already for problems with modest complexity, let alone the complexity of integrated assessment models for climate change.

The second option is to give up some generality of the uncertainty representation in return for an increase in mathematical tractability. This involves the search for special classes of imprecise probability that simplify the statistical reasoning considerably, while still retaining enough flexibility to capture at least some aspects of our large uncertainty about climate change. In this thesis we have chosen to go down this road. Our choice does not imply, however, that it would not be worthwhile investigating the other option of approaching a greatly simplified climate change assessment with a very general and flexible concept of imprecise probability. We see both approaches as complementary, and potentially very fruitful for an improved treatment of epistemic uncertainty in climate change assessments. However, they require two different types of analysis which cannot be pursued simultaneously within the scope of this thesis.

A special class of imprecise probabilities will be the more mathematically tractable the more it allows for a simple representation of its information content. Additive probability measures, for instance, are defined on the *field of events* $\mathcal{A}$ (see Appendix C.1 for a definition; an event $A \in \mathcal{A}$ is a subset of $\Omega$), but can be represented by a probability (mass) distribution on the universal set $\Omega$. This allows to process probabilities even in cases of uncountable $\Omega$. We do not expect imprecise probability representations to be of similar simplicity but want to move into this direction as far as possible without losing the gain from accounting for imprecision in the uncertainty assessment. Therefore, we restrict our investigation to *coherent lower probabilities* (see Definition C.7) and their associated *structures* (see Definition C.8) in the following. As pointed out in Appendix C.1, coherent lower probabilities and structures are less general than coherent lower previsions and convex sets of probabilities.

As is the case for additive probabilities, coherent lower probabilities $\underline{P} : \mathcal{A} \rightarrow [0,1]$ are defined on the field of events $\mathcal{A}$, which we will also call *event space* in the following. An assessment of coherent lower probabilities on the entire event space $\mathcal{A}$ becomes intractable for large universal sets $\Omega$. This can be seen easily by recalling that the field of all possible events in a finite universal set $\Omega_n = \{\theta_1, ..., \theta_n\}$ with $n$ atoms, i.e., the *power set* $\mathcal{P}(\Omega_n)$, contains $2^n$ elements. In this situation, we need to look for an efficient representation of the coherent lower probability, from which we can reconstruct lower and upper probability bounds $\underline{P}(A)$ and $\overline{P}(A)$ for arbitrary events $A \in \mathcal{A}$. The set of extreme points (see Definition C.6) of the associated structure $\Gamma(\underline{P})$ is a potential candidate since it contains all additive probabilities that set up the lower and upper probability bounds on the entire event space $\mathcal{A}$. However, it can be shown that a structure $\Gamma(\underline{P})$ for a universal set $\Omega_n$ with $n$ elements can have as much as $n!$ extreme points (see, e.g., Chateauneuf and Jaffray 1989 for the case of structures associated with 2-monotone lower probabilities). Since $n! \gg 2^n$ for universal sets with more than 5 elements, a representation of coherent lower probabilities in terms of the extreme points of their associated structure will be impractical in most cases.

Fortunately, there exists an alternative. Every coherent lower probability $\underline{P} : \mathcal{P}_n \rightarrow [0,1]$ for a *finite* universal set $\Omega_n$ allows for an additive representation on the event space in terms of its *Möbius inverse* $\nu : \mathcal{P}(\Omega_n) \rightarrow \mathbb{R}$ (see Definition C.12):

$$\forall\, A \in \mathcal{P}(\Omega_n) \quad \underline{P}(A) := \sum_{B \subseteq A} \nu(B)\,. \tag{3.1}$$

The Möbius inverse provides an efficient representation of a coherent lower probability if it is *sparse*, i.e., if it contains non-zero values $\nu(A) \neq 0$ only for a limited number $k$ of events $A$ in the power set $\mathcal{P}(\Omega_n)$ ($k \ll 2^n$). The Möbius inverse of a probability measure, for instance, has only non-zero values $\nu(\{\theta_i\}) \geq 0$ on the elementary events $\theta_i \in \Omega_n$. In this case, the Möbius inverse coincides with the probability mass distribution.

Therefore, the concept of the Möbius inverse is particularly interesting for the purpose of our analysis. It carries the potential to greatly simplify the treatment of coherent lower probabilities if problems of statistical inference like updating or propagating information can be rephrased in terms of manipulations of the Möbius inverse. There is one potential caveat, however, which we will need to watch out for during the course of our analysis. If the statistical inference converts a structure into a convex set of probabilities that cannot be fully described by its lower envelope any more, the restriction to coherent lower probabilities and their Möbius inverse will lead to a loss of information. Nevertheless, we have decided to emphasise mathematical tractability and seek for imprecise probability models that allow for an efficient representation in terms of a sparse Möbius inverse. We think that these models are easiest to implement in climate change analyses, and therefore provide a good starting-point for an investigation into the applicability of imprecise probabilities to the integrated assessment of climate change.

## 3.2 Imprecise probability models for various evidence

In the following, we will discuss various imprecise probability models that dwell on different types of information. As an example, we will use the wide range of different uncertainty statements about climate sensitivity in order to see how they can fit into the various imprecise probability models.

### 3.2.1 Interval estimates

Interval estimates of uncertain climate system properties are ubiquitous in the literature about climate change. This is illustrated nicely by the Third Assessment Report of the IPCC. Prominent examples are the radiative forcing strength of various anthropogenic forcing agents (Ramaswamy, 2001, Table 6.11), the global mean temperature change in the period 1990-2100, which is estimated to be in the range $\Delta T = [1.4\,\mathrm{K},\, 5.8\,\mathrm{K}]$, and the estimate of climate sensitivity, $T_{2x} = [1.5\,\mathrm{K},\, 4.5\,\mathrm{K}]$ (Cubasch and Meehl, 2001).

An interval estimate of an uncertain parameter $\theta$ can be given a probabilistic content if its bounds are interpreted as quantiles of the random variable $X_\theta$ capturing the uncertainty about $\theta$. However, this generally requires knowledge of the *parent distribution* of $X_\theta$, in which case the uncertainty could be stated directly in terms of probabilities. In classical statistics, *confidence intervals* are used to quantify a parameter range $[\underline{\theta}, \overline{\theta}]$, which contains the actual parameter $\theta^*$ at some confidence level $\alpha$. In this case, no probabilistic information about the random variable $X_\theta$ is conveyed beyond the confidence value that has to be given to the event $[\underline{\theta}, \overline{\theta}]$. However, none of the two cases is implied by the interval estimates discussed above. They specify a plausible range of values within which no further distinction between more or less likely values is attempted. This characterises the large uncertainty lying at the heart of such interval estimates. As a consequence, the exact meaning of the interval bounds remains opaque. They are usually chosen wide enough to consider values outside their range ‘implausible’, but they rarely delineate such values as ‘impossible’.

Imprecise probability models of interval uncertainty can hardly be specified if the meaning of the interval bounds is unclear. However, if the bounds are meant to exclude all values outside the interval, the interval estimate can be easily transformed into an imprecise probability model. The appropriate candidate is the *vacuous lower probability* model that contains all probability measures whose support is confined to the interval $S = [\underline{\theta}, \overline{\theta}]$. This model is described by a vacuous lower probability $\underline{P}_V : \mathcal{A} \to [0, 1]$ (Walley, 1991, Section 2.9.1) with

$$\underline{P}_V(A) = 0 \quad \text{if} \quad A \not\supseteq S \ , \qquad \text{and} \quad \underline{P}_V(A) = 1 \quad \text{if} \quad A \supseteq S \ . \tag{3.2}$$

The set of extreme points of the corresponding structure $\Gamma(\underline{P}_V)$ contains all Dirac measures with degenerate densities $\delta(\theta - \theta_0)$ on the interval $S$, i.e., $\theta_0 \in [\underline{\theta}, \overline{\theta}]$. The Möbius inverse $\nu_V : \mathcal{A} \to \mathbb{R}$ of $\underline{P}_V$ allows for an even simpler representation of the interval uncertainty. It assigns Möbius mass $\nu_V(S) = 1$ to the interval $S = [\underline{\theta}, \overline{\theta}]$, and no mass to the remaining events. It can easily be seen that this assignment characterises indeed the vacuous lower probability by $\underline{P}_V(A) = \sum_{B \subseteq A} \nu_V(B)$ for all $A \in \mathcal{A}$.

$\underline{P}_V$ constitutes a special case of a *belief function* (see Definition C.13). Belief functions have been introduced by Dempster (1967) and further explored by Shafer (1976). Since they will play a central role in our analysis, we briefly recapitulate the main concepts and properties about belief functions that are discussed in Appendix C. Belief functions constitute *totally monotone* coherent lower probabilities $bel : \mathcal{A} \rightarrow [0, 1]$. Such set functions have favourable properties. In particular, a set function is totally monotone if and only if its Möbius inverse $\nu : \mathcal{A} \rightarrow \mathbb{R}$, exhibits only non-negative entries $\nu(B) \geq 0$ which sum up to unity (Shafer, 1976). Thus, $\nu : \mathcal{A} \rightarrow \mathbb{R}$ can be interpreted as a *basic probability assignment* on $\mathcal{A}$. The sets $A \in \mathcal{A}$ with $\nu(A) > 0$ are called *focal elements*, and the tuple $(\mathcal{E}, \nu) := \{(E_1, \nu(E_1)), ..., (E_n, \nu(E_n))\}$ is called a (finite support) *random set* (see Definition C.14). In this analysis, we will only consider belief functions with a finite number of focal elements. Knowledge of the random set $(\mathcal{E}, \nu)$ suffices to determine *bel* and its conjugate set function *pl*, called *plausibility function* (see Definition C.13), which constitutes a coherent upper probability. It is

$$bel(A) := \sum_{B \subseteq A} \nu(B) \quad = \quad \sum_{i \,|\, E_i \subseteq A} \nu_i \,, \tag{3.3}$$

$$pl(A) := \sum_{B \cap A \neq \emptyset} \nu(A) \quad = \quad \sum_{i \,|\, E_i \cap A \neq \emptyset} \nu_i \,. \tag{3.4}$$

Equations (3.3) and (3.4) will be used frequently in this analysis to calculate lower and upper probabilities from a random set representation of the uncertainty.

### 3.2.2 Convex hull of diverse probability estimates

It has been argued that interval estimates of future climate change contain too little information to guide the implementation of climate protection policies (e.g., Schneider, 2001; Dessai and Hulme, 2003). Subjective probability estimates of experts were proposed as a remedy. In some respect, they constitute the opposite extreme of the interval uncertainty that was critisised. While interval estimates are born out of large uncertainty about many factors that influence the quantity in question, subjective probability estimates need to condense the available information about these factors into a single measure. In this situation, it is not unusual that estimates about the probability distribution of a random variable $X_\theta$ diverge. A good example is given by climate sensitivity. Morgan and Keith (1995) elicited 16 climate experts for their subjective probability distribution about the actual value of climate sensitivity. Although they declared the expert estimates as 'remarkably similar' with the exception of one outlier, the support, quantiles and means of the elicited distributions varied considerably. Removing the outlier, 5% quantiles were estimated in the range $q_{T5} \approx [-1.5\,\mathrm{K}, 1.8\,\mathrm{K}]$, means in the range $\bar{T}_{2x} = [1.9\,\mathrm{K}, 3.6\,\mathrm{K}]$, and 95% quantiles in the range $q_{T95} \approx [4\,\mathrm{K}, 8\,\mathrm{K}]$.

A set of diverse probability estimates $P_1, ..., P_n$ constitutes a special case of imprecise probability. Its *convex hull*, defined by

$$\mathcal{M}(P_1, ..., P_n) = \{\, P \mid \exists\, \lambda_1 \geq 0\,,\ \ldots\,,\ \lambda_n \geq 0 \quad \text{with} \ \sum_{i=1}^{n} \lambda_i = 1\,, \\ \text{so that} \quad \forall\, A \in \mathcal{A} \ \ P(A) = \sum_{i=1}^{n} \lambda_i\, P_i(A) \,\}\,, \tag{3.5}$$

is a closed convex set of probabilities with extreme points $P_1, ..., P_n$. As discussed in Appendix C, it suffices to work with the extreme points in problems of statistical inference like updating the information, projecting it onto prognostic model variables and subjecting it to a decision analysis. Hence, if the set of extreme points is small and specified, each extreme point can be processed separately with the standard methods of classical probability theory. A notable exception is decision analysis, which now has to deal with a *set* of expected utilities or the like. This requires new decision criteria that are fashioned for the general case of imprecise probability. Such criteria are readily available from the decision theoretic and economic literature and will be briefly discussed in Chapter 5.4.

### 3.2.3 An example: Probability estimates for climate sensitivity

In this analysis we will consider a set of probability estimates for climate sensitivity that were published in the recent literature (Andronova and Schlesinger, 2001; Forest et al., 2002; Knutti et al., 2002; Murphy et al., 2004). These estimates improve upon the purely subjective expert opininions elicited by Morgan and Keith (1995) in so far as they involved a comparison of observational data with the output of a climate model. Forest et al. (2002) used the MIT two-dimensional statistical dynamic model (Sokolov and Stone, 1998) to simulate the zonal mean climate for the 1860-1995 period, and to compare it with observational records for upper-air temperature, surface air temperature and interior ocean temperature. They conducted a series of model runs with perturbed values of climate sensitivity and effective ocean heat diffusivity and different assumptions about the net aerosol forcing strength in the 1980s. Hence, they have considered a very similar set of parameters than we do here. Forest et al. (2002) applied the $r^2$-statistics of an $F$ test to evaluate the residual between model output and observations and used it later on to generate a likelihood function on the three-dimensional parameter space. The likelihood was employed to update two different prior probabilities: one uniform prior to all three parameters, and one expert prior (Webster and Sokolov, 2000) that included the averaged expert estimates elicited by Morgan and Keith (1995). This led to two different posterior probabilities on the joint parameter space, whose marginals for climate sensitivity are shown in Fig. 3.1. It can be seen that the primary effect of the expert prior was to dampen the long tail of the posterior probability for climate sensitivity.

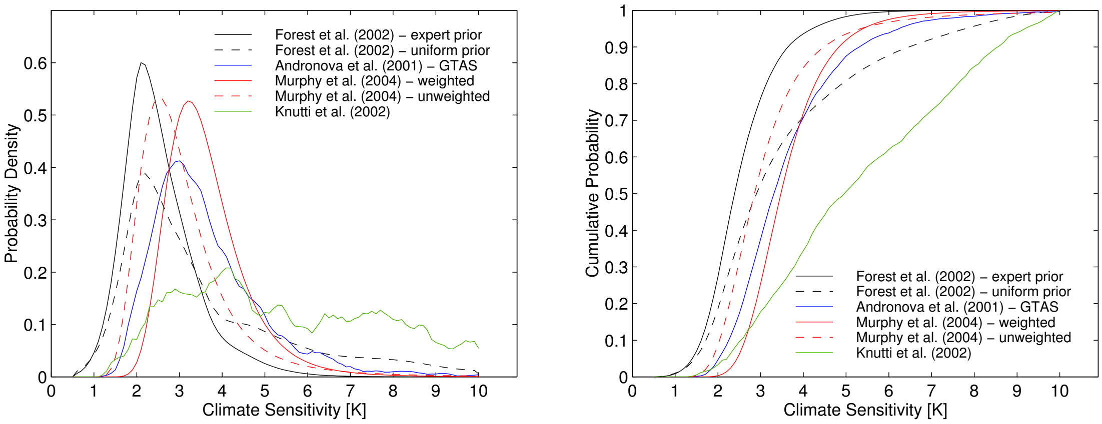


Figure 3.1: Probability density functions (left panel) and cumulative distribution functions (right panel) for climate sensitivity from the literature.

Knutti et al. (2002) also used a zonally averaged climate model of intermediate complexity to assess the range of climate sensitivity and indirect aerosol forcing that is consistent with observed surface air warming and ocean heat uptake in the 20th century. While Forest et al. (2002) based their analysis explicitely on a Bayesian framework, Knutti et al. (2002) introduce a Bayesian flavour rather implicitely. They assumed uniform, normal and log-normal sampling distributions for climate sensitivity and radiative forcing parameters in order to generate an ensemble of model responses and removed those responses that did not accord with observations. While their choice of sampling distributions coincided with the Bayesian choice of a prior, their removal of “inconsistent” model responses resembled a crisp likelihood update that allocates full likelihood to model realisations within a prespecified region around the data and no likelihood to the others. On this basis, they obtained a “data-filtered” ensemble of model realisations, from which they generate histogrammed frequency distributions for climate sensitivity and indirect aerosol forcing. If one takes a Bayesian viewpoint, and accepts the 0-1 likelihood formulation, these frequency distributions can be interpreted as probability distributions for the uncertain parameters. It turned out that the constraints provided by the global mean temperature and ocean heat uptake records had a strong impact on the probability estimate for aerosol forcing, but imposed little structure on the uniform prior distribution for climate sensitivity over the large range of $T_{2x} \in [1\,\mathrm{K}, 10\,\mathrm{K}]$. The empirical distribution function and the associated probability density function (PDF) for the data-filtered ensemble of climate sensitivities are shown in Fig. 3.1.

It is striking that the probability estimate of Knutti et al. (2002) is much less constrained by the comparison with the observational record than the estimate of Forest et al. (2002) for a uniform prior. Some of the difference can be attributed to the fact that Forest et al. (2002) used a more sophisticated likelihood formulation and considered in particular a richer data set against which the model response was evaluated.

However, Knutti et al. (2002) also included uncertainty about the historical radiative forcing trajectory beside the aerosol contribution, which was not accounted for by Forest et al. (2002). Since Knutti et al. (2002) have shown convincingly that this additional uncertainty is an important reason for their inability to further constrain climate sensitivity, the major part of the difference in the probability estimates of Knutti et al. (2002) and Forest et al. (2002) might be attributable to a different appreciation of radiative forcing uncertainty. Therefore, we include the estimate of Knutti et al. (2002) into our analysis as a viable probability representation that cannot be discarded given our current uncertainty about climate change in the 20th century.

The work of Andronova and Schlesinger (2001) does not fit easily in with the likelihood framework, because they dwelled on surrogate data along with the actual temperature observations to determine a probability distribution for climate sensitivity. They used a hemispheric upwelling/diffusion ocean energy balance model (Schlesinger et al., 1997) of slightly higher complexity than the model that we have constructed in Appendix A. The model response to various historical radiative forcing scenarios was compared to the global mean temperature record and inter-hemispheric temperature differences in the 1856-1997 period. Andronova and Schlesinger (2001) considered 16 different forcing models that all included the greenhouse gas contribution (labeled by G), but in- or excluded the forcing contributions of other agents, i.e., sulphate arerosol forcing (A), tropospheric ozone forcing (T), solar forcing (S), and volcanic forcing (V). For each forcing scenario, a maximum likelihood estimation of climate sensitivity and sulphate aerosol forcing (when included) was performed that reproduced the observations best. Andronova and Schlesinger (2001) then used a bootstrapping method on the residual between maximum likelihood response and observation to create a sample of surrogate observations, for which the maximum likelihood estimates of climate sensitivity and sulphate aerosol forcing were recalculated. The resulting samples of maximum likelihood estimates were used to generate an empirical distribution function for each of the 16 forcing scenarios.

If one can assume (near to) complete ignorance about whether the forcing agents A, T, S, or V have contributed significantly to the temperature increase in the past, it is possible to interpret all probability estimates that include and exclude A, T, S, V, respectively, as extreme points of an imprecise probability. This argument clearly does not apply to aerosol forcing A, since most studies suggest that it played a large role in offsetting the warming in the 20th century (see Section 2.2.2). In an earlier work (Kriegler and Held, 2005), we therefore considered only the scenarios GAS, GTAS, GA and GTA, with solar and tropospheric ozone forcing switched on and off. We discarded the scenarios with volcanic forcing (V) altogether, since it was noted by Andronova and Schlesinger (2000) themselves that their energy balance model combined with the high volcanic forcing estimates of Andronova et al. (1999) largely overestimated the volcanic spikes in the global mean temperature record. As a consequence, the volcanic spikes will have been strongly visible in the residual, which might have caused the bootstrapping method to produce spurious results. In this analysis, we will further restrict our selection from Kriegler and Held (2005). First of all, assuming ignorance about the significance of tropospheric ozone is not warranted, because this substance has been identified as the thirdmost important agent of anthropogenic greenhouse gas forcing in the last two decades. Ignorance prevails only in the case of solar forcing. As pointed out in Section 2.2.3, it is still unclear whether or not secular trends in solar forcing have existed over the 20th century. Hence, we could choose to include the probability estimates for GTAS and GTA in our analysis. However, the estimates of Andronova and Schlesinger (2001) without solar forcing allocate considerable probability mass to extraordinaryly high values $T_{2x} > 15\,\mathrm{K}$ of climate sensitivity. This result is not supported by any other model-based analysis so far. Moreover, most experts exclude such high values of climate sensitivity. Therefore, we will only consider the probability estimate for GTAS in this analysis (see Fig. 3.1). The GTAS forcing scenario was also emphasised by Andronova and Schlesinger (2000, 2001) as it seems to produce the best fit to the temperature record.

Recently, the first probability estimate of climate sensitivity from ensemble simulations with an atmospheric general circulation model coupled to a mixed layer ocean became available (Murphy et al., 2004). It revealed the computational limitations that plague ensemble calculations with complex climate models. Due to the high computing costs, the model parameters could be perturbed from their standard values only one at a time yielding a small ensemble of 53 model realisations for the present day and doubled $CO_2$ climates. Murphy et al. (2004) extended the ensemble to multi-parameter variations by assuming that the response patterns for the individual perturbations combine linearly. It is debatable whether the errors inflicted from this assumption do not outweigh the benefit of using a general circulation model instead of less complex climate models that were employed in the other studies. From the extended ensemble, Murphy et al. (2004) generate two probability estimates for climate sensitivity. The unweighted probability estimate was constructed by assuming a uniform prior distribution over the model parameter space without any assessment of the likelihood for reproducing the observed present-day climate. The second estimate took such information into account by weighing each ensemble member with a likelihood that depended on the r.m.s difference between model response pattern and observations for a set of climate variables. The r.m.s difference did not take into account spatial correlations as well as cross-correlations between variables, since the high computational costs inhibited long control simulations for each ensemble member, from which such correlations could have been deduced.

The unweighted and weighted probability estimates of Murphy et al. (2004) are shown in Fig. 3.1. It can be seen that including information about the difference between observed and modelled climate induces a shift to higher values of climate sensitivity. Nevertheless, the probability distribution functions of Murphy et al. (2004) lie in a relatively narrow band together with the estimates of Forest et al. (2002, expert prior) and Andronova and Schlesinger (2001, GTAS). The four estimates exhibit 90% confidence intervals in the range between 1.3 K and 6.3 K. The estimates of Forest et al. (2002, uniform prior) and particularly Knutti et al. (2002) deviate from the other probability distributions by allocating considerable probability mass to very high values of climate sensitivity. The corresponding 90% confidence intervals are $[1.4\,\mathrm{K}, 7.7\,\mathrm{K}]$ and $[2.2\,\mathrm{K}, 9.3\,\mathrm{K}]$, respectively. All estimated probability density functions (PDFs) have the same asymmetric shape with a long tail towards high values of climate sensitivity. The shape can be understood by recalling that climate sensitivity is proportional to the inverse of the climate feedback parameter $\lambda$ (see Equation 2.4). As is apparent in the analysis of Murphy et al. (2004), the typical shape of the PDF for climate sensitivity results from a normal distribution of the climate feedback parameter, which emerges from the superposition of the multiple feedback mechanisms contributing to $\lambda$ (Gregory et al., 2002).

This concludes our discussion of the six probability estimates that we will take into account for the formulation of our prior uncertainty about climate sensitivity. They are drawn from a comprehensive set of studies that derived probability distributions for climate sensitivity from a comparison of model behaviour and observational data. We note that more probability distributions appeared in the literature based on expert assessments (e.g., Tol and De Vos, 1998; Webster and Sokolov, 2000; Wigley and Raper, 2001), or purely on observations (Gregory et al., 2002). Moreover, our assessment constitutes only the snapshot of the current literature. During the IPCC preparation for the Fourth Assessment Report, more probability distributions of climate sensitivity derived from model-data comparisons are scheduled to appear (IPCC, 2004). Very recently, for instance, a frequency distribution for climate sensitivity has been published that represents the first result from a massive parallelisation experiment distributing an ensemble of general circulation model simulations to a large number of personal computers around the world (Stainforth et al., 2005).

### 3.2.4 Distribution bands and p-boxes

We have tried to point out that probability estimates for key determinants of climate change can differ significantly. This can already be demonstrated for the case of climate sensitivity which is relatively well understood compared to factors like, e.g., solar and aerosol forcing, meridional ocean circulation and feedbacks involving the terrestrial and marine biosphere, let alone technological and socio-economic determinants of climate change. What to do in situations where diverging probability estimates have surfaced in the literature or have emerged from expert elicitations? As pointed out in Section 3.2.2, one possible choice is to consider the convex hull of the diverse probability distributions. It allows the analyst to stick to the standard methods of classical probability theory, since it suffices to process the individual distributions separately. However, such a choice will generally be too restrictive for an assessment of the full uncertainty. If probability distributions from the literature diverge, why should all plausible probability estimates that are compatible with the current state of knowledge be restricted to convex combinations of the published distributions? In most cases, this will not be a very reasonable assumption.

An alternative would be to restrict the class of plausible probabilities only by some characteristics of the probability estimates from the literature. A natural choice are the lower and upper bounds on their cumulative *distribution functions* (CDF) (Billingsley, 1995, Section 14, Definition 14.2). Such uncertainty models have been called *distribution bands* in the literature (Basu and DasGupta, 1990). They contain a very large convex set of probabilities. It would emerge from the following two assumptions:

(A) Every probability whose CDF lies between the lower and upper bounds set up by the family of literature estimates is compatible with the current state of information.

(B) Every probability whose CDF is not included in these bounds is not supported by the current state of information.

Assumption (B) is rather reasonable if the distribution band is constructed from all probability estimates that are approved by the scientific community under the current state of knowledge[2]. In this case, every estimate outside the distribution band would be considered implausible until a scientific analysis has demonstrated the contrary. The situation is more difficult with Assumption (A). It basically says that every estimate should be considered plausible which cannot be excluded on the basis of Assumption (B). This is a very conservative statement, because we can imagine Dirac $\delta$-measures with their masses concentrated at point values that are rather implausible but are still fully contained in the distribution band. In this sense, a distribution band constitutes the antipode of a convex hull from diverse probability estimates. It contains too much rather than too few probabilities, and therefore overstates rather than understates the uncertainty.

It can be criticised that probability bounds from the literature, whether specified for CDFs or for other aspects of a probability measure, should be given the status of distinguishing between plausible and implausible probability estimates. Why should future estimates not violate these bounds? This is possible, of course. Assumptions (A) and (B) require to accept the notion that the uncertainty assessment is *conditional* on the given state of information. There might be future estimates that violate the bounds, but these have not been demonstrated yet. Moreover, if the analyst has serious doubts about the ability of the available estimates to paint a complete picture of the current uncertainty, there is nothing that prevents him from widening the bounds by subjective safety margins. In the extreme case of a complete lack of confidence, he would feel inclined to decrease all lower probability bounds to zero, and to increase all upper probability bounds to one. It is obvious, however, that this state of complete ignorance makes any statistical reasoning superfluous. If a statistical inference is seriously considered, then there will be some non-trivial set of probability bounds that can be considered a plausible prior assessment of the uncertainty.

---

[2] We acknowledge that this definition, like any other definition of 'scientific approval', is a difficult notion to work with, since it is both fuzzy and not necessarily related to truth.

It is also important to note that distribution bands can be directly elicited from experts. Expert elicitations often ask for probability assessments on cumulative events. If the experts are allowed to specify intervals on cumulative probability rather than precise values, the resulting uncertainty model is a distribution band. Since a distribution function is the inverse of the *quantile function* (Billingsley, 1995, Section 14, Definition 14.5), distribution bands are also in one-to-one correspondence to *quantile bands*. Hence, if the experts are asked to provide intervals for the plausible range of a set of quantiles, the result is again a distribution band. Therefore, distribution bands indeed constitute an important uncertainty model for the treatment of epistemic uncertainty in climate change assessments. In the following we will investigate this imprecise probability model for the case of *real-valued random variables* $X$, which are generally encountered in climate models and integrated assessment models of climate change. We have published the work that is presented in the next subsection in Kriegler and Held (2005).

### Belief function representation of p-boxes

Let the uncertainty about $X$ be described by a lower bounding function $\underline{F}_X : \mathbb{R} \to [0,1]$ and an upper bounding function $\overline{F}_X : \mathbb{R} \to [0,1]$ for a set of cumulative distribution functions $F_X(x) := P(X \le x)$ on the real line $\mathbb{R}$. Since the uncertain parameter $X$ is real-valued, a natural choice of event space for a probability measure is the *Borel field* $\mathcal{R}$ of $\mathbb{R}$[3] (Billingsley, 1995, Sect. 10 and 14).

The resulting distribution band is defined by the set of Borel-measurable probabilities

$$\Gamma_X(\underline{F},\overline{F}) := \{\, P \,|\, \forall\, x \in \mathbb{R} \quad \underline{F}(x) \le P(-\infty,x] \le \overline{F}(x)\,\}\,. \tag{3.6}$$

$\Gamma_X$ is convex, since for any two probabilities $P,Q \in \Gamma_X$ and $\lambda \in (0,1)$, also $\lambda P + (1-\lambda)Q \in \Gamma_X$.

An important special case of a distribution band is constituted by lower and upper step functions $\underline{SF}$, $\overline{SF} : \mathbb{R} \to \{0, a_1, ..., a_n, 1\}$. The resulting convex set of probabilities $\Gamma_X(\underline{SF},\overline{SF})$ is called a *probability box (p-box)* (Ferson et al., 2002). P-boxes naturally emerge, when the distribution of a continuous random variable on the real line is approximated by lower and upper step functions $\underline{SF} \le F \le \overline{SF}$, where $F \le \overline{SF}$ denotes pointwise domination. Such an approximation was employed to calculate bounds for the convolution of two random variables with unknown dependency (Williamson and Downs, 1990).

Likewise, a continuous distribution band $\Gamma_X(\underline{F},\overline{F})$ can be enclosed by a p-box $\Gamma_X(\underline{SF},\overline{SF}) \supset \Gamma_X(\underline{F},\overline{F})$. The smallest such p-box is bounded by a right-continuous step function $\underline{SF} \le \underline{F}$ from below and a left-continuous step function $\overline{SF} \ge \overline{F}$ from above (see Fig. 3.2). Right- and left-continuity of the bounds are necessary conditions

---

[3] $\mathcal{R}$ is the field of sets generated by arbitrary unions of the countable family of intervals $(a,b]\mathbb{R}$ with rational numbers $a \le b$. The Borel field is the largest $\sigma$-field on the real line, for which the existence of a probability measure can be guaranteed

for the p-box to be the smallest discrete approximation that encompasses the continuous distribution band.

In the following, we focus on p-boxes $\Gamma_X(\underline{SF}, \overline{SF})$ with right- and left-continuous bounding functions. Our goal is to identify a simple uncertainty representation for these p-boxes, which fully captures their information content. Since $\Gamma_X(\underline{SF}, \overline{SF})$ is a convex set of probabilities, we know that its *lower envelope* on the Borel field $\mathcal{R}$, defined by

$$\forall\, A \in \mathcal{R} \quad \underline{P}_X(A) := \inf_{P \in \Gamma_X(\underline{SF}, \overline{SF})} P(A)\,, \tag{3.7}$$

is a coherent lower probability (Walley, 1991, Theorem 3.3.3).

We will show in the following that $\underline{P}_X$ is a belief function (see Definition C.13). The close relationship between p-boxes and belief functions was already noted by Yager (1986). Our new result establishes equivalence between p-boxes and their belief function representation on the real line. Assume that the bounding step functions of the p-box have the form

$$\underline{SF}(x) = \begin{cases} \underline{SF}(x_{*i}) & x_{*i} \leq x < x_{*i+1} \\ 0 & x < x_{*1} \\ 1 & x_{*n} \leq x \end{cases} \quad , \; \overline{SF}(x) = \begin{cases} \overline{SF}(x^*_{j+1}) & x^*_j < x \leq x^*_{j+1} \\ 0 & x \leq x^*_1 \\ 1 & x^*_m < x \end{cases} \quad . \tag{3.8}$$

with $x_{*1} < ... < x_{*n} \in \mathbb{R}$ and $x^*_1 < ... < x^*_m \in \mathbb{R}$. The following algorithm can be used to construct a *random set* $(\mathcal{E}, \nu)$ (see Definition C.14) from $\underline{SF}$ and $\overline{SF}$ (see Fig. 3.2).

**Algorithm 3.1** *1. Initialize indices $k = 1$ (running over the focal elements of the random set to be constructed), $i = 1$ (running over $x_{*i}$), $j = 1$ (running over $x^*_j$). Let $p_k$ denote the cumulative probability already accounted for in step $k$. Assign $p_0 = 0$.*

*2. Construct focal element $E_k = (x^*_j,\, x_{*i}]$.*

*3. If $j = m$, choose arbitrary $x^*_{m+1} > x^*_m$ , thus $\overline{SF}(x^*_{m+1}) = 1$.*

- *(a) $\underline{SF}(x_{*i}) < \overline{SF}(x^*_{j+1})$: $\nu_k = \underline{SF}(x_{*i}) - p_{k-1}$ , $p_k = \underline{SF}(x_{*i})$. Raise indices $k \to k+1$, $i \to i+1$. Return to Step 2.*
- *(b) $\underline{SF}(x_{*i}) > \overline{SF}(x^*_{j+1})$: $\nu_k = \overline{SF}(x^*_{j+1}) - p_{k-1}$ , $p_k = \overline{SF}(x^*_{j+1})$. Raise indices $k \to k+1$, $j \to j+1$. Return to Step 2.*
- *(c) $\underline{SF}(x_{*i}) = \overline{SF}(x^*_{j+1})$: $\nu_k = \overline{SF}(x^*_{j+1}) - p_{k-1}$ .*
  *If $\underline{SF}(x_{*i}) = \overline{SF}(x^*_{j+1}) = 1$, stop.*
  *If $\underline{SF}(x_{*i}) = \overline{SF}(x^*_{j+1}) < 1$, set $p_k = \overline{SF}(x^*_{j+1})$. Raise indices $k \to k+1$, $i \to i+1$, $j \to j+1$. Return to Step 2.*

For each step $k$, it is $x^*_j \leq x_{*i}$ (since $\overline{SF} \geq \underline{SF}$), and $\nu_k > 0$ (since $\overline{SF}$, $\underline{SF}$ monotone increasing). The algorithm will always reach the points $x_{*n}, x^*_{m+1}$ with $\underline{SF}(x_{*n}) = \overline{SF}(x^*_{m+1}) = 1$ and stop.

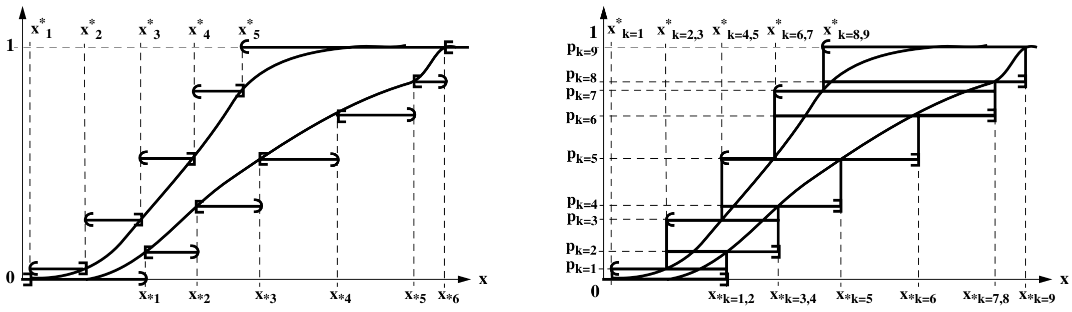


Figure 3.2: Illustration of the p-box approximation of a distribution band (left panel), and the construction of a random set from a p-box by use of Algorithm 3.1 (right panel).

**Lemma 3.1** *Algorithm 3.1 constructs a random set* $(\mathcal{E},\nu)$*, which has the following properties:*

*(I)* $(\mathcal{E},\nu)$ *contains* $q < n+m$ *half-closed intervals* $E_k = (x^*_{j(k)}, x_{*i(k)}]$ *as focal elements.*

*(II)* $(\mathcal{E},\nu)$ *includes no pair of focal elements* $E_k$, $E_l$ *with* $x^*_{j(k)} < x^*_{j(l)} < x_{*i(l)} < x_{*i(k)}$.

*(III)* $\forall\, x \in \mathbb{R}$*, the associated belief and plausibility functions fulfil*
$bel_{\mathcal{E}}(-\infty, x] = \underline{SF}(x)$, $pl_{\mathcal{E}}(-\infty, x] = \overline{SF}(x)$.

The proof of Lemma 3.1 is given in Appendix D. Similar algorithms have been presented in the literature (Ferson et al., 2002; Regan et al., 2004). The main difference is constituted by the fact that Algorithm 3.1 generates half-closed intervals $E_k = (x^*_k, x_{*k}]$, while other formulations usually choose the corresponding closed interval $\tilde{E}_k = [x^*_k, x_{*k}]$. Since for all $k$ $E_k \subset \tilde{E}_k$, we have $bel_{\mathcal{E}} > bel_{\tilde{\mathcal{E}}}$ and $pl_{\mathcal{E}} < pl_{\tilde{\mathcal{E}}}$ for some sets in $\mathcal{R}$. Consider, e.g., the set $(-\infty, x^*_l]$. Due to property (II), it is $pl_{\mathcal{E}}(-\infty, x^*_l] = \sum_{k<l} \nu_k < \sum_{k \leq l} \nu_k = pl_{\tilde{\mathcal{E}}}(-\infty, x^*_l]$. Since $pl_{\mathcal{E}}(-\infty, x^*_l] = \overline{SF}(x^*_l)$, the choice of closed intervals instead of half-closed intervals leads to $pl_{\tilde{\mathcal{E}}}(-\infty, x] > \overline{SF}(x)$ at the points $x = \{x^*_1, ..., x^*_q\}$. For application purposes, this additional imprecision does not matter much, since $pl_{\tilde{\mathcal{E}}}(-\infty, x]$ and $\overline{SF}(x)$ agree almost everywhere on the real line. However, if we want to show that the information content of a p-box $\Gamma_X(\underline{SF}, \overline{SF})$ is completely captured by the random set constructed from Algorithm 3.1, we need to be more precise.

**Theorem 3.1** *Let* $\Gamma_X(\underline{SF}, \overline{SF})$ *be a p-box on the real line bounded by a left-continuous step function* $\overline{SF} : \mathbb{R} \to [0,1]$ *from above and a right-continuous step function* $\underline{SF} : \mathbb{R} \to [0,1]$ *from below (see Definition 3.6).*

*Let* $\underline{P}_X : \mathcal{R} \to [0,1]$ *be the lower envelope of* $\Gamma_X$ *on the Borel field* $\mathcal{R}$ *as defined in Equation (3.7).*

*Let* $(\mathcal{E},\nu)$ *be the random set constructed from* $\underline{SF}$ *and* $\overline{SF}$ *by Algorithm 3.1, and* $bel_{\mathcal{E}} : \mathcal{R} \to [0,1]$ *the associated belief function.*

*Then,* $\forall\, A \in \mathcal{R} \quad bel_{\mathcal{E}}(A) = \underline{P}_X(A)$.

Theorem 3.1 is proved in Appendix D. As a direct consequence of Theorem 3.1, the p-box $\Gamma_X(\underline{SF}, \overline{SF})$ coincides with the convex set of probabilities

$$\Gamma_X(bel_{\mathcal{E}}) := \{\, P \,|\, \forall\, A \in \mathcal{R} \quad bel_{\mathcal{E}}(A) \leq P(A) \,\}$$

that is encompassed by the belief function $bel_{\mathcal{E}}$. This can be seen by noting that every $P_X \in \Gamma_X(bel_{\mathcal{E}})$ has to be an element of $\Gamma_X(\underline{SF}, \overline{SF})$, since $\forall\, x \in \mathbb{R}$ it is $bel(-\infty, x] := \underline{SF}(x) \leq P_X(-\infty, x] \leq pl(-\infty, x] := \overline{SF}(x)$. In turn, every $P_X \in \Gamma_X(\underline{SF}, \overline{SF})$ has to be an element of $\Gamma_X(bel_{\mathcal{E}})$, since $bel_{\mathcal{E}}$ is the lower envelope of $\Gamma_X(\underline{SF}, \overline{SF})$.

Thus, the p-box $\Gamma_X(\underline{SF}, \overline{SF})$ can be represented indeed by a belief function $bel_{\mathcal{E}}$. However, not every $bel_{\mathcal{E}}$ constitutes a p-box. The following corollary of Theorem 3.1 establishes necessary and sufficient conditions for $bel_{\mathcal{E}}$ to be the representation of a p-box.

**Corollary 3.1** *Let* $bel_{\mathcal{E}} : \mathcal{R} \rightarrow [0, 1]$ *be a belief function with (finite support) random set* $(\mathcal{E}, \nu)$*, which defines a right-continuous* $\underline{SF} : \mathbb{R} \rightarrow [0, 1]$ *and left-continuous* $\overline{SF} : \mathbb{R} \rightarrow [0, 1]$ *by* $\underline{SF}(x) := bel(-\infty, x]$ *and* $\overline{SF}(x) := pl(-\infty, x] = 1 - bel(x, +\infty)$ *for all* $x \in \mathbb{R}$*. Then,*

$$\Gamma_X(bel_{\mathcal{E}}) \subseteq \Gamma_X(\underline{SF}, \overline{SF}) \;,$$

*where equality holds if and only if* $(\mathcal{E}, \nu)$ *has Properties (I) and (II) in Lemma 3.1.*

The proof of Corollary 3.1 can be found in Appendix D.

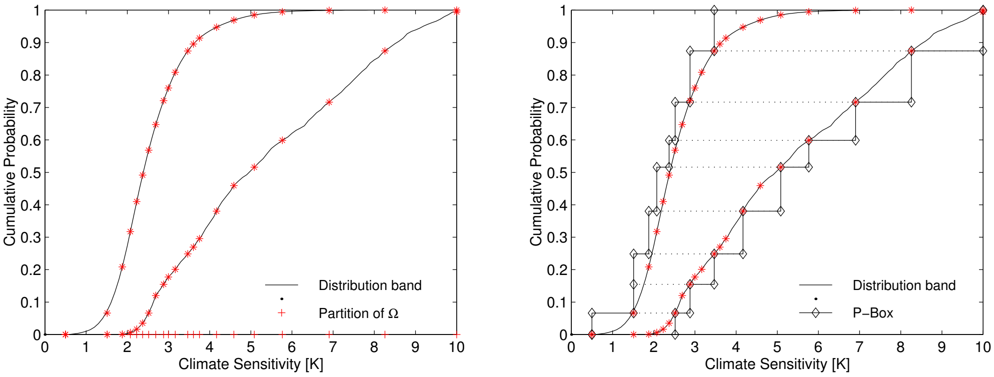
### Application to climate sensitivity

We will use Algorithm 3.1 to construct the random set of a p-box $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF})$ for climate sensitivity which encloses the distribution band of the recently published probability estimates assembled in Section 3.2.3. As shown in Fig. 3.1.b, the upper bound of the distribution band is constituted by a single CDF which was estimated in Forest et al. (2002) under the assumption of an expert prior. The lower bound is set up by two CDFs which stem from the weighted estimate of Murphy et al. (2004) and the filtered ensemble of Knutti et al. (2002), respectively. Before we can apply Algorithm 3.1, we need to approximate the continuous distribution band by a right-continuous step function $\underline{SF}$ from below and a left-continuous step function $\overline{SF}$ from above. Obviously, the choice of step functions is an important one. The p-box approximation always entails a loss of information, since it encompasses more probabilities than the original distribution band.

Following Williamson and Downs (1990), it is common practice to fix the number $n$ of focal elements beforehand, and then define an equiprobable partition of the unit interval for generating best possible bounding functions of the continuous distribution band that jump between the associated probability levels of $0, 1/n, ..., n-1/n, 1$ (Tonon, 2004; Ferson et al., 2002). In this process, the number of focal elements is usually chosen heuristically based on the trade-off between the accurate representation of the continuous band ($n$ large) and computational efficiency ($n$ small). In Kriegler and Held (2005), we have proposed a different approach to enhance the accuracy of the p-box approximation for a fixed number $n$ of focal elements. This can be achieved by adjusting the step height of the approximating step functions to reflect the shape of the continuous bounding functions. We specified a nonlinear program for finding the optimal $n$-partition of the unit probability interval that minimizes the area between lower and upper step functions enclosing the distribution band. The resulting step functions coincide with the p-box approximation that yields the smallest possible increase in area between bounding functions w.r.t to the original distribution band.

In this analysis, we will employ yet another approach. The reason is that we want to amalgamate the p-box representation with another imprecise probability model later on and, more importantly, update the resulting imprecise prior with the likelihood function calculated in Chapter 2. Both steps require the specification of a *finite partition* of the continuous climate model parameter space $\Omega = \mathbb{R}(T_{2x}) \times \mathbb{R}(\kappa_v) \times \mathbb{R}(Q_{\text{S90}})$. Our approach in Kriegler and Held (2005) would automatically generate a partition for climate sensitivity on the basis of which the most accurate p-box approximation of the continuous distribution band with $n$ focal elements can be constructed. However, it will become apparent in Chapter 4 that the updating procedure can introduce large imprecision to the posterior probability estimate, if the choice of partition does not reflect the shape of the likelihood function $\mathcal{L}(\cdot;\hat{y}) : \Omega \to \mathbb{R}_0^+$. Since this effect will outweigh the initial gain in accuracy provided by the approach of Kriegler and Held (2005), we will adapt our choice of partition to the likelihood function rather than to the prior distribution band.

The method to construct the likelihood-adapted partition is presented in detail in Section 3.4.1. Here, we only report the result for the domain $\Omega(T_{2x}) = (0.5\,\text{K}, 10\,\text{K}]$ of climate sensitivity. The partition separates this domain into 20 half-closed intervals, on which we have to construct the p-box approximation of the continuous distribution band. Fig. 3.3.a shows the spacing of the interval bounds on the $T_{2x}$-axis along with the lower and upper cumulative probability that the distribution band exhibits at these points. The p-box approximation is constructed by an upper step function $\overline{SF} > \overline{F}$, and a lower step function $\underline{SF} < \underline{F}$ which jump to higher probability levels at the interval bounds. We have derived the exact form of these step functions heuristically from the condition to approximate the continuous distribution band as close as possible *on the given partition* with a limited number $n \leq 10$ of probability levels. The restriction to 10 or less probability levels limits the number of focal elements that we will have to include in the random set representation of the p-box. It will become evident in the further analysis that a strong limitation of the number of focal elements at this stage is crucial. Since we will add additional information to the p-box model, combine the random sets for $T_{2x}$, $\kappa_v$ and $Q_{\text{S90}}$ to an imprecise probability model on the joint parameter space, and update it later on with the likelihood function, the total number of focal elements will increase by three orders of magnitude.



a. Distribution band

b. P-box approximation

Figure 3.3: Distribution band $\Gamma_{T_{2x}}(\underline{F}, \overline{F})$ and p-box approximation $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF})$ for climate sensitivity. Also shown are the partitioning points of the domain $\Omega(T_{2x}) = (0.5\,\mathrm{K}, 10\,\mathrm{K}]$ (red bars), and the lower and upper cumulative probabilities at these points (red stars). Dotted lines in Panel b indicate the random set $(\mathcal{E}_{T_{2x}}, \nu)$.

| | Focal element | Möbius mass | | Focal element | Möbius mass |
|---|---|---|---|---|---|
| $E_1$ | $(0.50\,\mathrm{K}, 2.52\,\mathrm{K}]$ | 0.066 | $E_6$ | $(2.37\,\mathrm{K}, 5.77\,\mathrm{K}]$ | 0.083 |
| $E_2$ | $(1.51\,\mathrm{K}, 2.88\,\mathrm{K}]$ | 0.089 | $E_7$ | $(2.52\,\mathrm{K}, 6.90\,\mathrm{K}]$ | 0.118 |
| $E_3$ | $(1.51\,\mathrm{K}, 3.47\,\mathrm{K}]$ | 0.094 | $E_8$ | $(2.88\,\mathrm{K}, 8.26\,\mathrm{K}]$ | 0.158 |
| $E_4$ | $(1.88\,\mathrm{K}, 4.17\,\mathrm{K}]$ | 0.132 | $E_9$ | $(3.47\,\mathrm{K}, 10.00\,\mathrm{K}]$ | 0.126 |
| $E_5$ | $(2.08\,\mathrm{K}, 5.09\,\mathrm{K}]$ | 0.135 | | | |

Table 3.1: Random set $(\mathcal{E}_{T_{2x}}, \nu)$ of the p-box $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF})$ for climate sensitivity.

The lower and upper step functions generated in this way are shown in Fig. 3.3.b. They define a p-box $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF})$ which fully encloses the continuous distribution band. We apply Algorithm 3.1 to construct the random set $(\mathcal{E}_{T_{2x}}, \nu)$ from knowledge of $\underline{SF} : \Omega(T_{2x}) \to [0, 1]$ and $\overline{SF} : \Omega(T_{2x}) \to [0, 1]$. As highlighted in Fig. 3.3.b, $(\mathcal{E}_{T_{2x}}, \nu)$ includes the 9 half-closed intervals which are formed between the probability steps of the lower and upper step function. They carry a Möbius mass that is equal to the step height. The resulting random set is tabulated in Table 3.1.

We recall that the p-box $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF})$ includes all probability estimates for climate sensitivity whose distribution functions $F$ fulfil $\underline{SF}(x) \le F(x) \le \overline{SF}(x)$ for all $x \in \mathbb{R}$. It was shown in Theorem 3.1 that the belief function $bel_{T_{2x}} : \mathcal{R} \to [0, 1]$ associated with the random set $(\mathcal{E}_{T_{2x}}, \nu)$ constitutes the lower envelope of these probabilities on the Borel field $\mathcal{R}$. This result is important because it allows us to calculate the minimum and maximum probability that can be obtained from probabilities contained in $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF})$ for any event $A \in \mathcal{R}$ in a simple manner. We demonstrate the calculation for the IPCC estimate of $T_{2x} \in [1.5\,\mathrm{K}, 4.5\,\mathrm{K}]$. The lower and upper probability for the

IPCC estimate are given by (see Definitions 3.3 and 3.4)

$$\begin{aligned} bel_{T_{2x}}(T_{2x} \in [1.5\,\mathrm{K}, 4.5\,\mathrm{K}]) &= \sum_{i \,|\, E_i \subseteq [1.5\,\mathrm{K}, 4.5\,\mathrm{K}]} \nu_i = 0.31 \;, \\ pl_{T_{2x}}(T_{2x} \in [1.5\,\mathrm{K}, 4.5\,\mathrm{K}]) &= \sum_{i \,|\, E_i \cap [1.5\,\mathrm{K}, 4.5\,\mathrm{K}] \neq \emptyset} \nu_i = 1 \;. \end{aligned}$$

By analogous calculations, the probability for $T_{2x} < 1.5\,\mathrm{K}$ is found in the interval $[0, 0.07]$, and for $T_{2x} > 4.5\,\mathrm{K}$ in the interval $[0, 0.62]$. The numbers show that $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF})$ supports the lower bound, but not the upper bound of the IPCC estimate. This reflects the fact that the probability estimates from the literature allocate considerable probability mass to high climate sensitivities $T_{2x} > 4.5\,\mathrm{K}$ (Andronova and Schlesinger, 2001; Knutti et al., 2002; Forest et al., 2002; Murphy et al., 2004).

We note that the lower and upper probability of an interval event $A = [x_*, x^*]$ can be deduced directly from knowledge of the lower bound $\underline{F} : \Omega(T_{2x}) \rightarrow [0, 1]$ and upper bound $\overline{F} : \Omega(T_{2x}) \rightarrow [0, 1]$ of the continuous distribution band by[4]

$$\underline{P}[x_*, x^*] = \max[0, \underline{F}(x^*) - \overline{F}(x_*)] \;, \qquad \overline{P}[x_*, x^*] = \overline{F}(x^*) - \underline{F}(x_*) \;.$$

Hence, we can assess the information loss accompanying the p-box approximation for the case of the IPCC estimate. The continuous distribution band yields a probability interval $[0.37, 0.97]$ for the event $T_{2x} \in [1.5\,\mathrm{K}, 4.5\,\mathrm{K}]$, which is only slightly stricter than the interval generated by the p-box. However, we should not conclude from this single comparison that the overall information loss associated with the p-box approximation is negligible. Since the number of focal elements contained in $(\mathcal{E}_{T_{2x}}, \nu)$ is small, the lower and upper probabilities defined by Equations (3.3) and (3.4) are highly discontinuous on the Borel field $\mathcal{R}$. The close agreement between the continuous distribution band and the p-box approximation can disappear, e.g., when narrowing the bounds of the interval estimate. Consider, for instance, the event $T_{2x} \in [1.52\,\mathrm{K}, 4.5\,\mathrm{K}]$. Since it is very similar to the IPCC estimate, the corresponding interval probability from the continuous distribution band undergoes only a negligible change. However, the lower probability from the p-box approximation decreases abruptly from 0.31 to 0.13, because two focal elements are not contained in $[1.52\,\mathrm{K}, 4.5\,\mathrm{K}]$ that were contained in $[1.5\,\mathrm{K}, 4.5\,\mathrm{K}]$. Therefore, it is important to explore the neighbourhood of an event $A$ for discontinuities, when working with a small number of focal elements. Since the probability intervals from the p-box will always provide a conservative outer bound on the continuous distribution band, the probability interval with the tightest bounds in the neighbourhood of $A$ will yield the best indication of the range of probability values that the distribution band allows on and around $A$. The impact of discontinuities will be reduced, if the number of focal elements is increased. As a rule of thumb, it is often suggested to work with approximately one hundred focal elements on one-dimensional

---

[4] Since the bounds $\underline{F}$ and $\overline{F}$ are continuous, the probability of an interval $A = [x_*, x^*]$ will not depend on whether it is open, closed or half-closed.

continuous spaces (Ferson et al., 2002). In our analysis, we cannot meet this benchmark because we have to cover a three-dimensional parameter space. In this respect, we are suffering from the curse of dimensionality that also plagues probabilistic analyses based on Monte Carlo simulations. In contrast to Monte Carlo methods, however, the lack of computational power to resolve uncertainty on high-dimensional spaces affects imprecise probability analyses by inducing overly conservative probability intervals, which might contain less information than warranted, but which do not contain unwarranted information.

### 3.2.5 Lower and upper probability mass functions

There are many other imprecise probability models besides distribution bands and p-boxes. Any specification of *compatible* lower probability bounds on a collection of Borel fields $\mathcal{B} \subset \mathcal{R}$ defines a set of probabilities

$$\Gamma(\underline{P}_{|\mathcal{B}}) := \{P \,|\, \forall\, B \in \mathcal{B} \;\; P(B) \geq \underline{P}(B)\} \,.$$

In this context compatibility means that the constraints have to be fulfilled by at least one probability measure, i.e., $\Gamma(\underline{P}_{|\mathcal{B}}) \neq \emptyset$. In the case of a p-box, the lower probability bounds were specified on a finite series of cumulative events ($P(-\infty, x_i] \geq \underline{SF}(x_i)$) and their complements ($P(x_i, \infty) \geq 1 - \overline{SF}(x_i)$ with $1 \leq i \leq q$). Although the bounds are specified only on a subset $\mathcal{B}$ of the Borel field, they have implications for the lower probability of events $A \notin \mathcal{B}$. $\underline{P}(A)$ cannot be smaller than the infimum probability mass that is allocated by the probability measures contained in $\Gamma(\underline{P}_{|\mathcal{B}})$ to the event $A$. Hence, the lower envelope of $\Gamma(\underline{P}_{\mathcal{B}})$ extends the implication of the bounds $\underline{P}_{|\mathcal{B}}$ onto the entire Borel field $\mathcal{R}$. It is called *natural extension* of $\underline{P}_{\mathcal{B}}$ (see Definition C.5; Walley 1991, Chapter 3.1). We have shown in Theorem 3.1 that the natural extension of the probability bounds on a finite number of cumulative events and their complements is a belief function.

Beside a series of cumulative events, another natural choice of collection $\mathcal{B}$ for specifying lower probability bounds is constituted by a *partition* $\{A_1, ..., A_k\}$ of the universal set $\Omega = \cup_{j=1}^{n} A_j$, together with the complementary events $\{A_1^c, ..., A_k^c\}$ . In this case, the lower probability on the atoms $A_j$ are specified by a *lower probability mass function* $\underline{p} = \{\underline{p}_1, ..., \underline{p}_k\}$. Compatibility requires $\sum_{j=1}^{k} \underline{p}_j \leq 1$, because otherwise an additive probability measure could not fulfil all constraints simultaneously. The lower probability bounds on the complements $A_j^c$ can be defined in terms of an *upper probability mass function* $\overline{p} = \{\overline{p}_1, ..., \overline{p}_k\}$, since $\underline{P}(A_j^c) = 1 - \overline{p}_j$. Obviously, compatibility requires $\sum_{j=1}^{k} \overline{p}_j \geq 1$ and $\overline{p}_j \geq \underline{p}_j$ for all $1 \leq j \leq k$. The resulting imprecise probability model $\Gamma(\underline{p}, \overline{p})$ has been called an *atomic lower and upper probability (ALUP)* model in the literature (Herron et al., 1997).

Fig. 3.4 shows the ALUP model for climate sensitivity that can be generated from the diverse probability estimates discussed in Section 3.2.3. Here, the choice of partition $\{A_1, ..., A_k\}$ is the same that was used in Section 3.2.4 to construct the p-box

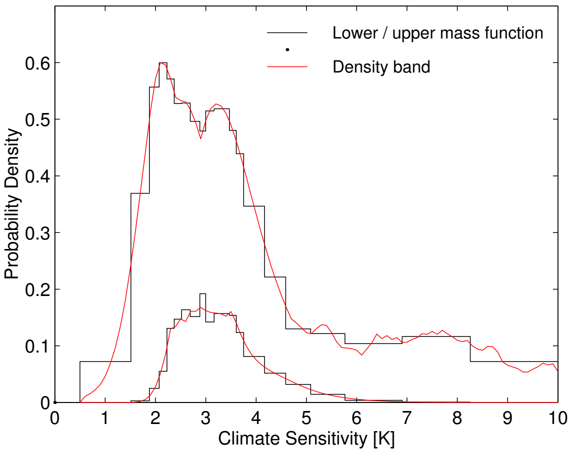


Figure 3.4: Lower and upper probability mass functions on the atoms of $\Omega(T_{2x})$ (converted to densities by division with atom length), and lower and upper density functions for climate sensitivity as set up by the probability estimates from the literature (Section 3.2.3).

approximation of the continuous distribution band. The lower and upper probability masses on the atoms of the partition constitute the minimum and maximum values that can be obtained from one of the six probability estimates of Andronova and Schlesinger (2001), Forest et al. (2002), Knutti et al. (2002) and Murphy et al. (2004). Obviously, the convex hull $co(P_1, ..., P_6)$ of these probability estimates are included in the structure

$$\Gamma(\underline{p}, \overline{p}) = \{P \,|\, \forall\, 1 \leq j \leq k \quad \underline{p}_j \leq P(A_j) \leq \overline{p}_j\}\,, \tag{3.9}$$

but $\Gamma(\underline{p}, \overline{p})$ contains much more probability measures than $co(P_1, ..., P_6)$. As was the case for the distribution band, the ALUP model is less informative than the convex hull. It is also important to note that it describes a different set of probabilities than the p-box constructed in Section 3.2.4. This can be seen by comparing, e.g., the lower probabilities $bel_{T_{2x}}(A_j)$ implied by the p-box with the lower mass function of the ALUP model. Since no focal element of the p-box is contained in a single atom, we find $bel_{T_{2x}}(A_j) = 0$ for all atoms $A_j$. Hence, the p-box contains probability measures which violate the ALUP constraints. It can be shown that the reverse is also true.

The resolution of the partition $\{A_1, ..., A_k\}$ can be increased up to the resolution of the real line $\mathbb{R}$ . In this limit, the lower and upper mass functions become lower and upper probability densities, and the ALUP model turns into a so-called *density band* or *density bounded class*, respectively. This model has been studied in robust Bayesian analysis (Berger, 1993). A density band $\Gamma(\underline{\rho}, \overline{\rho})$ is completely determined by a lower PDF $\underline{\rho} : \Omega \to \mathbb{R}_0^+$ and an upper PDF $\overline{\rho} : \Omega \to \mathbb{R}_0^+$. The density band obtained from the probability estimates for climate sensitivity discussed above is shown in Fig. 3.4. It encompasses a convex set of probabilities that is fully included in the ALUP model, since the lower and upper density functions guarantee the ALUP constraints on the partition $\{A_1, ..., A_k\}$, but provide additional constraints on any refinement of $\{A_1, ..., A_k\}$. Density band and ALUP model relate to each other like distribution band and p-box.

The natural extension of ALUP model and density band onto the entire Borel field $\mathcal{R}$ can be calculated directly from the knowledge of the lower and upper mass functions and densities, respectively (e.g. Wasserman and Kadane, 1990). For arbitrary $A \in \mathcal{R}$,

$$\underline{P}(A) \quad = \quad \max[\int_A \underline{\rho}(x)dx \, , 1 - \int_{A^c} \overline{\rho}(x)dx] \, , \tag{3.10}$$

and an analogous relation holds for lower and upper probability mass functions. The natural extension (3.10) describes the lower envelope of the set of probabilities $\Gamma(\underline{\rho}, \overline{\rho})$. It can be shown that it is a 2-monotone lower probability (see Definition C.10; e.g., Wasserman and Kadane 1990). The same is true for the ALUP model. Therefore, both models allow for a powerful representation in terms of lower probabilities. In particular, the lower and upper expectation of a function $f : \Omega \to \mathbb{R}$ under the ALUP model / density band can be calculated directly from the knowledge of $\underline{P}$ by means of the Choquet integral (see Appendix C). However, $\underline{P}$ is not a belief function in general. Its Möbius inverse can contain negative Möbius masses on some focal elements that cancel out positive probability masses on other focal elements. Hall and Lawry (2004) have provided an algorithm to construct a belief function approximation for the ALUP model. The algorithm basically rescales the Möbius inverse of the lower probability $\underline{P}$ to eliminate negative Möbius masses in exchange for a loss of information. However, this algorithm easily produces Möbius inverses that contain a large number of focal elements.

Since we are looking for imprecise probability models with relatively sparse Möbius inverses, the ALUP model would constitute a difficult choice. The model simplifies greatly if we ignore the information about the upper probability masses $\{\overline{p}_1, ..., \overline{p}_k\}$ that can be allocated to the atoms of the partition. The lower mass function $\underline{p} = \{\underline{p}_1, ..., \underline{p}_k\}$ alone defines a set of probabilities $\Gamma(\underline{p})$, whose lower and upper envelope is given by $\underline{P}(\Omega) = \overline{P}(\Omega) = 1$, and

$$\underline{P}(A) = \sum_{j|A_j \subseteq A} \underline{p}_j \, , \qquad \overline{P}(A) = \sum_{j|A_j \subseteq A} \underline{p}_j + 1 - \sum_{j=1}^{k} \underline{p}_j \quad \text{for } A \subset \Omega \, . \tag{3.11}$$

This imprecise probability model is equivalent to the so-called *ε-contamination model* (e.g., Herron et al., 1997). The latter model is defined in terms of the $\varepsilon$-mixture

$$\mathcal{P}(P, \varepsilon) = (1 - \varepsilon) \, P + \varepsilon \, \mathcal{M}(\Omega)$$

of an additive probability measure $P : \mathcal{R} \to [0, 1]$ with the set of all possible probability measures $\mathcal{M}(\Omega)$. The name stems from the idea that the model describes the set of arbitrary $\varepsilon$-perturbations of the probability $P$. Since $\mathcal{M}(\Omega)$ is characterised by lower probability $\underline{P}(A) = 0$ and upper probability $\overline{P}(A) = 1$ for arbitrary events $A \subset \Omega$, the lower and upper envelope of the $\varepsilon$-contamination model on the partition $\{A_1, ..., A_k\}$ are given by

$$\underline{P}(A) = \sum_{j|A_j \subseteq A} (1-\varepsilon)\, p_j \ , \qquad \overline{P}(A) = \sum_{j|A_j \subseteq A} (1-\varepsilon)\, p_j + \varepsilon \quad \text{for } A \subset \Omega \ , \tag{3.12}$$

where $p = \{p_1, ..., p_k\}$ specifies the probability mass function of $P$ on $\{A_1, ..., A_k\}$. By setting $\underline{p}_j := (1-\varepsilon)\, p_j$ for all $1 \leq j \leq k$, the lower and upper envelope of the $\varepsilon$-contamination model in Equation (3.12) can be identified with the lower and upper envelope of $\Gamma(\underline{p})$ in Equation (3.11). Hence, we will call $\Gamma(\underline{p})$ $\varepsilon$-contamination model in the following. As mentioned above, compatibility requires $\varepsilon := 1 - \sum_{j=1}^{k} \underline{p}_j \geq 0$.

From inspection of Equation (3.11), the Möbius inverse of the lower envelope $\underline{P} : \mathcal{R} \to [0, 1]$ can be easily identified. The focal elements with non-zero Möbius assignments are constituted by the atoms $A_1, ..., A_{\tilde{k}}$ with $\underline{p}_j > 0$ ($\tilde{k} \leq k$), and by the universal set $\Omega$, on which the remaining probability mass $\varepsilon$ has to be allocated. Hence, the random set $(\mathcal{F}, \underline{p})$ of the Möbius inverse is given by

$$(\mathcal{F}, \underline{p}) := \left\{ (A_1, \underline{p}_1), ..., (A_{\tilde{k}}, \underline{p}_{\tilde{k}}), (\Omega, \varepsilon := 1 - \sum_{j=1}^{k} \underline{p}_j) \right\} \ . \tag{3.13}$$

Since all focal elements carry positive probability mass, the associated lower probability is a belief function $bel_{\underline{p}} : \mathcal{R} \to [0, 1]$. By recalling Relation (3.3) between a belief function and its associated random set, it can be seen that $bel_{\underline{p}}$ is equal to the lower envelope of the $\varepsilon$-contamination model as specified in Equation (3.11). Therefore, we conclude without formal proof that the $\varepsilon$-contamination model is represented by a belief function whose random set (3.13) can be derived directly from knowledge of the partition $\{A_1, ..., A_k\}$ and the lower probability mass function $\underline{p} = \{\underline{p}_1, ..., \underline{p}_k\}$.

The $\varepsilon$-contamination model for climate sensitivity as obtained from the probability estimates in the literature is shown in Fig 3.5. By definition, the lower mass function on the partition $\{A_1, ...A_k\}$ must be identical with the ALUP model. However, the upper bound on the probability masses has become much weaker, since we have ignored the information provided by the upper mass function (shown by the dashed red line). The resulting loss of information is aggrevated by the fact that the lower probability mass

|  | Focal element | Möbius mass |  | Focal element | Möbius mass |
|---|---|---|---|---|---|
| $A_4$ | (2.08 K, 2.23 K] | 0.008 | $A_{11}$ | (3.17 K, 3.47 K] | 0.048 |
| $A_5$ | (2.23 K, 2.37 K] | 0.019 | $A_{12}$ | (3.47 K, 3.60 K] | 0.021 |
| $A_6$ | (2.37 K, 2.52 K] | 0.022 | $A_{13}$ | (3.60 K, 3.75 K] | 0.018 |
| $A_7$ | (2.52 K, 2.69 K] | 0.028 | $A_{14}$ | (3.75 K, 4.17 K] | 0.034 |
| $A_8$ | (2.69 K, 2.88 K] | 0.029 | $A_{15}$ | (4.17 K, 4.59 K] | 0.022 |
| $A_9$ | (2.88 K, 3.00 K] | 0.023 | $A_{16}$ | (4.59 K, 5.09 K] | 0.016 |
| $A_{10}$ | (3.00 K, 3.17 K] | 0.024 | $A_{17}$ | (5.09 K, 5.77 K] | 0.010 |
| $\Omega(T_{2x})$ | (0.50 K, 10.0 K] | 0.680 |  |  |  |

Table 3.2: Random set $(\mathcal{F}, \underline{p})$ of the $\varepsilon$-contamination model for climate sensitivity.

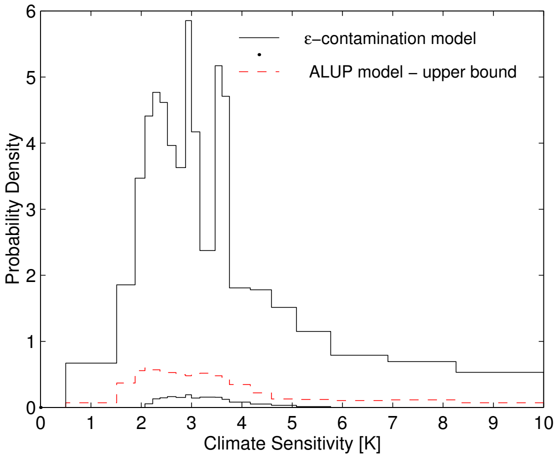


Figure 3.5: Lower and upper probability mass function (converted to densities) of the $\varepsilon$-contamination model for climate sensitivity. For comparison, the upper bound of the ALUP model is also shown.

function $\underline{p} = \{\underline{p}_1, ..., \underline{p}_k\}$ binds less than a third of the total probability mass, so that the 'contamination factor' $\varepsilon$ is large. The spikes in the density plot arise from the fact that $\varepsilon$ can be moved freely within the domain $\Omega(T_{2x})$, and therefore might be fully contained in every atom independently of its size. Therefore, the added amount of density must be inverse proportonial to the atom size. The random set $(\mathcal{F}, \nu)$ of the $\varepsilon$-contamination model is tabulated in Table 3.2. Thereby, we have neglected atoms with lower probability mass $\underline{p}_j < 0.005$ (see Section 3.4.2).

### 3.2.6 Summary of imprecise probability models

So far we have only discussed imprecise probability models that can be represented by a coherent lower probability. However, as pointed out in our survey of imprecise probability theory in Appendix C, not all convex sets of probabilities can be described in this way. A classical example is the *probability ratio model* (Walley, 1991, Chapter 4.6). It is described by lower and upper bounds on the probability (density) ratio of elementary events $\theta \in \Omega$) w.r.t. to a distinguished elementary event $\theta^* \in \Omega$. While we have seen that bounds on probability distributions and probability masses can be captured in terms of coherent lower probabilities, this is not possible for bounds on probability ratios. In this sense, probability ratio information has a different structure than lower probability information. Taking into account both types of information simultaneously would require the use of the general formalism of coherent lower previsions that was laid out by Walley (1991). As discussed in Section 3.1, we have decided to restrict ourselves to coherent lower probabilities for reasons of mathematical tractability. Another possible choice would be, of course, to just dwell on probability ratio information. The lower and upper ratio bounds allow for an efficient representation of the uncertainty,

| Model | $T_{2x} \in [0\,\mathrm{K}, 1.5\,\mathrm{K}]$ | $T_{2x} \in [1.5\,\mathrm{K}, 4.5\,\mathrm{K}]$ | $T_{2x} \in [4.5\,\mathrm{K}, 10.0\,\mathrm{K}]$ |
|---|---|---|---|
| $co(P_1, ..., P_6)$ | [0, 0.07] | [0.43, 0.90] | [0.03, 0.56] |
| $\Gamma(\underline{F}, \overline{F})$ | [0, 0.07] | [0.37, 0.97] | [0.02, 0.56] |
| $\Gamma(\underline{SF}, \overline{SF})$ | [0, 0.07] | [0.31, 1.00] | [0, 0.62] |
| $\Gamma(\underline{\rho}, \overline{\rho})$ | [0, 0.07] | [0.35, 0.97] | [0.03, 0.58] |
| $\Gamma(\underline{p}, \overline{p})$ | [0, 0.07] | [0.28, 0.97] | [0.03, 0.65] |
| $\Gamma(\underline{p})$ | [0, 0.68] | [0.27, 0.98] | [0.02, 0.73] |
| $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$ | [0, 0.07] | [0.31, 0.98] | [0.02, 0.62] |

Table 3.3: Probability intervals for climate sensitivity below, within, and above the range of the IPCC estimate for various imprecise probability models discussed in the text.

and there exists a simple way to update them with a likelihood function. Moreover, probability ratio information can emerge naturally from expert elicitations, since experts might feel more comfortable to specify probabilities relative to a reference event. Therefore, we earmark the application of the probability ratio model to climate change assessments as an important area of future research.

In the preceding sections we have discussed six imprecise probability models: the convex hull $co(P_1, ..., P_n)$ of a finite number of additive probabilities, the continuous distribution band $\Gamma(\underline{F}, \overline{F})$ and the corresponding discrete p-box $\Gamma(\underline{SF}, \overline{SF})$, the density band $\Gamma(\underline{\rho}, \overline{\rho})$ and the corresponding discrete ALUP model $\Gamma(\underline{p}, \overline{p})$, and the $\varepsilon$-contamination model $\Gamma(\underline{p})$. Interval estimates $S \subset \Omega$ constitute the simplest form of a p-box with just one focal element. They are also the simplest form of an $\varepsilon$-contamination model with $\varepsilon = 1$, and the set of all possible probability distributions $\mathcal{M}(S)$ restricted to the interval $S$. We have generated these models for the uncertainty about climate sensitivity from a set of diverse probability estimates in the literature. Our goal was to construct classes of plausible prior probabilities in the light of literature estimates. It is important to note, however, that the construction of these imprecise probability models does not rest on probability estimates from the literature. They can also emerge from expert elicitations, observational constraints and plausibility considerations (see Section 3.4.4).

Table 3.3 compares the different models for climate sensitivity in terms of what they tell us about the imprecise probability of the IPCC estimate $T_{2x} \in [1.5\,\mathrm{K}, 4.5\,\mathrm{K}]$. Note that the last imprecise probability model listed in Table 3.3 has not been discussed so far, but will be introduced in Section 3.3. The partial hierarchy of the imprecise probability models from strong to weak information, i.e.,

$$\begin{array}{lllllll} co(P_1, ..., P_6) & \subset & \Gamma(\underline{F}, \overline{F}) & \subset & \Gamma(\underline{SF}, \overline{SF}) \;, & & \\ co(P_1, ..., P_6) & \subset & \Gamma(\underline{\rho}, \overline{\rho}) & \subset & \Gamma(\underline{p}, \overline{p}) & \subset & \Gamma(\underline{p}) \;, \end{array}$$

is clearly reflected in the probability bounds. Moreover, all probability models except of the $\varepsilon$-contamination model support the lower bound of the IPCC estimate. The situation is different for the upper bound. As already inherent in the probability estimates of Andronova and Schlesinger (2001), Forest et al. (2002), Knutti et al. (2002) and Murphy et al. (2004), considerable weight is given to the possibility of $T_{2x} > 4.5\,\mathrm{K}$. We cannot end this section without noting that there are many more imprecise probability models that we have not discussed here. Two of these many models will be briefly mentioned in Section 3.4.4. An illuminating survey can be found in Walley (1991, Chapter 4).

## 3.3 Joint belief function representation of p-box and $\varepsilon$-contamination model

We have discussed various types of evidence that can be cast into the rich language of imprecise probability models. At this point we need to decide which class, or which combination of classes, to use for our prototypical imprecise probability analysis of climate change. We assert that the convex hull of diverse probability estimates from the literature contains too much information to be a sensible choice. It will be hard to justify that all plausible probability measures to be included in the prior need to be convex combinations of the published estimates. Hence, we are looking for less informative models.

The restriction to lower probabilities implies that we are searching for an uncertainty model described by lower probability bounds on a collection $\mathcal{B}$ of events $B \in \mathcal{R}$. We have seen above that lower probability bounds on a series of cumulative events $(-\infty, x_1] \subset ... \subset (-\infty, x_m]$, $x_1 < ... < x_m \in \mathbb{R}$, and $(y_1, \infty) \supset ... \supset (y_{m'}, \infty)$, $y_1 < ... < y_{m'} \in \mathbb{R}$, generate a p-box $\Gamma(\underline{SF}, \overline{SF})$. However, this class can contain very little information on the probability masses that have to be allocated to the atoms of a partition $\{A_1, ..., A_k\}$ of a subset $S$ of the the real line. In many cases, the resulting lower probability bounds $\underline{P}(A_j)$, $0 \leq j \leq k$ will be considered implausibly low. Moreover, this lack of information can lead to very imprecise results in problems of statistical inference (Kozine and Krymsky, 2003; Kriegler and Held, 2005). Therefore, we believe it to be necessary to consider also a lower probability mass function $\underline{p} = \{\underline{p}_1, ..., \underline{p}_k\}$ on the atoms of a partition $\{A_1, ..., A_k\}$ of a Borel set $S \subseteq \mathbb{R}$. $S$ shall denote the largest possible support of plausible probability estimates for the uncertain quantity in question. As discussed above, such a lower probability mass function defines an $\varepsilon$-contamination model $\Gamma(\underline{p})$.

Hence, a sensible collection of events $B \in \mathcal{B}$ for which lower probability bounds should be specified is given by

$$\mathcal{B} = \{(-\infty, x_1], ..., (-\infty, x_m], (y_1, \infty), ..., , (y_{m'}, \infty), A_1, ..., A_k\} \tag{3.14}$$

for some choice of $x_1 < ... < x_m \in \mathbb{R}$, $y_1 < ... < y_{m'} \in \mathbb{R}$ and partition $\{A_1, ..., A_k\}$ of $S \subseteq \mathbb{R}$. The collection $\mathcal{B}$ can be considered a conservative choice, since it includes only a very limited selection of possible events. As illustrated by the comparison of imprecise probability models in Section 3.2.6, the models become the less informative, the smaller the number of events $B \in \mathcal{B}$. We also note that expert elicitations frequently elicit the expert opinion on (a subset of) the events contained in $\mathcal{B}$.

For all these reasons, we choose to specify the prior uncertainty about the climate model parameters by lower probability bounds on the collection $\mathcal{B}$ of events. The corresponding structure $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$ contains all probability measures $P : \mathcal{R} \to [0,1]$ which obey the constraints

$$\begin{aligned} \underline{SF}(x_i) \leq P(-\infty, x_i] \ , \ 1 \leq i \leq m \ , \qquad 1 - \overline{SF}(y_{i'}) \leq P(y_{i'}, \infty) \ , \ 1 \leq i' \leq m' \ , \\ \underline{p}_j \leq P(A_j) \ , \ 1 \leq j \leq k \ , \end{aligned} \tag{3.15}$$

on the events $B \in \mathcal{B}$. $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$ constitutes the *intersection* of a p-box $\Gamma(\underline{SF}, \overline{SF})$, and an $\varepsilon$-contamination model $\Gamma(\underline{p})$. As discussed in Section 3.2.4 and 3.2.5, the latter two classes can be represented efficiently in terms of the random set associated with their lower envelope. Algorithm 3.1 constructs the random set $(\mathcal{E}, \nu) = \{(E_1, \nu_1), ..., (E_n, \nu_n)\}$ of the lower envelope $bel_{\mathcal{E}}$ of the p-box $\Gamma(\underline{SF}, \overline{SF})$ (see Theorem 3.1). As shown in Lemma 3.1, the focal elements $E_i \in \mathcal{E}$, constitute half-closed intervals of the real line. Likewise, the lower envelope $bel_{\underline{p}} : \mathcal{R} \to [0,1]$ of the $\varepsilon$-contamination model $\Gamma(\underline{p})$ is determined by the the random set $(\mathcal{F}, \underline{p}) = \Big\{(A_1, \underline{p}_1), ..., (A_{\tilde{k}}, \underline{p}_{\tilde{k}}), (S, \varepsilon := 1 - \sum_{j=1}^{k} \underline{p}_j)\Big\}$, where the focal elements $A_j \in \mathcal{F}$ are the atoms with $\underline{p}_j > 0$ ($\tilde{k} \leq k$), and $S = A_1 \cup ... \cup A_k$ is the largest support for probabilities in $\Gamma(\underline{p})$. Both lower envelopes $bel_{\mathcal{E}}$ and $bel_{\underline{p}}$ are belief functions.

In the following, we will try to identify the lower envelope $\underline{P} : \mathcal{R} \to [0,1]$ of the intersection $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$ and to construct the random set $(\mathcal{G}, \nu^*)$ that describes its Möbius inverse. Such an attempt is only meaningful if $\Gamma(\underline{SF}, \overline{SF})$ and $\Gamma(\underline{p})$ are *compatible*, i.e., if their exists at least one probability measure $P : \mathcal{R} \to [0,1]$ that is contained in both sets of probabilities[5]. Such a $P$ will exist if and only if

$$bel_{\mathcal{E}}(A) \leq pl_{\underline{p}}(A) \qquad \Leftrightarrow \qquad bel_{\underline{p}}(A) \leq pl_{\mathcal{E}}(A) \quad \text{for all} \ \ A \in \mathcal{R} \ , \tag{3.16}$$

where the equivalence follows from the conjugacy between lower and upper probability. In the further analysis, we will always assume compatibility of $\Gamma(\underline{SF}, \overline{SF})$ and $\Gamma(\underline{p})$. Compatibility will be guaranteed, if we construct the p-box and the $\varepsilon$-contamination class from the same set of probabilities. In addition, we will have to make the following assumption about the focal elements $E_i \in \mathcal{E}$ and atoms $\{A_1, ..., A_k\}$.

**Assumption 3.1** *The support $S = A_1 \cup ... A_k$ constitutes a half-closed interval of the real line, and all focal elements $E_i \in \mathcal{E}$ are also elements of the power set of the partition $\{A_1, ..., A_k\}$.*

As a direct consequence of this assumption, it is either $A_j \subseteq E_i$ or $A_j \cap E_i = \emptyset$ for arbitrary pairs $1 \leq i \leq n, 1 \leq j \leq k$. Moreover, the union of all focal elements $E = \cup_{i=1}^{n} E_i$ needs to be included in the support $S$ of the partition. The following Lemma establishes a closed form expression for the lower envelope of the intersection $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$ of a p-box and an $\varepsilon$-contamination model.

---

[5] In the language of Walley (1991), the set of probabilities $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$ needs to *avoid sure loss.*

**Lemma 3.2** *Let* $bel_{\mathcal{E}} : \mathcal{R} \to [0,1]$ *be the lower envelope of a p-box* $\Gamma(\underline{SF}, \overline{SF})$ *and* $(\mathcal{E}, \nu)$ *its associated random set. Let* $\{A_1, ..., A_k\}$ *be a partition of the largest possible support* $S \subseteq \mathcal{R}$*, on which the focal elements* $E_i \in \mathcal{E}$ *fulfil Assumption 3.1. Let* $bel_{\underline{p}} : \mathcal{R} \to [0,1]$ *be the lower envelope of an* $\varepsilon$*-contamination model* $\Gamma(\underline{p})$*, with the lower mass function* $\underline{p}$ *defined on this partition, and* $(\mathcal{F}, \nu)$ *its associated random set.*

*Let* $\mathcal{U}$ *be the set of all possible unions* $E = \cup_{i=1}^{s} E_i$*,* $1 \le s \le n$*, of* $E_i \in \mathcal{E}$*, including the focal elements* $E_i$ *themselves. Construct a set function* $\underline{P}^* : \mathcal{R} \to [0,1]$ *from* $bel_{\mathcal{E}}$ *and* $bel_{\underline{p}}$ *by*

$$\begin{aligned} \underline{P}^*(A) &= \max_{E \in \mathcal{U},\, E \subseteq A} \Big( \max[bel_{\mathcal{E}}(E) + bel_{\underline{p}}(A \cap E^c),\, bel_{\underline{p}}(A)] \Big) \qquad (3.17) \\ &= \max_{E \in \mathcal{U},\, E \subseteq A} \left( \max[\sum_{E_i \subseteq E} \nu_i + \sum_{A_j \subseteq (A \cap E^c)} \underline{p}_j,\, \sum_{A_j \subseteq A} \underline{p}_j] \right) \qquad \textit{for all} \quad A \in \mathcal{R}\,. \end{aligned}$$

*Let* $\Gamma(\underline{SF}, \overline{SF})$ *and* $\Gamma(\underline{p})$ *be compatible, i.e.,* $\Gamma(\underline{SF}, \overline{SF}, \underline{p}) = \Gamma(\underline{SF}, \overline{SF}) \cap \Gamma(\underline{p}) \neq \emptyset$*. Let* $\underline{P}_E : \mathcal{R} \to [0,1]$ *be the lower envelope of* $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$*, i.e.,*

$$\underline{P}_E(A) = \inf_{\Gamma(\underline{SF}, \overline{SF}, \underline{p})} P(A) \qquad \textit{for all} \quad A \in \mathcal{R}\,.$$

*Then,* $\underline{P}^*$ *constitutes a lower probability bound of* $\underline{P}_E$*, i.e.,* $\underline{P}_E(A) \ge \underline{P}^*(A)$ *for all* $A \in \mathcal{R}$*.*

The proof of Lemma 3.2 is given in Appendix D. Here, we only note that $bel_{\mathcal{E}}(E) + bel_{\underline{p}}(A \cap E^c) \le 1$ for arbitrary $E \in \mathcal{U}$ and $A \in \mathcal{R}$. Due to compatibility, it is

$$bel_{\mathcal{E}}(E) + bel_{\underline{p}}(A \cap E^c) \le bel_{\mathcal{E}}(E) + pl_{\mathcal{E}}(A \cap E^c) \le pl_{\mathcal{E}}(A) \le 1\,,$$

where we have used the property $pl(A \cup B) \ge bel(A) + pl(B)$ for disjoint sets $A, B$ (Walley, 1991, Property 2.7.4.(e)). Compatibility also requires $bel_{\underline{p}}(A_j) \le pl_{\mathcal{E}}(A_j)$, implying that every atom $A_j \in \{A_1, ..., A_k\}$ with $\underline{p}_j \ge 0$ needs to be contained in at least one focal element $E_i \in \mathcal{E}$.

Lemma 3.2 establishes an analytical expression for the lower probability $\underline{P}^* : \mathcal{R} \to [0,1]$ that confines the convex set of probabilities $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$. We will show in the following that $\underline{P}^*$ constitutes not only a lower bound for $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$, but is in fact identical with its lower envelope $\underline{P}_E$. This can be seen from the Möbius inverse $\nu^* : \mathcal{R} \to [0,1]$ of $\underline{P}^* : \mathcal{R} \to [0,1]$, albeit it is far from obvious. We are also interested in the Möbius inverse from the application point of view, since it allows us to process the information contained in the lower probability $\underline{P}^*$ more efficiently (see Section 3.1). The following algorithm constructs the random set $(\mathcal{G}, \nu^*)$ that characterizes the Möbius inverse $\nu^*$ of $P^*$.

**Algorithm 3.2** *1. Let* $1 \le i \le n$ *denote the running index over the* $n$ *focal elements* $E_i$ *contained in* $\mathcal{E}$*. Let* $1 \le j \le k$ *be the running index over the* $k$ *atoms* $A_j$ *of* *the partition $\{A_1, ..., A_k\}$. The algorithm is designed under the premise that the focal elements $E_i \in \mathcal{E}$ fulfil Assumption 3.1. Let $(\mathcal{G}, \nu^*)$ denote the random set that will be constructed with this algorithm.*

*Consider the $\tilde{k} \leq k$ atoms $A_j \in \mathcal{F}$ with $\underline{p}_j > 0$. Include all these focal elements $A_j \in \mathcal{F}$ with Möbius assignment*

$$\nu_j^* = \underline{p}_j \ , \quad 1 \leq j \leq \tilde{k} \qquad \textit{in} \quad (\mathcal{G}, \nu^*) \ .$$

*Label the collection $\{(A_1, \nu_1^*), ..., (A_{\tilde{k}}, \nu_{\tilde{k}}^*)\}$ with $(\mathcal{G}, \nu^*)^0$.*

2. *Construct all unions of the focal elements $E_1, ..., E_n \in \mathcal{E}$. This will generate a family of sets $\mathcal{U}$,*

$$\begin{array}{rrl}
 & E_i \ , & 1 \leq i \leq n \ , \\
\textit{for} \quad i \neq j \ , & E_{i(2)} = E_i \cup E_j \ , & 1 \leq i(2) \leq \dfrac{n!}{n-2!2!} \ , \\
\textit{for} \quad i \neq j, j \neq l, i \neq l \ , & E_{i(3)} = E_i \cup E_j \cup E_l \ , & 1 \leq i(3) \leq \dfrac{n!}{n-3!3!} \ , \\
... \ , & ... \ , & ... \ , \\
 & E_{i(n)} = E_1 \cup ... \cup E_n \ , & i(n) = 1 \ ,
\end{array}$$

*which contains at most $2^n - 1$ elements. The number can be smaller, if some unions are identical.*

3. *For each $E \in \mathcal{U}$, calculate*

$$\underline{P}^*(E) = \max_{\tilde{E} \in \mathcal{U}, \, \tilde{E} \subseteq E} \left( \max[\sum_{E_i \subseteq \tilde{E}} \nu_i + \sum_{A_j \subseteq \left(E \cap \tilde{E}^c\right)} \underline{p}_j, \sum_{A_j \subseteq E} \underline{p}_j] \right) \ .$$

4. *Construct a hierarchy of focal elements $E_i$ and their unions as follows. For each $E \in \mathcal{U}$, count the number $h(E)$ of sets $E' \in \mathcal{U}$ with $E' \subseteq E$. Since each set contains itself, it is $h(E) \geq 1$. Let $h_1 < ... < h_t$ be the ordered set of numbers $h(E)$ that are obtained in this way. For each level $1 \leq s \leq t$ of the hierarchy, collect the sets $E \in \mathcal{U}$ that contain $h_s$ sets $E' \in \mathcal{U}$. We will call these sets "level-s sets" in the following, and label them $E_{i(s)}^s$ with running index $1 \leq i(s) \leq n(s)$. The level-1 sets at the bottom of the hierarchy are constituted by the focal elements $E_i \in \mathcal{E}$ which do not contain other focal elements.*

5. *In the following loop, more focal elements will be added successively to the collection $(\mathcal{G}, \nu^*)^0$ with increasing level $s$ of the hierarchy imposed on the sets $E \in \mathcal{U}$. Initialize the loop by setting $s = 1$.*

6. *If* $n(s) > 0$*, calculate for* $1 \leq i(s) \leq n(s)$
$$\nu^*_{i(s)} = P^*(E^s_{i(s)}) - \sum_{l \,\middle|\, \substack{G_l \in \mathcal{G}^{s-1} \\ G_l \subseteq E^s_{i(s)}}} \nu_l \,.$$
*Add all sets* $E^s_{i(s)}$ *with Möbius mass* $\nu^*_{i(s)} \neq 0$ *as focal elements to the collection* $(\mathcal{G}, \nu^*)^{s-1}$*. Label the expanded collection with* $(\mathcal{G}, \nu^*)^s$*.*
7. *If* $s = t$ *stop the construction of the random set, and set* $(\mathcal{G}, \nu^*) := (\mathcal{G}, \nu^*)^t$*. If* $s < t$*, set* $s = s + 1$*, and repeat Step 5.*

Algorithm 3.2 constructs a random set $(\mathcal{G}, \nu^*)$, which can contain as much as $k + 2^n - 1$ focal elements. The following Lemma establishes that $(\mathcal{G}, \nu^*)$ describes indeed the Möbius inverse of $\underline{P}^* : \mathcal{R} \to [0, 1]$.

**Lemma 3.3** *Let* $\underline{P}^* : \mathcal{R} \to [0, 1]$ *be the lower probability constructed from* $bel_{\mathcal{E}}$ *and* $bel_{\underline{p}}$ *by means of Definition (3.17). Let the focal elements* $E_1, ..., E_n \in \mathcal{E}$ *of the belief function* $bel_{\mathcal{E}}$ *fulfil Assumption 3.1.*

*Let* $(\mathcal{G}, \nu^*)$ *be the random set constructed from Algorithm 3.2. Then,* $(\mathcal{G}, \nu^*)$ *describes the Möbius inverse of* $\underline{P}^*$*, i.e.,*
$$\underline{P}^*(A) = \sum_{l | G_l \subseteq A} \nu^*_l \qquad \textit{for all} \quad A \in \mathcal{R} \,.$$

The proof of Lemma 3.3 is given in Appendix D. As mentioned above, compatibility requires that all atoms $A_j$ contained in $(\mathcal{G}, \nu^*)$ need to be subsets of $E^\dagger = E_1 \cup ... \cup E_n$. Therefore,
$$\sum_{l | G_l \in \mathcal{G}} \nu^*_l = \sum_{l | G_l \subseteq E^\dagger} \nu^*_l = \underline{P}^*(E^\dagger) = 1 \,,$$
i.e., the Möbius masses in $(\mathcal{G}, \nu^*)$ add to unity.

We will show in the following that $(\mathcal{G}, \nu^*)$ exhibits very useful properties for application purposes.

**Lemma 3.4** *Let* $(\mathcal{G}, \nu^*)$ *be constructed from the random sets* $(\mathcal{E}, \nu)$ *of the p-box and* $(\mathcal{F}, \underline{p}_j)$ *of the* $\varepsilon$*-contamination model by means of Algorithm 3.2. Let the focal elements* $E_i \in \mathcal{E}$ *fulfil Assumption 3.1.*

*Consider an arbitrary level-s set* $E^s_{i(s)} \in \mathcal{U}$*. If there exists two other sets* $E^{s'}_*$, $E^{s''}_{**} \in \mathcal{U}$ *with* $E^s_{i(s)} = E^{s'}_* \cup E^{s''}_{**}$ *and* $E^{s'}_* \cap E^{s''}_{**} = \emptyset$*, then it is* $\nu^*_{i(s)} = 0$*.*

Lemma 3.4 is proved in Appendix D. Since all focal elements $E_i$ in $\mathcal{E}$ are half-closed intervals (see Lemma 3.1), Lemma 3.4 shows that all their unions with non-zero Möbius assignment $\nu^*$ need to be half-closed intervals as well. This is an important result for application purposes because it excludes the emergence of non-convex focal elements in the random set $(\mathcal{G}, \nu^*)$. Such non-convexities could cause severe difficulties in optimisation problems that need to be solved, e.g., when projecting imprecise probability onto prognostic model variables (see Chapter 5) or when conducting a decision analysis under imprecise probability. Moreover, Lemma 3.4 assures that the number of focal elements in $\mathcal{G}$ will be substantially lower than $k + 2^n - 1$, since all non-convex level-$s$ sets $E^s_{i(s)}$ are not contained in $\mathcal{G}$. This assessment remains true even in the special case where most focal elements intersect each other, and the number of non-convex level-$s$ sets $E^s_{i(s)}$ will be small (see, e.g., the random set for climate sensitivity in Fig. 3.3). In this particular case, most unions will be identical, and therefore, merge into a small number $\ll 2^n$ of convex focal elements contained in $\mathcal{G}$. We conclude that the Möbius inverse of $\underline{P}^*$, although richer than the Möbius inverse of a p-box, remains *sparse* when compared to the cardinality of the power set $\mathcal{P}(A_1, ..., A_k)$ or the number $2^n - 1$ of all possible unions of focal elements $E_i \in \mathcal{E}$.

**Lemma 3.5** *Let the assumptions of Lemma 3.4 be fulfilled. For each level-s set* $E^s_{i(s)} \in \mathcal{U}$, *it is* $\nu^*_{i(s)} \geq 0$.

Lemma 3.5 is proved in Appendix D. It provides the last cornerstone to establish the following Theorem.

**Theorem 3.2** *Let* $bel_{\mathcal{E}} : \mathcal{R} \to [0, 1]$ *be the lower envelope of* $\Gamma(\underline{SF}, \overline{SF})$, *and* $(\mathcal{E}, \nu)$ *its associated random set. Let* $\{A_1, ..., A_k\}$ *be a partition, on which the focal elements* $E_i \in \mathcal{E}$ *fulfil Assumption 3.1.*

*Let* $bel_{\underline{p}} : \mathcal{R} \to [0, 1]$ *be the lower envelope of* $\Gamma(\underline{p})$, *with the lower mass function* $\underline{p}$ *defined on the partition* $\{A_1, ..., A_k\}$ *and* $(\mathcal{F}, \nu)$ *its associated random set.*

*Let* $\underline{P}^* : \mathcal{R} \to [0, 1]$ *be the lower probability given by Definition (3.17).*

*If the p-box* $\Gamma(\underline{SF}, \overline{SF})$ *and the* $\varepsilon$*-contamination model* $\Gamma(\underline{p})$ *are compatible (see Inequalities 3.16), then* $\underline{P}^*$ *constitutes the lower envelope of* $\Gamma(\underline{SF}, \overline{SF}, \underline{p}) = \Gamma(\underline{SF}, \overline{SF}) \cap \Gamma(\underline{p})$. *Moreover,* $\underline{P}^*$ *is a belief function.*

The proof of Theorem 3.2 can be found in Appendix D. We conclude that Algorithm 3.2 provides us with a method to establish a belief function representation for the intersection of a p-box $\Gamma(\underline{SF}, \overline{SF})$ and an $\varepsilon$-contamination model $\Gamma(\underline{p})$. This constitutes an important achievement, since it allows us to combine lower probability constraints on cumulative events with lower probability masses on the atoms of a partition $\{A_1, ..., A_k\}$. In many instances of statistical reasoning with imprecise probabilities, both types of information are needed to avoid overly large imprecision in the inference process. Moreover, the random set $(\mathcal{G}, \nu^*)$ shares the nice properties of the random set $(\mathcal{E}, \nu)$ of the p-box to some extent. Although richer in the variety of elements, it contains only convex focal elements, and their number will still be far below the cardinality $2^k$ of the partition or the number $2^n - 1$ of unions of focal elements $E_i \in \mathcal{E}$. Therefore, $(\mathcal{G}, \nu^*)$ provides a suitable model from the application perspective. In the following, we will use Algorithm 3.2 to construct the random sets $(\mathcal{G}, \nu^*)$ for the uncertain climate model parameters $T_{2x}$, $Q_{\mathrm{S90}}$ and $\kappa_v$.

## 3.4 Belief functions for the climate model parameters

In order to quantify the uncertainty about $T_{2x}$, $Q_{\mathrm{S90}}$ and $\kappa_v$ we need to specify a universal set $\Omega \subset \mathbb{R}(T_{2x}) \times \mathbb{R}(\kappa_v) \times \mathbb{R}(Q_{\mathrm{S90}})$ of *possible values* for the climate model parameters. We choose

$$\begin{aligned}\Omega = (\underline{T}_{2x} = 0.5\,\mathrm{K}, \overline{T}_{2x} = 10\,\mathrm{K}] \times (\underline{\kappa}_v = 0.05\,\mathrm{cm}^2\,\mathrm{s}^{-1}, \overline{\kappa}_v = 5\,\mathrm{cm}^2\,\mathrm{s}^{-1}] \\ \times (\underline{Q}_{\mathrm{S90}} = -1.9\,\mathrm{W\,m}^{-2}, \overline{Q}_{\mathrm{S90}} = 0\,\mathrm{W\,m}^{-2}]\,. \end{aligned} \tag{3.18}$$

For the case of climate sensitivity, a similar choice was made by Knutti et al. (2002) and Forest et al. (2002). Other studies without a predefined range of climate sensitivity can reject values below 0.5 K with high confidence (Gregory et al., 2002; Murphy et al., 2004; Schneider von Deimling et al., 2005). Moreover, such small climate sensitivities are assigned a very low likelihood $< e^{-40}\,\mathcal{L}_{\max}$ from the comparison of the historical temperature record and the response of our energy balance model in Chapter 2. The situation is less clear for the upper bound $\overline{T}_{2x} = 10\,\mathrm{K}$. Incidents of climate sensitivities above 10 K have surfaced in perturbed parameter experiments with atmosphere general circulation models coupled to a mixed layer ocean (Stainforth et al., 2005). Moreover, such cases cannot be excluded by model-data comparisons with high confidence (for an extreme example see Andronova and Schlesinger 2001). This is also true for our analysis in Chapter 2, where we have calculated a likelihood $\approx e^{-2.2}\,\mathcal{L}_{\max}$ for $T_{2x} = 10\,\mathrm{K}$. However, as mentioned in Section 3.2.3, most experts view such high values of climate sensitvity as unrealistic, and we therefore exclude them from our analysis.

The domain for total sulphate aerosol forcing $Q_{\mathrm{S90}}$ is bounded from above by $\overline{Q}_{\mathrm{S90}} = 0\,\mathrm{W\,m}^{-2}$. A positive heating contribution would contradict the basic properties of sulphate aerosols which mainly reflect the incoming solar radiation. This is reflected in the Third Assessment Report of the IPCC which establishes a negative upper bound of $Q_{\mathrm{S90}} < -0.2\,\mathrm{W\,m}^{-2}$ (Ramaswamy, 2001). However, we also want to account for the possibility that the net heating effect of black carbon was underestimated in the specification of the historical radiative forcing trajectory in Section 2.2.2. To counterbalance an undervaluation of black carbon, the cooling contribution of sulphate aerosols would need to be reduced. We allow for such a compensation up to the point of a vanishing sulphate cooling. Given our estimate of $-0.15\,\mathrm{W\,m}^{-2}$ for the net forcing from carbonaceous aerosol species in the 1990s, this provides an upper bound for the total aerosol forcing of $Q_{Aer} < -0.15\,\mathrm{W\,m}^{-2}$. Forest et al. (2002) exclude a smaller cooling contribution from aerosols with 98% confidence. The model-data comparison in Chapter 2 yields a likelihood of $\approx e^{-3.9}\,\mathcal{L}_{\max}$ for $Q_{\mathrm{S90}} = 0\,\mathrm{W\,m}^{-2}$.

Our choice of lower bound $\underline{Q}_{\mathrm{S90}} = -1.9\,\mathrm{W\,m}^{-2}$ is not supported by the IPCC assessment of a combined direct and indirect sulphate aerosol forcing between $-2.8\,\mathrm{W\,m}^{-2}$ and $-0.2\,\mathrm{W\,m}^{-2}$ (Ramaswamy, 2001). A recent literature survey has even reported aerosol model simulations that allow for a global cooling effect up to $4\,\mathrm{W\,m}^{-2}$ (Anderson et al., 2003). However, our analysis in Chapter 2 has identified only a very low likelihood for a sulphate aerosol cooling larger than $\underline{Q}_{\mathrm{S90}} = -1.9\,\mathrm{W\,m^{-2}}$ ($< e^{-22}\,\mathcal{L}_{\max}$). Since every prior assessment of $Q_{\mathrm{S90}} < -1.9\,\mathrm{W\,m^{-2}}$ would be nullified by such a small factor in the updating procedure, it makes little sense to include such prior estimates deliberately in our analysis. The probability estimates from the literature that will be considered for the construction of the prior imprecise probability model in Section 3.4.3 do not support values of $Q_{\mathrm{S90}} < -1.9\,\mathrm{W\,m^{-2}}$.

Plausible values for the effective vertical ocean heat diffusivity $\kappa_v$ are constrained by observations of the ocean heat uptake in the second half of the 20th century (Levitus et al., 2000). It will be shown in Section 3.4.4 that these observations do not support ocean heat diffusivities above $5\,\mathrm{cm^2\,s^{-1}}$ in the particular modelling framework derived in Appendix A. The lower bound for $\kappa_v$ will be motivated on a different count in that section. Since we have considered only surface temperature information for the estimation of the likelihood function in Chapter 2, the observational constraints from ocean heat uptake measurements are not reflected in the likelihood values for the lower bound $\underline{\kappa}_v = 0.05\,\mathrm{cm^2\,s^{-1}}$ ($\approx e^{-1.1}\,\mathcal{L}_{\max}$) and upper bound $\overline{\kappa}_v = 5\,\mathrm{cm^2\,s^{-1}}$ ($\approx e^{-3.7}\,\mathcal{L}_{\max}$).

### 3.4.1 Likelihood-adapted partition of the parameter space

Since the uncertain parameter space $\Omega$ is continuous, we need to construct a finite partition $\{A_1, ..., A_k\}$ of $\Omega = \cup_{j=1}^{k} A_j$, on which the random sets for the p-box, the $\varepsilon$-contamination model, and the intersection of the two models can be specified. As was discussed in Section 3.2.3, the choice of partition is usually motivated by the objective to approximate a continuous imprecise probability model as accurately as possible. In this analysis, however, we are concerned with the accuracy of the imprecise *posterior* probability approximation that emanates from updating a discretised prior model in place of its continuous counterpart. The updating method presented in Chapter 4 will require to discretise the likelihood function as well. Our experiments have shown that this step is key to a meaningful approximation of the imprecise posterior probability that would result in the continuous case. As will be discussed in Chapter 4, the updating procedure can introduce large artificial imprecision to the posterior probability estimate, if the choice of partition does not reflect the shape of the continuous likelihood function $\mathcal{L}(\cdot;\hat{y}) : \Omega \to \mathbb{R}_0^+$. Therefore, we will search for a partition $\{A_1, ..., A_k\}$ that is adapted to the likelihood function.

It can be seen from Fig. 2.7 that the likelihood function incorporates a strong positive correlation between climate sensitivity and ocean heat diffusivity and a strong positive correlation between climate sensitivity and sulphate aerosol cooling. Due to these correlations, a partitioning of $\Omega$ by hyperplanes that are orthogonal to the $T_{2x}$, $\kappa_v$, and $Q_{\mathrm{S90}}$-axes cannot provide the best adaptation to the likelihood function. However, if we adopted another choice of partitioning hyperplanes, we would lose the possibility of constructing the imprecise prior model for each model parameter separately, because the atoms $A_j$ would be no longer constituted by simple product sets $A_{T_{2x}} \times A_{\kappa_v} \times A_{Q_{\mathrm{S90}}} \subset \Omega$. This would complicate the construction of the imprecise prior probability greatly. Firstly, it is more difficult to specify and represent a continuous distribution or density band on multi-dimensional spaces. Secondly, and more importantly, Algorithm 3.1 for the construction of a random set $(\mathcal{E}, \nu)$ of a p-box $\Gamma(\underline{SF}, \overline{SF})$, as well as the statements provided in Lemma 3.1 and Theorem 3.1, are restricted to the real line $\mathbb{R}$. We speculate that they can be extended to higher dimensions, but this will require further investigations. Within the scope of this prototypical analysis, we will not attempt to generalise the formalism to higher dimensions or explore more sophisticated grid methods that might help to circumvent the problem. These things need to be left to future research. Therefore, we will limit our search to partitions $\{A_1, ..., A_k\}$, whose atoms constitute product sets of marginal events on the one-dimensional spaces $\Omega(T_{2x})$, $\Omega(\kappa_v)$, $\Omega(Q_{\text{S90}})$, although such partitions cannot provide the best adaptation to the shape of the likehood function.

A partition $\{A_1, ..., A_k\}$ of the cubic parameter space $\Omega = (\underline{T}_{2x}, \overline{T}_{2x}] \times (\underline{\kappa}_v, \overline{\kappa}_v] \times (\underline{Q}_{\text{S90}}, \overline{Q}_{\text{S90}}]$, whose atoms are product sets of marginal events, is generated by a set of grid points

$$(T_{2x}, \kappa_v, Q_{\text{S90}}) \;\; \in \;\; G(l1, l2, l3) := \{T_1, ..., T_{l1}\} \times \{\kappa_1, ..., \kappa_{l2}\} \times \{Q_1, ..., Q_{l3}\} .$$

with $T_1 := \underline{T}_{2x}$, $T_{l1} := \overline{T}_{2x}$, $\kappa_1 := \underline{\kappa}_v$, $\kappa_{l2} := \overline{\kappa}_v$, $Q_1 := \underline{Q}_{\text{S90}}$, and $Q_{l3} := \overline{Q}_{\text{S90}}$. The grid $G(l1, l2, l3)$ specifies $l1$, $l2$ and $l3$ hyperplanes that are orthogonal to the dimension of $T_{2x}$, $\kappa_v$ and $Q_{\text{S90}}$, respectively. These hyperplanes define a partition of $\Omega$ with $k = (l1 - 1) \cdot (l2 - 1) \cdot (l3 - 1)$ atoms $A_{qrs} = (T_q, T_{q+1}] \times (\kappa_r, \kappa_{r+1}] \times (Q_s, Q_{s+1}]$. For the purpose of this analysis we have chosen the number of grid points to be $l1 = 15$ for climate sensitivity, $l2 = 11$ for ocean heat diffusivity and $l3 = 11$ for sulphate aerosol forcing. The proportion of these numbers reflects the relative importance of the individual parameters for determining the climate response. The absolute values were chosen heurestically under the requirement that the overall number of atoms should be on the order of $10^3$ for reasons of computational feasibility. It becomes apparent that computational feasibility exerts a strong limitation on the resolution of the partition in a multi-dimensional space. As indicated by the discussion at the end of Section 3.2.4, the number of atoms in each dimension should ideally be at least an order of magnitude higher, which would yield about $10^6$ atoms for the entire parameter space $\Omega$. Hence, the curse of dimensionality forces us to adopt only a coarse resolution of the uncertainty space. The resulting loss of information causes an imprecise probability analysis to widen the bounds for the lower and upper probability of events $A \subset \Omega$. However, these bounds will always contain the probability interval that would be obtained with any finer partition of the continuous parameter space. Therefore, adopting a coarse resolution can produce overly conservative, but not spurious estimates. It is interesting to note that lower and upper probability bounds will rapidly approach zero and one, respectively, if too much information has been given up in the process of the analysis. This leads to a state of complete ignorance, which tells the analyst that his information base was too weak to conduct the statistical inference in question.

In order to generate a likelihood-adapted partititon $\{A_1, ..., A_k\}$, we are looking for a grid $G^*(l1, l2, l3)$ which yields a similar change of likelihood

$$\Delta\mathcal{L}_{qrs} = \sup_{\substack{T_q \le T_{2x} \le T_{q+1} \\ \kappa_r \le \kappa_v \le \kappa_{r+1} \\ Q_s \le Q_{\mathrm{S90}} \le Q_{s+1}}} \mathcal{L}(T_{2x}, \kappa_v, Q_{\mathrm{S90}}; \hat{y}) - \inf_{\substack{T_q \le T_{2x} \le T_{q+1} \\ \kappa_r \le \kappa_v \le \kappa_{r+1} \\ Q_s \le Q_{\mathrm{S90}} \le Q_{s+1}}} \mathcal{L}(T_{2x}, \kappa_v, Q_{\mathrm{S90}}; \hat{y})$$

over each atom $A_{qrs}$. More formally, we require for all $1 \le q \le l1-1$, $1 \le r \le l2-1$, and $1 \le s \le l3-1$ that

$$\Delta\mathcal{L}_{qrs} = \frac{1}{k} \sum_{q=1}^{l1-1} \sum_{r=1}^{l2-1} \sum_{s=1}^{l3-1} \Delta\mathcal{L}_{qrs} \,. \tag{3.19}$$

This will require the atoms to have smaller than average volume in areas of strong likelihood changes and larger than average volume in areas of a flat likelihood function. Since the *quadratic mean* is minimised at the point where Condition (3.19) is fulfilled for all atoms, a likelihood-adapted grid $G^*(l1, l2, l3)$ can be found by solving the nonlinear program

$$G^*(l1, l2, l3) = \arg \min_{G(l1,l2,l3)} \sum_{q=1}^{l1-1} \sum_{r=1}^{l2-1} \sum_{s=1}^{l3-1} \Delta\mathcal{L}_{qrs}^2 \,. \tag{3.20}$$

Program (3.20) is very difficult to solve for arbitrary likelihood functions $\mathcal{L}(\cdot; \hat{y}) : \Omega \rightarrow \mathbb{R}_0^+$. The optimisation will be greatly simplified if $\mathcal{L}$ is quasi-concave. In this case, the cuts $C(L) = \{\theta | \mathcal{L}(\theta, \hat{y}) \ge L\}$, with $L \in \mathbb{R}^+$, are convex, and the infimum likelihood $\mathcal{L}_{qrs,*}$ over an atom $A_{qrs}$ is obtained at the vertices of $A_{qrs}$, while the supremum likelihood $\mathcal{L}_{qrs}^*$ will be situated on the edges of $A_{qrs}$ (except for the atom $A^*$ containing the maximum likelihood). Since the likelihood function $\mathcal{L}(\theta; \hat{y})$ calculated in Chapter 2 for 13650 grid points in $\Omega$ is approximately quasi-concave (see Fig. 2.7), we can indeed restrict our search for $\mathcal{L}_{qrs,*}$ and $\mathcal{L}_{qrs}^*$ to the edges and vertices of the atoms $A_{qrs}$. However, the simplified program is still computationally expensive, because it requires line searches for the supremum likelihood on all edges of the $k$ atoms in each step of the optimisation procedure. In our application, for example, we want to construct a partition with $14 \cdot 10 \cdot 10 = 1400$ atoms that have 4994 distinct edges. As a consequence, the calculation of the sum in the objective function of Program (3.20) involves 4994 line searches. Since the number of free hyperplanes to be adjusted to the shape of the likelihood function is $(l1-2) + (l2-2) + (l3-2) = 32$, we have to expect on the order of thousand iterations to solve Program (3.20).

In this situation we have decided to reduce the complexity of the optimisation problem by choosing a more heuristic approach for identifying a likelihood-adapted partition. In a first step, we have restricted us to the projections $\mathcal{L}(T_{2x}; \hat{y})$, $\mathcal{L}(\kappa_v; \hat{y})$ and $\mathcal{L}(Q_{\mathrm{S90}}; \hat{y})$ of the likelihood function $\mathcal{L}(\theta; \hat{y})$ onto the individual dimensions $X \in \{T_{2x}, \kappa_v, Q_{\mathrm{S90}}\}$ of the parameter space (see Fig. 2.8). They represent the upper envelope $\overline{\mathcal{L}}(X)$ of all intersects of the likelihood function that are parallel to the $T_{2x}$, $\kappa_v$ and $Q_{\mathrm{S90}}$-axes, respectively. Therefore, they capture the descent of the ridges of the likelihood function from the point of maximum likelihood. We solved a one-dimensional version of Program (3.20) for each parameter $T_{2x}$, $\kappa_v$ and $Q_{\mathrm{S90}}$ separately. In the one-dimensional case, a set of grid points $\{x_1, ..., x_l\}$ partitions the universal set $\Omega = (x_1, x_l]$ into $l-1$ half-closed intervals $A_j$. Since the projected likelihood functions are quasi-concave, the infimum and supremum likelihood over an $A_j$ will be obtained at the interval bounds of $A_j$ (except for the atom $A^*$ which contains the maximum likelihood). This greatly simplifies the evaluation of the objective function, and the one-dimensional version of Program 3.20 can be solved efficiently for $T_{2x}$, $\kappa_v$ and $Q_{\mathrm{S90}}$.

The resulting partitions of the domains $\Omega(T_{2x})$, $\Omega(\kappa_v)$, and $\Omega(Q_{\mathrm{S90}})$ are shown in Fig. 3.6 (black points in left panels). It can be seen that the likelihood changes are indeed equally spaced over the atoms. We joined the one-dimensional partitions to a three-dimensional partition of the entire parameter space $\Omega = \Omega(T_{2x}) \times \Omega(\kappa_v) \times \Omega(Q_{\mathrm{S90}})$ and checked its performance of adapting to the full likelihood function by calculating $\Delta\mathcal{L}_{qrs}$ for all atoms $A_{qrs}$. The average change of likelihood over the atoms was only $\bar{\Delta}\mathcal{L}_{qrs} = 0.05\mathcal{L}_{\max}$. However, this small average change is produced by the portion of atoms on the outskirts of the partition where the likelihood function is flat, and $\Delta\mathcal{L}_{qrs}$ is close to zero. In the center around the maximum likelihood, we have found atoms where the likelihood dropped from almost maximum to near zero. This shows that the upper envelopes $\overline{\mathcal{L}}(X)$ do not capture the steep slopes around the point of maximum likelihood.

Therefore, we have also looked at the intersects $\mathcal{L}^*(X)$ through the point of maximum likelihood $T_{2x} = 3.26\,\mathrm{K}$, $\kappa_v = 0.55\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, and $Q_{\mathrm{S90}} = -0.97\,\mathrm{W\,m}^{-2}$ that run parallel to the $X = T_{2x}$, $\kappa_v$, and $Q_{\mathrm{S90}}$-axes. As shown in Fig. 3.6 (left panels), they drop much faster to zero than the ridges $\overline{\mathcal{L}}(X)$ of the likelihood. To include this steep drop into the optimisation of the partition, we have recalculated the one-dimensional version of Program 3.20 for the sum $\overline{\mathcal{L}}(X) + \mathcal{L}^*(X)$ in each dimension $X = T_{2x}$, $\kappa_v$, and $Q_{\mathrm{S90}}$. The resulting partition is also shown in Fig. 3.6 (red points in left panels). It can be seen that the atom sizes around the point of maximum likelihood have decreased considerably, while the atoms on the outskirts of the domain were inflated. We have checked again the performance of the joint partition of $\Omega$ against the three-dimensional likelihood function. The average change of likelihood over the atoms was now $\bar{\Delta}\mathcal{L}_{qrs} = 0.12\mathcal{L}_{\max}$, which is considerably higher than before. Part of the increase can be explained by the larger portion of atoms that have been placed on the steep slopes of the likelihood function. However, the inflation of atoms on the outskirts of the partition has also increased $\Delta\mathcal{L}_{qrs}$ significantly in this area.

In this situation, we have chosen to combine the partitions that can be obtained from $\overline{\mathcal{L}}(X)$ and $\overline{\mathcal{L}}(X) + \mathcal{L}^*(X)$. The combined partition was constructed by replacing the $\overline{\mathcal{L}}(X)$-atoms around the maximum likelihood with the denser coverage of atoms that was obtained from solving the one-dimensional version of Program (3.20) for the sum $\overline{\mathcal{L}}(X) + \mathcal{L}^*(X)$. In this way, we generate a partition that concentrates on the maximum likelihood area without thinning the number of atoms on the outskirts. Fig. 3.6 (right panels) shows the combined partitions for each dimension $T_{2x}$, $\kappa_v$, and $Q_{\mathrm{S90}}$. As a consequence of the denser coverage around the maximum likelihood, the number of atoms in the joint partition of $\Omega$ has increased from 1400 to 2592 (18 atoms for climate

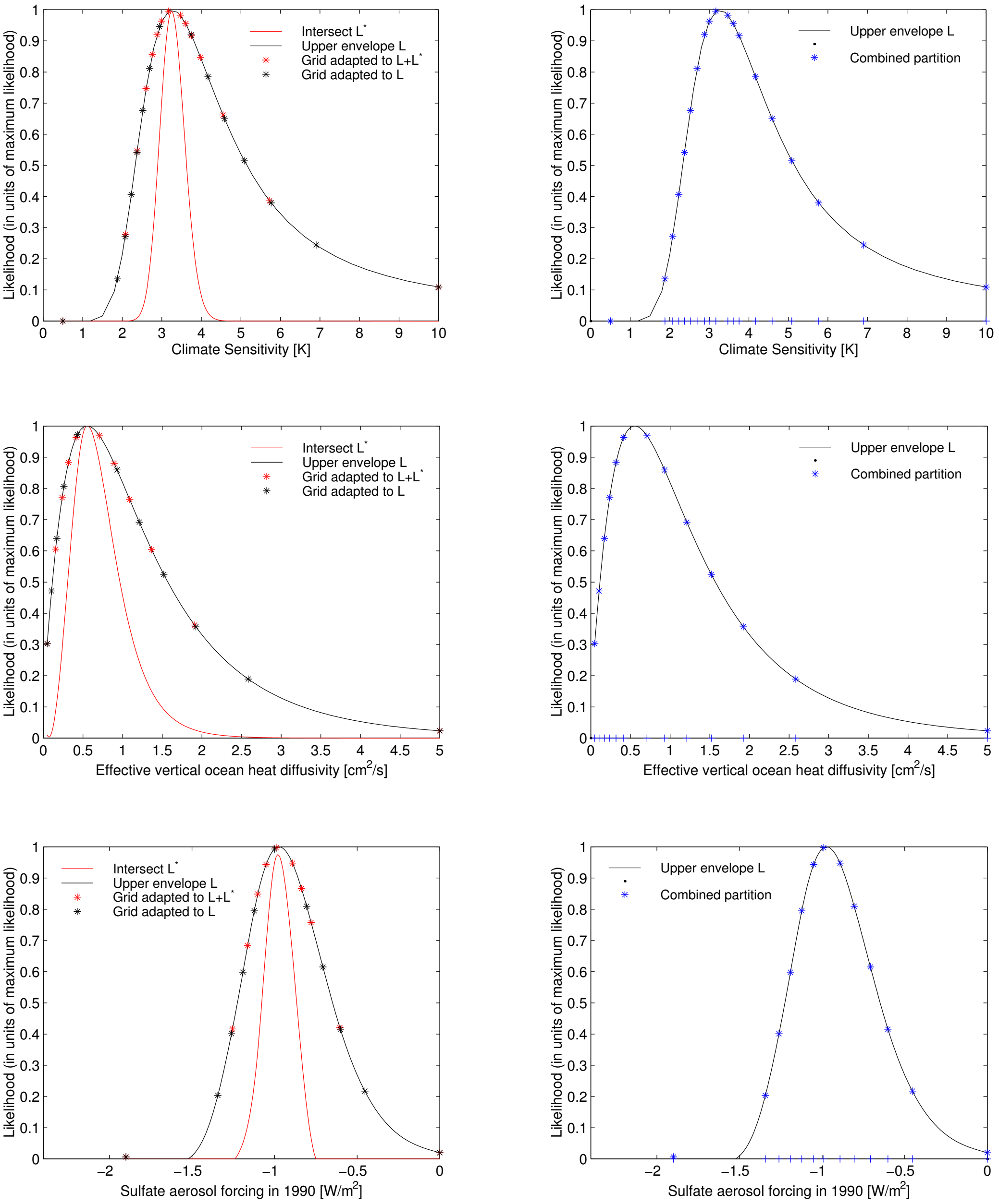


Figure 3.6: Likelihood-adapted partitions for the domains $\Omega(T_{2x})$, $\Omega(\kappa_v)$ and $\Omega(Q_{\mathrm{S90}})$. The left panels show the optimally adapted grid (black points) to the upper envelope $\overline{\mathcal{L}}(X)$ (black line, $X \in \{T_{2x}, \kappa_v, Q_{\mathrm{S90}}\}$), as well as the optimally adapted grid (red points) that emerges from inclusion of the intersect $\mathcal{L}^*(X)$ (red line). The right panels show the combined set of grid points that define the partitions to be used in the following analysis.

sensitivity, and 12 atoms for ocean heat diffusivity and aerosol forcing each). We have checked the performance of the partition in adapting to the likelihood function, and found an average change of likelihood of $\bar{\Delta}\mathcal{L}_{qrs} = 0.07\mathcal{L}_{\max}$. This value change is close to the value for the original partition, albeit the portion of atoms in the steep slopes around the maximum likelihood has been increased. We have compared the partitions also in terms of the minimum likelihood that is allocated to the atoms. It turns out that the average minimum likelihood of the combined partition is a factor two smaller than for the maximum-likelihood centered partition generated from $\overline{\mathcal{L}}(X)+\mathcal{L}^*(X)$, but exceeds the average minimum over the $\overline{\mathcal{L}}(X)$-partition by a factor five. We conclude that the combined partition indeed provides a good compromise between a maximum-likelihood centered and an upper-envelope based partition. Hence, we will use the combined partition as shown in Fig. 3.6 (right panels) in the following. It will be discussed below that further atoms need to be added to this partition to capture the structure in the tails of the prior uncertainty for $T_{2x}$, $\kappa_v$, and $Q_{\text{S90}}$.

### 3.4.2 A belief function for climate sensitivity

We generate a belief function representation for the prior uncertainty about climate sensitivity from recent probability estimates in the literature (Andronova and Schlesinger, 2001; Knutti et al., 2002; Forest et al., 2002; Murphy et al., 2004). The published probability distributions and the various methods of their derivation are discussed in Section 3.2.3. Fig. 3.1 shows the six estimates that we have selected to include in the prior uncertainty. The choice of selection is motivated in Section 3.2.3. As explained above, we want to construct the convex set of of all probability measures whose

1. cumulative distribution functions lie between the lower envelope $\underline{F}$ and upper envelope $\overline{F} : \Omega(T_{2x}) \rightarrow [0,1]$ of the six distribution functions from the literature (see Definition 3.6), and whose

2. probability masses on a likelihood-adapted partition $\{A_1, ..., A_k\}_{T_{2x}}$ are larger or equal than the lower envelope $\underline{p} : \{A_1, ..., A_k\}_{T_{2x}} \rightarrow [0,1]$ of the probability masses which are allocated by the six literature estimates to the atoms of the partition.

It was shown in Section 3.2.4 that a continuous distribution band $\Gamma(\underline{F}, \overline{F})$ can be approximated by a p-box $\Gamma(\underline{SF}, \overline{SF}) \supseteq \Gamma(\underline{F}, \overline{F})$, whose lower and upper step functions confine the envelope of the distribution band from below and above, respectively. For the reasons explained in Section 3.4.1, we want to construct the right-continuous lower step function $\underline{SF} : \{x_{1*}, ..., x_{n*}\} \rightarrow \{\underline{SF}_1, ..., \underline{SF}_n\}$ and left-continuous upper step function $\overline{SF} : \{x_1^*, ..., x_m^*\} \rightarrow \{\overline{SF}_1, ..., \overline{SF}_m\}$ (see Definition 3.8) on the likelihood-adapted partition $\{A_1, ..., A_{18}\}_{T_{2x}}$ which was generated above. This requires that the points $\{x_{1*}, ..., x_{n*}\}$ and $\{x_1^*, ..., x_m^*\}$ have to be chosen from the set of grid points $\{\underline{T}_{2x}, a_1, ..., a_{k-1}, \overline{T}_{2x}\}$ which define the partition $\Omega(T_{2x}) = (\underline{T}_{2x}, a_1] \cup (a_1, a_2] \cup ... \cup (a_{k-1}, \overline{T}_{2x}]$. It turns out that the likelihood-adapted partition for climate sensitivity does not provide enough grid points to approximate the tails of the continuous distribution bands without an unreasonably large error. Therefore, we extend the partition by inserting two more grid points in the tails of the distribution band. As a consequence, the adjusted partition for climate sensitivity now contains 20 atoms $\{A_1, ..., A_{20}\}_{T_{2x}}$.

Fig. 3.3.a in Section 3.2.4 shows the lower and upper cumulative probability values that the continuous distribution band $\Gamma_{T_{2x}}(\underline{F}, \overline{F})$ exhibits at the grid points of the atoms. This information is used to construct the p-box approximation $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}) \supset \Gamma_{T_{2x}}(\underline{F}, \overline{F})$ for climate sensitivity (shown in Fig. 3.3.b). Algorithm 3.1 was used to construct the random set $(\mathcal{E}_{T_{2x}}, \nu)$ which provides a representation of the p-box $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF})$ (see Theorem 3.1). It is constituted by 9 half-closed intervals of the real line that are included in the power set of the partition $\{A_1, ..., A_{20}\}_{T_{2x}}$ (see Fig. 3.3.b and Table 3.1).

In a next step we construct the random set $(\mathcal{F}_{T_{2x}}, \underline{p})$ of the $\varepsilon$-contamination model $\Gamma(\underline{p})$ that is generated by the lowest probability masses $\{\underline{p}_1, ..., \underline{p}_{20}\}$ from the selected set of literature estimates. Equation 3.13 shows that $(\mathcal{F}_{T_{2x}}, \underline{p})$ can be deduced immediately from the knowledge of the partition $\{A_1, ..., A_{20}\}_{T_{2x}}$, and the lower mass function $\underline{p} : \{A_1, ..., A_{20}\}_{T_{2x}} \to [0, 1]$. The six probability estimates in Andronova and Schlesinger (2001), Knutti et al. (2002), Forest et al. (2002), and Murphy et al. (2004) allocate positive probability mass unanimously only to the 16 atoms in the range $T_{2x} \in (1.88\,\mathrm{K}, 6.90\,\mathrm{K}]$. Since the minimum probability mass allocated to the outer atoms $(1.88\,\mathrm{K}, 2.08\,\mathrm{K}]$ and $(5.77\,\mathrm{K}, 6.9\,\mathrm{K}]$ is small, i.e., $\underline{p}_j < 0.005$, we exclude this information from our analysis. The small loss of information is outweighed by the considerable gain in computational efficiency from working with a smaller number of focal elements. Hence, the random set for $\Gamma(\underline{p})$ comprises the 14 atoms in the range $(2.08\,\mathrm{K}, 5.77\,\mathrm{K}]$ carrying altogether lower probability mass 0.32 and the domain $\Omega(T_{2x})$ carrying the remaining probability mass $\varepsilon = 0.68$ . $(\mathcal{F}_{T_{2x}}, \underline{p})$ is tabulated in Table 3.2, and the corresponding lower and upper mass functions $\underline{p}$ and $\underline{p}+\varepsilon$ are shown in Fig. 3.5.

We can now apply Algorithm 3.2 to combine the random set $(\mathcal{E}_{T_{2x}}, \nu)$ of the p-box and the random set $(\mathcal{F}_{T_{2x}}, \underline{p})$ of the $\varepsilon$-contamination model to the random set $(\mathcal{G}_{T_{2x}}, \nu^*)$ of their intersection $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}, \underline{p})$. We have shown in Section 3.3 that $(\mathcal{G}_{T_{2x}}, \nu^*)$ provides a representation of the convex set of probabilities which observe the p-box constraints and the lower bound on the probability masses simultaneously. The random set $(\mathcal{G}_{T_{2x}}, \nu^*)$ constructed by means of Algorithm 3.2 is tabulated in Table 3.4. As we have predicted in Lemma 3.5, all Möbius masses are positive. Hence, $(\mathcal{G}_{T_{2x}}, \nu^*)$ defines a belief function constituting the lower envelope of the structure $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}, \underline{p})$ (see Theorem 3.2).

The random set $(\mathcal{G}_{T_{2x}}, \nu^*)$ contains 29 focal elements, which represents only a mild increase from the original number of $9+14 = 23$ focal elements collected in $(\mathcal{E}_{T_{2x}}, \nu)$ and $(\mathcal{F}_{T_{2x}}, \underline{p})$. In particular, the number is an order of magnitude smaller than the number $2^9 + 14 - 1 = 531$ of potential focal elements that could be identified by Algorithm 3.2. This underlines our assertion in Section 3.3 that the benign properties of the intersection $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}, \underline{p})$ prevent an explosion of non-zero Möbius assignments associated with its lower envelope $\underline{P}^* : \mathcal{R} \to [0, 1]$. To put the number of 29 focal elements into perspective, we can contrast it with the number $k = 20$ of atoms in the partition. Every probability measure on the power set $\mathcal{P}(A_1, ..., A_k)$ would need to be represented by 20 probability masses $p_1, ..., p_{20}$. To the other extreme, the power set $\mathcal{P}(A_1, ..., A_k)$ contains $2^{20} - 1 \approx 10^6$ non-empty sets which could carry in principle non-zero Möbius

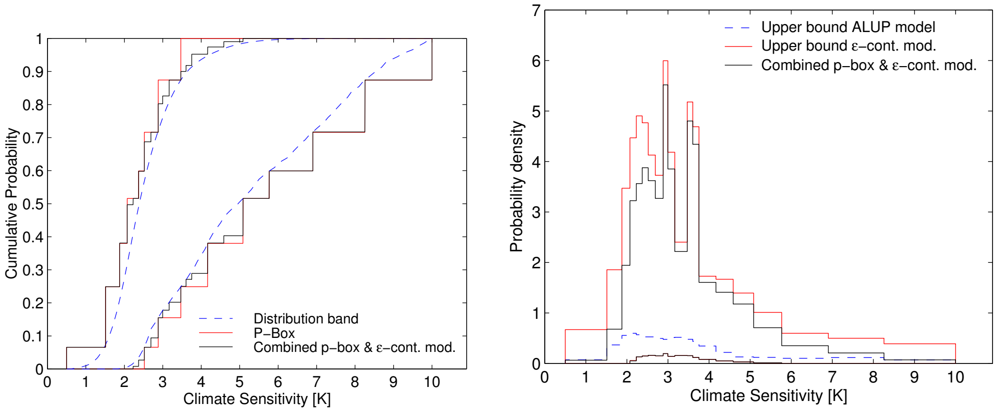
| | Focal element | Möbius mass | | Focal element | Möbius mass |
|---|---|---|---|---|---|
| $G_1$ | $(2.08\,\mathrm{K}, 2.23\,\mathrm{K}]$ | 0.008 | $G_{15}$ | $(0.50\,\mathrm{K}, 2.52\,\mathrm{K}]$ | 0.017 |
| $G_2$ | $(2.23\,\mathrm{K}, 2.37\,\mathrm{K}]$ | 0.019 | $G_{16}$ | $(3.47\,\mathrm{K}, 10.0\,\mathrm{K}]$ | 0.005 |
| $G_3$ | $(2.37\,\mathrm{K}, 2.52\,\mathrm{K}]$ | 0.022 | $G_{17}$ | $(0.50\,\mathrm{K}, 2.88\,\mathrm{K}]$ | 0.032 |
| $G_4$ | $(2.52\,\mathrm{K}, 2.69\,\mathrm{K}]$ | 0.028 | $G_{18}$ | $(2.52\,\mathrm{K}, 8.26\,\mathrm{K}]$ | 0.003 |
| $G_5$ | $(2.69\,\mathrm{K}, 2.88\,\mathrm{K}]$ | 0.029 | $G_{19}$ | $(2.88\,\mathrm{K}, 10.0\,\mathrm{K}]$ | 0.063 |
| $G_6$ | $(2.88\,\mathrm{K}, 3.00\,\mathrm{K}]$ | 0.023 | $G_{20}$ | $(1.51\,\mathrm{K}, 4.17\,\mathrm{K}]$ | 0.041 |
| $G_7$ | $(3.00\,\mathrm{K}, 3.17\,\mathrm{K}]$ | 0.024 | $G_{21}$ | $(1.88\,\mathrm{K}, 5.77\,\mathrm{K}]$ | 0.028 |
| $G_8$ | $(3.17\,\mathrm{K}, 3.47\,\mathrm{K}]$ | 0.048 | $G_{22}$ | $(2.08\,\mathrm{K}, 6.90\,\mathrm{K}]$ | 0.014 |
| $G_9$ | $(3.47\,\mathrm{K}, 3.60\,\mathrm{K}]$ | 0.021 | $G_{23}$ | $(2.37\,\mathrm{K}, 8.26\,\mathrm{K}]$ | 0.061 |
| $G_{10}$ | $(3.60\,\mathrm{K}, 3.75\,\mathrm{K}]$ | 0.018 | $G_{24}$ | $(2.52\,\mathrm{K}, 10.0\,\mathrm{K}]$ | 0.058 |
| $G_{11}$ | $(3.75\,\mathrm{K}, 4.17\,\mathrm{K}]$ | 0.034 | $G_{25}$ | $(1.51\,\mathrm{K}, 5.09\,\mathrm{K}]$ | 0.097 |
| $G_{12}$ | $(4.17\,\mathrm{K}, 4.59\,\mathrm{K}]$ | 0.022 | $G_{26}$ | $(0.50\,\mathrm{K}, 4.17\,\mathrm{K}]$ | 0.017 |
| $G_{13}$ | $(4.59\,\mathrm{K}, 5.09\,\mathrm{K}]$ | 0.016 | $G_{27}$ | $(1.88\,\mathrm{K}, 6.90\,\mathrm{K}]$ | 0.104 |
| $G_{14}$ | $(5.09\,\mathrm{K}, 5.77\,\mathrm{K}]$ | 0.010 | $G_{28}$ | $(2.08\,\mathrm{K}, 8.26\,\mathrm{K}]$ | 0.094 |
| | | | $G_{29}$ | $(1.51\,\mathrm{K}, 5.77\,\mathrm{K}]$ | 0.045 |

Table 3.4: Random set $(\mathcal{G}_{T_{2x}}, \nu^*)$ of the intersection $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}, \underline{p})$ of p-box and $\varepsilon$-contamination model for climate sensitivity.

masses. Hence, the Möbius inverse constructed from Algorithm 3.2 can be considered sparse indeed.

Fig. 3.7 shows the distribution and probability mass bands that are spanned by the probabilities contained in $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}, \underline{p})$. It can be seen that the p-box information is tightened by the inclusion of the lower probability masses on the atoms, but only in areas with atoms $A_j$ carrying $\underline{p}_j > 0$. Since p-box and $\varepsilon$-contamination model have been constructed from the same set of probabilities $co(P_{1,T_{2x}}, ..., P_{6,T_{2x}})$, the lower and upper distribution function emerging from $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}, \underline{p})$ need to fully include the distribution band set up by $co(P_{1,T_{2x}}, ..., P_{6,T_{2x}})$. However, a tightening of the p-box approximation can happen in so far as the lower mass function information recovers some of the information that was lost in the approximation process. The information gain from combining the two models can also be seen in the mass band. Here, the p-box information helps to tighten the upper bound on the probability masses $\overline{p}_j$. This effect is particularly large in areas where no atoms $A_j$ with $\underline{p}_j > 0$ existed. Table 3.3 compares the information content of $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}, \underline{p})$ with the convex hull, the continuous distribution band, the p-box and the $\varepsilon$-contamination model for the example of the IPCC estimate for climate sensitivity. It can be seen that the intersection of p-box and $\varepsilon$-contamination model always collects the tighter bounds from the two models.

For the subsequent analysis, we will adjust the random set $(\mathcal{G}_{T_{2x}}, \nu^*)$ listed in Table 3.4 slightly. The focal elements $G_{16} = (3.47\,\mathrm{K}, 10.0\,\mathrm{K}]$ and $G_{18} = (2.52\,\mathrm{K}, 8.26\,\mathrm{K}]$ carry only small probability mass $\leq 0.005$. We remove the two focal elements from $(\mathcal{G}_{T_{2x}}, \nu^*)$, and redistribute their masses on the supersets $G_{24} = (2.89\,\mathrm{K}, 10.0\,\mathrm{K}]$ and $G_{23} = (2.37\,\mathrm{K}, 8.26\,\mathrm{K}]$, respectively. Inspection of the relationships (3.3) and (3.4) between a belief function, its conjugate plausibility function and the Möbius inverse reveals that such a redistribution of probability mass on supersets results in losening the probability bounds for all events $A \in \mathcal{R}$. Hence, we have induced a small loss of information, which is easily outweighed by the gain in computational efficiency from



a. Distribution band

b. Mass band

Figure 3.7: Distribution band and mass band (converted to probability densities) generated from the random set $(\mathcal{G}_{T_{2x}}, \nu^*)$ of the intersection $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}, \underline{p})$ (black lines). The left panel also shows the corresponding bounding functions of the p-box and the continuous distribution band. The right panel compares the upper mass function with the $\varepsilon$-contamination model and the ALUP model.

reducing the number of focal elements. The adjusted random set $(\mathcal{G}_{T_{2x}}, \nu^*)$ contains 27 focal elements. In the following, we will use this random set as the representation of the prior uncertainty about climate sensitivity.

### 3.4.3 A belief function for sulphate aerosol forcing

We construct a belief function representation for the prior uncertainty about the *combined direct and indirect sulphate aerosol forcing around the year 1990* in analogy to the case of climate sensitivity. Andronova and Schlesinger (2001), Knutti et al. (2002) and Forest et al. (2002) have considered aerosol forcing besides climate sensitivity as uncertain parameter in their probabilistic analyses that were discussed in Section 3.2.3. They all provided marginal probability distributions for the aerosol forcing strength, albeit with varying sets of aerosol species that were accounted for in these estimates. Andronova and Schlesinger (2001) have focused on the sulphate aerosol forcing in the year 1990 as we want to do here. Moreover, they used the same parameterisation of the direct and indirect forcing effect of sulphate aerosols that was employed for the calculation of the likelihood function in Chapter 2. The ratio between indirect and direct forcing effect was set to 0.8 / 0.3, which is slightly different than the ratio of 0.8 / 0.4 that we have adopted in Equation (2.9). However, this difference is small and, therefore, the probability estimate of Andronova and Schlesinger (2001) for the total sulphate aerosol forcing can be directly related to the uncertain parameter $Q_{S90}$ that we consider here. The probability distribution for the GTAS forcing scenario (Greenhouse gases + Tropospheric Ozone + Sulphate Aerosols + Sun) is shown in Fig. 3.8. We have motivated this choice of forcing scenario in Section 3.2.3.

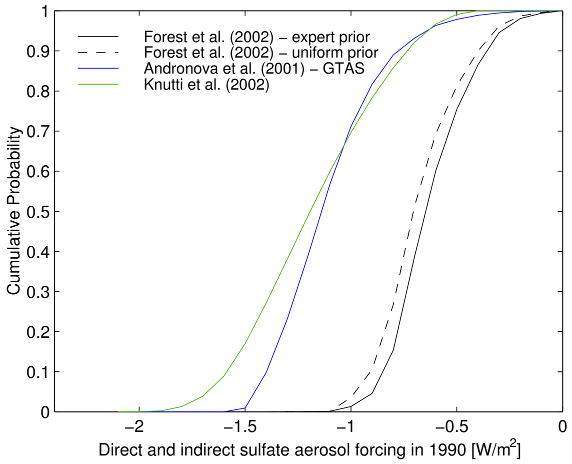


Figure 3.8: Cumulative distribution functions for direct and indirect sulphate aerosol forcing from the literature.

The situation is different for the probability estimates of Knutti et al. (2002) and Forest et al. (2002). In their publication, Knutti et al. (2002) focused on constraining the indirect aerosol forcing effect of aerosols. We have obtained the corresponding probability estimate for the combined direct and indirect aerosol forcing directly from the authors. As can be seen in Fig. 3.8, it is shifted towards larger negative values in the tail of the distribution relative to the GTAS estimate of Andronova and Schlesinger (2001). A reason may be that the forcing includes the possibility of a significant cooling from carbonaceous aerosols. Knutti et al. (2002) consider forcings in their Monte Carlo analysis that can be as large as $Q_{\mathrm{Car}} = -0.4\,\mathrm{W\,m^{-2}}$ (Knutti et al., 2003, Table 1). We include the forcing estimate of Knutti et al. (2002) into our uncertainty about $Q_{\mathrm{S90}}$ to account for a possible underestimation of the cooling contribution from organic carbon and biomass burning aerosols in our forcing scenario. If this was the case, the effective sulphate aerosol forcing $Q_{\mathrm{S90}}$ would need to be increased (in absolute terms) to reproduce the total forcing of all aerosol species.

However, a similar argument can be set up for decreasing the effective cooling from sulphate aerosols. Black carbon is an absorber of solar radiation and, therefore, provides a net heating of the atmosphere. As discussed in Section 2.2.2, the magnitude of this heating is subject to scientific debate. Some authors have even suggested that it might counterbalance most of the cooling from sulphate aerosols (Jacobson, 2001; Hansen et al., 2002). In this case, we would have strongly underestimated the heating contribution of black carbon, since it is already outweighed by the negative forcing from organic and biomass burning aerosols in the scenario adopted in Section 2.2.2. To account for this possibility, we also include the probability estimates of Forest et al. (2002) for the net total aerosol forcing in the 1980s in our prior uncertainty model. Their two estimates under different assumptions about the prior probability (see Section 3.2.3)

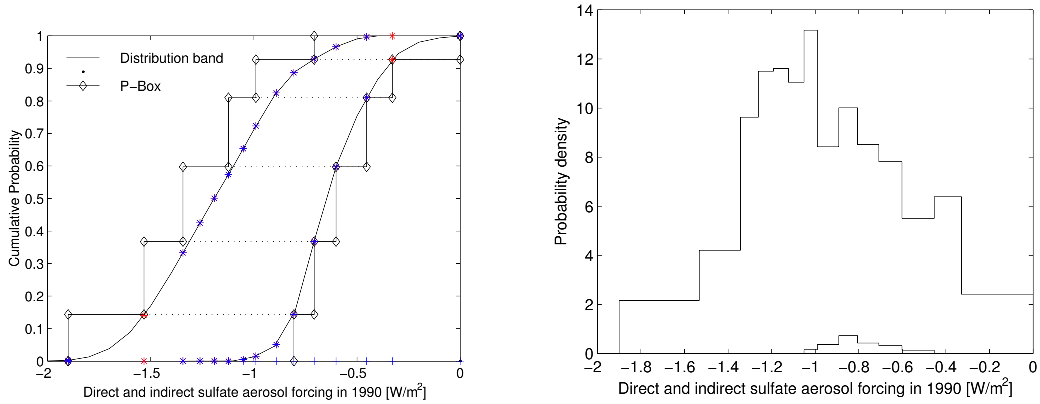
| $\varepsilon$-contamination model & p-box | | | Intersection of both models | | |
|---|---|---|---|---|---|
| | Focal element (in $\mathrm{W\,m^{-2}}$) | Möbius mass | | Focal element (in $\mathrm{W\,m^{-2}}$) | Möbius mass |
| $B_1$ | $(-1.05, -0.99]$ | 0.010 | $G_1$ | $(-1.05, -0.99]$ | 0.010 |
| $B_2$ | $(-0.99, -0.89]$ | 0.036 | $G_2$ | $(-0.99, -0.89]$ | 0.036 |
| $B_3$ | $(-0.89, -0.81]$ | 0.062 | $G_3$ | $(-0.89, -0.81]$ | 0.062 |
| $B_4$ | $(-0.81, -0.71]$ | 0.042 | $G_4$ | $(-0.81, -0.71]$ | 0.042 |
| $B_5$ | $(-0.71, -0.60]$ | 0.034 | $G_5$ | $(-0.71, -0.60]$ | 0.034 |
| $B_6$ | $(-0.60, -0.45]$ | 0.020 | $G_6$ | $(-0.60, -0.45]$ | 0.020 |
| $E_1$ | $(-1.90, -0.81]$ | 0.144 | $G_7$ | $(-1.90, -0.81]$ | 0.036 |
| $E_2$ | $(-1.53, -0.71]$ | 0.224 | $G_8$ | $(-1.53, -0.71]$ | 0.074 |
| $E_3$ | $(-1.34, -0.60]$ | 0.230 | $G_9$ | $(-1.34, -0.60]$ | 0.046 |
| $E_4$ | $(-1.12, -0.45]$ | 0.212 | $G_{10}$ | $(-1.12, -0.45]$ | 0.008 |
| $E_5$ | $(-0.99, -0.33]$ | 0.117 | $G_{11}$ | $(-0.71, 0.00]$ | 0.019 |
| $E_6$ | $(-0.71, 0.00]$ | 0.073 | $G_{12}$ | $(-1.90, -0.71]$ | 0.108 |
| | | | $G_{13}$ | $(-1.53, -0.60]$ | 0.150 |
| | | | $G_{14}$ | $(-1.34, -0.45]$ | 0.184 |
| | | | $G_{15}$ | $(-1.12, -0.33]$ | 0.117 |
| | | | $G_{16}$ | $(-1.12, 0.00]$ | 0.054 |

Table 3.5: Random sets $(\mathcal{F}_{Q_{\mathrm{S90}}}, \underline{p})$ of the $\varepsilon$-contamination model (upper left, focal element $\Omega(Q_{\mathrm{S90}})$ with mass 0.80 omitted), $(\mathcal{E}_{Q_{\mathrm{S90}}}, \underline{p})$ of the p-box (lower left), and $(\mathcal{G}_{Q_{\mathrm{S90}}}, \nu^*)$ of their intersection $\Gamma_{Q_{\mathrm{S90}}}(\underline{SF}, \overline{SF}, \underline{p})$ (right side) for sulphate aerosol forcing.

are shown in Fig. 3.8. They are shifted towards a considerably smaller forcing contribution of sulphate aerosols as compared to the estimates of Knutti et al. (2002) and Andronova and Schlesinger (2001). They even give some support to the possibility of a vanishing cooling effect of sulphate aerosols.

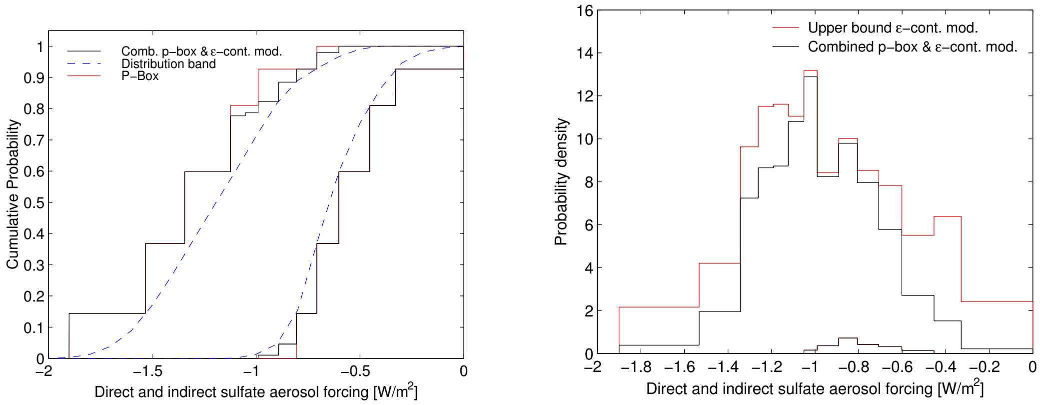
Since the following construction of the prior imprecise probability model for $Q_{\mathrm{S90}}$ evolves completely analogous to the case of climate sensitivity, we comment the individual steps only briefly. We use the convex hull $co(P_{1,Q_{\mathrm{S90}}}, ..., P_{4,Q_{\mathrm{S90}}})$ of the four probability estimates from the literature to generate a continuous distribution band $\Gamma_{Q_{\mathrm{S90}}}(\underline{F}, \overline{F})$. Fig. 3.9.a shows the lower and upper distribution functions of the band. The blue points denote the lower and upper cumulative probabilities that are obtained at the grid points of the likelihood-adapted partition $\Omega(Q_{\mathrm{S90}}) = (\underline{Q_{\mathrm{S90}}} = -1.9\,\mathrm{W\,m^{-2}}, b_1] \cup (b_1, b_2] \cup ... \cup (b_{k-1}, \overline{Q_{\mathrm{S90}}} = 0\,\mathrm{W\,m^{-2}}]$, with $k = 12$ atoms, which was generated in Section 3.4.1. It turns out that the spacing of grid points on the outskirts of the domain is too sparse to approximate the tails of the distribution band in a reasonable manner. Therefore, we insert two more grid points into the support of the tails (red points in Fig. 3.9.a), which extends the partition to 14 atoms $\{B_1, ..., B_{14}\}_{Q_{\mathrm{S90}}}$. The p-box approximation $\Gamma_{Q_{\mathrm{S90}}}(\underline{SF}, \overline{SF})$ of the continuous distribution band is now constructed on the extended set of grid points for a choice of $n = 6$ probability levels (see Fig. 3.9.a). We use Algorithm 3.1 to generate the random set $(\mathcal{E}_{Q_{\mathrm{S90}}}, \nu)$ that captures the information content of the p-box. The resulting random set with 6 focal elements is listed in Table 3.5.

Fig. 3.9.b shows the lower mass function $\underline{p} : \{B_1, ..., B_{14}\}_{Q_{\mathrm{S90}}} \rightarrow [0, 1]$ on the likelihood adapted partition that is defined by the set of probability estimates from the literature. Only the 6 atoms in the range $(-1.05\,\mathrm{W\,m^{-2}}, -0.45]$ carry a significant



a. P-Box

b. $\varepsilon$-contamination model



c. Distribution band of intersection

d. Mass band of intersection

Figure 3.9: P-Box approximation of the continuous distribution band (Panel a) and $\varepsilon$-contamination model (Panel b) for sulphate aerosol forcing. Lower panels show the distribution and mass band of the intersection $\Gamma_{Q_{\mathrm{S90}}}(\underline{SF}, \overline{SF}, \underline{p})$ of the two models in comparison.

lower probability mass $\underline{p} > 0.005$. The corresponding random set $(\mathcal{F}_{Q_{\mathrm{S90}}}, \nu)$ is listed in Table 3.5. Its focal elements are comprised by 6 atoms carrying altogether lower probability mass $\underline{p} = 0.20$ and the domain $\Omega(Q_{\mathrm{S90}})$ carrying the remaining probability mass $\varepsilon = 0.80$.

We employ Algorithm 3.2 to generate the random set $(\mathcal{G}_{Q_{\mathrm{S90}}}, \nu^*)$ of the intersection $\Gamma_{Q_{\mathrm{S90}}}(\underline{SF}, \overline{SF}, \underline{p})$ from knowledge of $(\mathcal{E}_{Q_{\mathrm{S90}}}, \nu)$ and $(\mathcal{F}_{Q_{\mathrm{S90}}}, \nu)$. The result is tabulated in Table 3.5. $(\mathcal{G}_{Q_{\mathrm{S90}}}, \nu^*)$ contains 16 focal elements with positive Möbius mass as predicted in Lemma 3.5. It defines a belief function constituting the lower envelope of $\Gamma_{Q_{\mathrm{S90}}}(\underline{SF}, \overline{SF}, \underline{p})$. Note that the number of focal elements is only slightly larger than the number of atoms of the partition. Fig. 3.9.c compares the information content of the intersection $\Gamma_{Q_{\mathrm{S90}}}(\underline{SF}, \overline{SF}, \underline{p})$ to p-box and continuous distribution band on the series of cumulative events. By definition, the intersection provides identical or tighter bounds for the distribution band than the p-box. An analogous comparison is made in Fig. 3.9.d for the lower and upper mass functions on the partition $\{B_1, ..., B_{14}\}_{Q_{\mathrm{S90}}}$. Compared to the $\varepsilon$-contamination model, the inclusion of the p-box information tightens the upper bounds on the probability masses particularly on the outskirts of the domain $\Omega(Q_{\mathrm{S90}})$. In the following, we will use the random set $(\mathcal{G}_{Q_{\mathrm{S90}}}, \nu^*)$ as the representation of the prior uncertainty about $Q_{\mathrm{S90}}$.

### 3.4.4 A belief function for ocean heat diffusivity

A belief function for the prior uncertainty about *effective vertical ocean heat diffusivity* $\kappa_v$ cannot be constructed in the same manner as for $T_{2x}$ and $Q_{\mathrm{S90}}$, because $\kappa_v$ is a model dependent parameter. It captures the overall thermal conductivity of the ocean column which is determined by a multitude of processes. Only a small portion of the conductivity stems from molecular heat diffusion, while the larger part is due to turbulent mixing in the water column. Therefore, the value of $\kappa_v$ will depend on the modeller's choice which mixing processes to resolve explicitly and which processes to include in the $\kappa_v$ parameterisation. In our 1-D diffusion ocean model derived in Appendix B all diffusive processes are subsumed in $\kappa_v$. Hence, we can assess a plausible magnitude of $\kappa_v$ from estimates for the globally and vertically averaged effective heat diffusivity in the ocean which lie around $\kappa_v = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$ (Munk, 1966; Simmons et al., 2004).

This plausibility assessment does not suffice to construct an imprecise probability model for the prior uncertainty about effective ocean heat diffusivity $\kappa_v$. However, the value of $\kappa_v$ can be directly related to the ocean heat uptake in the second half of the 20th century which has been estimated from ocean temperature measurements recently (Levitus et al., 2000, 2001). In our modelling framework the ocean heat uptake in a given period $[t_o, t_1]$ is specified by (see Section 2.1.2 and Appendix B, Equation B.26)

$$\begin{aligned} H_O(t_o, t_1) \quad &:= \quad \int_{t_o}^{t_1} \left( C_{AS}\, \dot{T}_S(t') + F_O(t') \right) \, dt' \qquad (3.21) \\ &= \quad C_{AS}\ \left(T_S(t_e) - T_S(t_o)\right) + f_{SO}\, c_V \sqrt{\frac{\kappa_v}{\pi}} \int_{t_o}^{t_1} \int_0^{t'} \frac{\dot{T}_S(t'')}{\sqrt{t'-t''}} \, dt''\, dt' \ , \end{aligned}$$

where we have neglected the higher order bottom correction terms. They do not influence the heat uptake for several hundred years after the development of a temperature anomaly $T_S$ at the sea surface. The first term specifies the heat uptake of the mixed layer with a depth of 60 m and a heat capacity of $C_{AS} = 7.80\,\mathrm{Wyr\,m^{-2}K^{-1}}$ (see Table 2.1). The second term captures the integral of the heat flux into the interior ocean. It depends only on $\kappa_v$, since the parameters $f_{SO}$ and $c_V$ are fixed by topographical and physical considerations, respectively (see Appendix A.4.1). Hence, knowledge of the historical SSTs allows us to directly calculate a mapping from vertical heat diffusivity $\kappa_v$ onto ocean heat uptake for an arbitrary period $[t_o, t_1]$.

Fig. 3.10 shows the relationship between $\kappa_v$ and the heat uptake in the period 1955 to 1996. It was derived by inserting the global mean SST record based on the HadSST

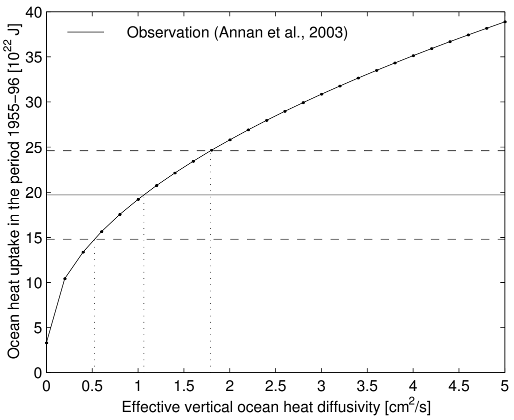


Figure 3.10: Relationship between effective vertical ocean heat diffusivity $\kappa_v$ and ocean heat uptake in the period 1955-1996 under the assumption of the global mean SST record based on the HadSST data set. The central value and standard deviation of the observation based estimate of Annan and Hargreaves (2003) (see Text) are shown for comparison.

data set (Jones et al. 2001, see Fig. 2.3.a) in Equation (3.21). As can be seen directly from Equation (3.21) the heat uptake increases in proportion to the square root of $\kappa_v$, i.e., $H_O(1955, 1996) = a + b\sqrt{\kappa_v}$. Here, $a = 3.3 \cdot 10^{22}\,\mathrm{J}$ specifies the amount of heat that is taken up by the mixed layer[6]. Observational errors in the instrumental SST record will introduce uncertainty about the heat uptake of both the mixed layer (coefficient $a$) and the interior ocean (coefficient $b$). The amount of this uncertainty is hard to assess because the global mean SST record derived from the HadSST data set was not attached with error estimates. What is more important, the inclusion of these errors into Equation (3.21) requires an assumption about their serial correlation, since we cannot expect the observation errors to be a set of IID normal random variables. However, a plausible assumption about this serial correlation was not available to us. In this situation, we abstain from specifying an error in the relationship depicted in Fig. 3.10, albeit we acknowledge that it can be sensitive to the measurement error in the instrumental temperature record. This issue needs to be investigated in further research which is beyond the prototypical analysis presented here. Plausible values of vertical diffusivity can be assessed by comparing its implication for ocean heat uptake with observations of the heat content in the world ocean (Levitus et al., 2000). Fig. 3.10 includes an estimate of heat uptake over the period 1955-96 (Annan and Hargreaves, 2003) that has been derived from the data of Levitus et al. (2000). It can be seen that its central value at $19.7 \cdot 10^{22}\,\mathrm{J}$ suggest an ocean heat diffusivity of $\kappa_v \approx 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$ in agreement with the canonical estimate of Munk (1966).

If the observational estimate of ocean heat uptake was only afflicted with probabilistic uncertainty, we could directly deduce a probability distribution for ocean heat

---

[6] A heat uptake of $1.14 \cdot 10^{22}\,\mathrm{J}$ equals an average heat uptake of $1\,\mathrm{W\,yr\,m^{-2}}$ over the world ocean area.

diffusivity $\kappa_v$ from knowledge of the transfer function $H_O(\kappa_v)$. Since imprecise probability constitutes a generalisation of additive probability, this special case could be accommodated easily in the framework of our analysis. All we need to do is to coarsen the probabilistic information to a probability mass function $p : \{C_1, ..., C_{12}\}_{\kappa_v} \rightarrow [0, 1]$ on the likelihood-adapted partition of the support of $\kappa_v$. It is important to note that such a mass function can be interpreted as a random set $(\mathcal{D}, p) = \{(C_1, p_1), ..., (C_{12}, p_{12})\}$ which allocates the probability masses $p_j$ exclusively to the atoms $A_j$. Since all the masses are positive, the random set describes a belief function (see Appendix C). However, in the special case of $(\mathcal{D}, p)$ belief and plausibility function collapse to a single additive probability measure (see Equations 3.3 and 3.4). Moreover, the random set $(\mathcal{D}, p)$ fulfils the conditions in Corollary 3.1, and therefore represents a p-box. It was indeed this special case of p-boxes that appeared first to construct lower and upper bounds for the convolution of random variables with unknown dependency (Williamson and Downs, 1990). Hence, we could include the probabilistic information $\{(C_1, p_1), ..., (C_{12}, p_{12})\}$ into our analysis without any need for an adjustment of the mathematical framework.

However, the situation is less favourable because the world ocean heat uptake is not a direct observable. Levitus et al. (2000) reconstructed a spatial field of temperature anomalies in the ocean column down to 3000 m, and deduced a time series of heat content anomalies in the world ocean from it. Simply adding the resulting heat fluxes in the period 1955-96 requires an assumption about the time correlation of the observational errors. This is in essence the same problem that we have encountered above for the SST record, but it is aggravated here because the relative error of the observations is much larger. Levitus et al. (2001) fitted a linear trend model to the heat content anomaly data to arrive at a central estimate of $18.2 \cdot 10^{22}$ J. To the best of our knowledge, they did not specify error bars for this estimate. Annan and Hargreaves (2003) tried to reproduce the linear fit under the assumption of IID deviations from the data, and found a different estimate of $(16.4 \pm 1.9) \cdot 10^{22}$ J for the heat uptake in the period 1955-96. However, they also noted that a linear trend model combined with the assumption of IID errors must be clearly rejected. It cannot explain the reduction of heat content in the 1960s and 80s that interspersed the increasing heat content in the 2nd half of the 20th century. As a remedy, Annan and Hargreaves (2003) proposed a linear trend model together with the assumption of an AR(1) process (see Equation 2.10) for the deviations from the data. Under this assumption they found an estimate of $(19.7 \pm 4.9) \cdot 10^{22}$ J, which exhibited a markedly larger standard deviation. The AR(1) assumption is not perfect, of course. Like surface air temperature, the time series of ocean heat content anomaly will also be modulated by changes in the radiative forcing.

Nevertheless, we will use the heat uptake estimates of Levitus et al. (2001) and Annan and Hargreaves (2003) to construct an imprecise probability for the prior uncertainty about ocean heat diffusivity $\kappa_v$. The central values of the estimates span a range $H_O(1955, 1996) \in [\underline{\mu} = 16.4 \cdot 10^{22}\,\text{J},\ \overline{\mu} = 19.7 \cdot 10^{22}\,\text{J}]$. The standard deviations for the IID and AR(1) assumptions of Annan and Hargreaves (2003) set up a domain $\sigma \in [\underline{\sigma} = 1.9 \cdot 10^{22}\,\text{J}, \overline{\sigma} = 4.9 \cdot 10^{22}\,\text{J}]$. Under the assumption of a normally distributed

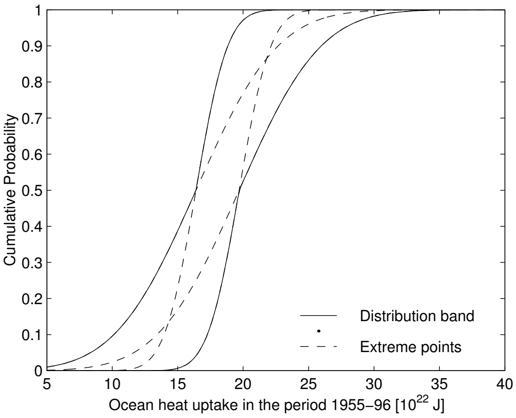
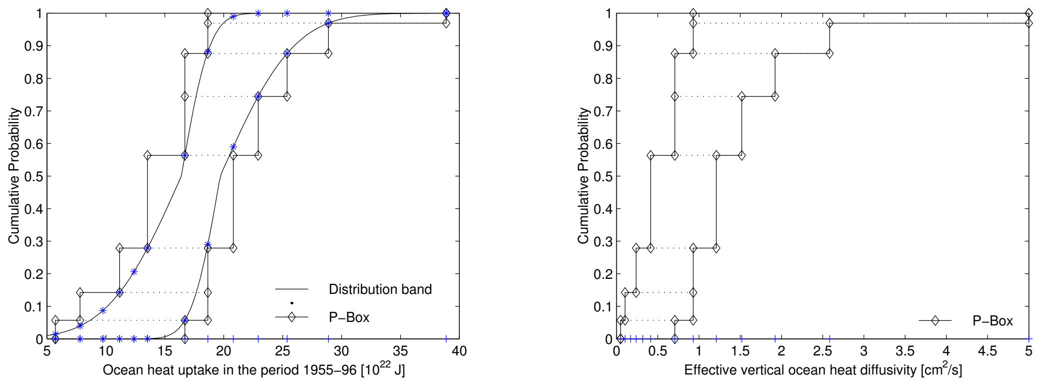


Figure 3.11: Distribution band for ocean heat uptake that is spanned by the four normal distributions constituting the extreme points of the class $\mathcal{N}_{H_O}(\underline{\mu} \leq \mu \leq \overline{\mu}, \underline{\sigma} \leq \sigma \leq \overline{\sigma})$.

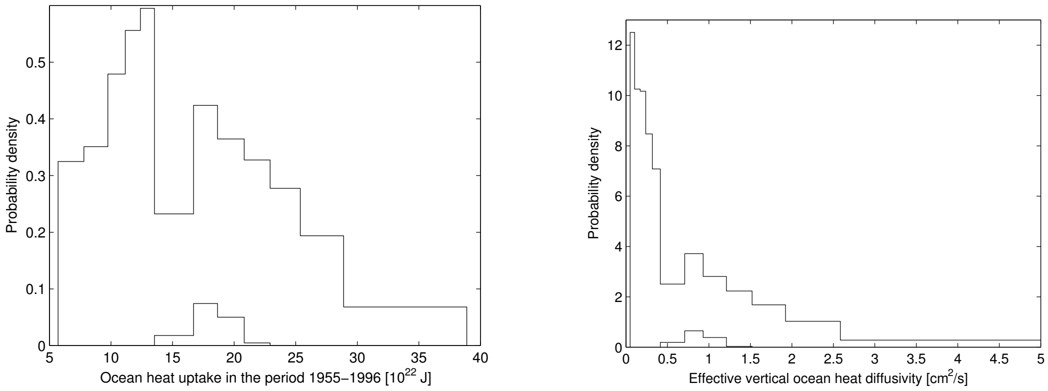
estimation error, these two intervals define a parametric imprecise probability model $\mathcal{N}_{H_O}(\underline{\mu} \leq \mu \leq \overline{\mu}, \underline{\sigma} \leq \sigma \leq \overline{\sigma})$ containing all normal distributions whose mean and standard deviation lie in the respective intervals. Such parametric classes of probabilities have been investigated in the framework of robust Bayesian analysis (Berger, 1993). They cannot be represented by belief functions in general. This does not matter for our analysis, because we want to choose a larger class of probabilities to describe the uncertainty about the heat uptake in 1955-96. In analogy to the case of climate sensitivity and sulphate aerosol forcing, we are looking for the set of all probabilities whose distribution lies between the lower and upper distribution functions of $\mathcal{N}_{H_O}$, and whose mass function dominates the lower mass function of $\mathcal{N}_{H_O}$ on the partition $\{(C_1, \nu_1), ..., (C_{12}, \nu_{12})\}$.

Fig. 3.11 shows the distribution band $\Gamma_{H_O}(\underline{F}, \overline{F})$ that is set up by the extreme points of the normal distributions contained in $\mathcal{N}_{H_O}$. We have used Equation 3.21 to transform the likelihood adapted grid for $\kappa_v$ onto the domain of heat uptake $H_O(1955, 1996)$. Thereby, we have neglected the uncertainty in the SST record that would prevent us from specifying a deterministic mapping between the two quantities. The following analysis proceeds in complete analogy to the cases of climate sensitivity and sulphate aerosol forcing, and therefore will not be commented any further here. Fig. 3.12.a shows the p-box approximation $\Gamma_{H_O}(\underline{SF}, \overline{SF})$ that we have constructed on the likelihood-adapted partition as transformed to the range of heat uptake. It cuts the infinite support of the distribution band at the points $\underline{H}_O = 5.7 \cdot 10^{22}$ J and $\overline{H}_O = 38.9 \cdot 10^{22}$ J corresponding to a domain that ranges over two orders of magnitude of ocean heat diffusivity. A heat uptake above $38.9 \cdot 10^{22}$ J has an upper probability smaller than $10^{-4}$ and therefore can be neglected easily. A heat uptake below $5.7 \cdot 10^{22}$ J exhibits a larger upper probability of 0.015, but will be excluded on the basis of physical considerations. They would imply the unrealistic situation that the mixed layer of 60 m depth takes up



a. Heat uptake - P-Box

b. Heat diffusivity - P-Box



c. Heat uptake - $\varepsilon$-contamination model

d. Heat diffusivity - $\varepsilon$-contamination model

Figure 3.12: P-Box approximation of the continuous distribution band (Panel a), and $\varepsilon$-contamination model (Panel c) for ocean heat uptake in the period 1955-96. Right panels b, d show the associated p-box and $\varepsilon$-contamination model for effective ocean heat diffusivity $\kappa_v$.

more heat over a period of 50 years than the entire interior ocean.

The random set representation of the p-box for the heat uptake in 1955-1996 is also indicated in Fig. 3.12.a. By means of Equation 3.21, it can be retransformed to a random set $(\mathcal{E}, \nu)_{\kappa_v}$ (see Table 3.6). Its associated p-box $\Gamma_{\kappa_v}(\underline{SF}, \overline{SF})$ is shown in Fig. 3.12.b. We proceed in a similar manner for the $\varepsilon$-contamination model that is set up by the extreme points of the normal distributions contained in $\mathcal{N}_{H_O}$. Fig. 3.12.c shows its defining lower mass function on the transformed partition of the heat uptake domain. Fig. 3.12.d depicts the corresponding lower mass function on the domain of vertical diffusivity. The associated random set $(\mathcal{F}, )_{\kappa_v}$ of the $\varepsilon$-contamination model $\Gamma_{\kappa_v}(\underline{p})$ is listed in Table 3.6. Only four out of 12 atoms in the partition $\{(C_1, \nu_1), ..., (C_{12}, \nu_{12})\}$ carry positive lower probability mass $\underline{p}_j > 0.005$ which altogether amounts to $\underline{p} = 0.32$. As before, we construct the random set $(\mathcal{G}, \nu^*)_{\kappa_v}$ for the intersection $\Gamma_{\kappa_v}(\underline{SF}, \overline{SF}, \underline{p})$ of p-box and $\varepsilon$-contamination model by means of Algorithm 3.2. It contains 15 focal elements which are listed in Table 3.6. Fig. 3.13 compares the information contained in

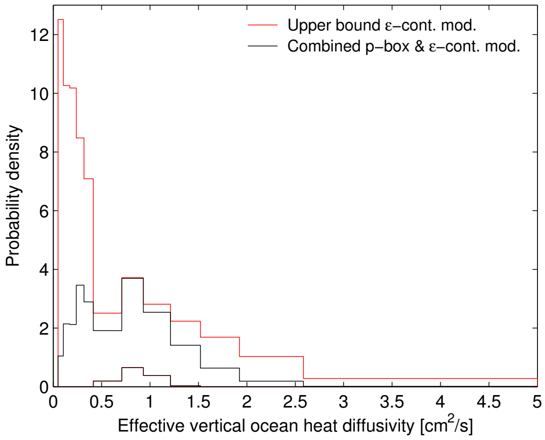


Figure 3.13: Mass band (converted to probability densities) generated from the random set $(\mathcal{G}_{\kappa_v}, \nu^*)$ of the intersection $\Gamma_{\kappa_v}(\underline{SF}, \overline{SF}, \underline{p})$. The upper mass function of the $\varepsilon$-contamination model is shown for comparison.

| $\varepsilon$-contamination model & p-box | | | Intersection of both models | | |
|---|---|---|---|---|---|
| | Focal element (in cm$^2$ s$^{-1}$) | Möbius mass | | Focal element (in cm$^2$ s$^{-1}$) | Möbius mass |
| $C_6$ | $(0.41, 0.71]$ | 0.057 | $G_1$ | $(0.41, 0.71]$ | 0.057 |
| $C_7$ | $(0.71, 0.93]$ | 0.145 | $G_2$ | $(0.71, 0.93]$ | 0.145 |
| $C_8$ | $(0.93, 1.21]$ | 0.108 | $G_3$ | $(0.93, 1.21]$ | 0.108 |
| $C_9$ | $(1.21, 1.52]$ | 0.010 | $G_4$ | $(1.21, 1.52]$ | 0.010 |
| $E_1$ | $(0.05, 0.71]$ | 0.057 | $G_5$ | $(0.10, 0.93]$ | 0.019 |
| $E_2$ | $(0.10, 0.93]$ | 0.085 | $G_6$ | $(0.71, 1.92]$ | 0.050 |
| $E_3$ | $(0.24, 0.93]$ | 0.136 | $G_7$ | $(0.93, 5.00]$ | 0.005 |
| $E_4$ | $(0.41, 1.21]$ | 0.285 | $G_8$ | $(0.24, 1.21]$ | 0.111 |
| $E_5$ | $(0.71, 1.52]$ | 0.181 | $G_9$ | $(0.41, 1.52]$ | 0.146 |
| $E_6$ | $(0.71, 1.92]$ | 0.132 | $G_{10}$ | $(0.05, 0.93]$ | 0.057 |
| $E_7$ | $(0.93, 2.58]$ | 0.093 | $G_{11}$ | $(0.71, 2.58]$ | 0.093 |
| $E_8$ | $(0.93, 5.00]$ | 0.030 | $G_{12}$ | $(0.10, 1.21]$ | 0.066 |
| | | | $G_{13}$ | $(0.41, 1.92]$ | 0.082 |
| | | | $G_{14}$ | $(0.24, 1.52]$ | 0.025 |
| | | | $G_{15}$ | $(0.71, 5.00]$ | 0.025 |

Table 3.6: Random sets $(\mathcal{F}_{\kappa_v}, \underline{p})$ of the $\varepsilon$-contamination model (upper left, focal element $\Omega(\kappa_v)$ with mass 0.68 omitted), $(\mathcal{E}_{\kappa_v}, \underline{p})$ of the p-box (lower left), and $(\mathcal{G}_{\kappa_v}, \nu^*)$ of their intersection $\Gamma_{\kappa_v}(\underline{SF}, \overline{SF}, \underline{p})$ (right side) for ocean heat diffusivity.

$\Gamma_{\kappa_v}(\underline{SF}, \overline{SF}, \underline{p})$ with the $\varepsilon$-contamination model. It can be seen that the addition of the p-box information tightens the upper mass function considerably for small diffusivities $\kappa_v$. For the subsequent analysis, we will adjust the random set $(\mathcal{G}_{\kappa_v}, \nu^*)$ listed in Table 3.4 slightly. The focal element $G_7$ carries only small probability mass 0.005. We remove it $(\mathcal{G}_{\kappa_v}, \nu^*)$ and redistribute its mass on the superset $G_{15} = (0.71\,\mathrm{K}, 5.0\,\mathrm{K}]$ (cmp. Section 3.4.2).

## 3.5 Combination of marginal belief functions

In the preceding section we have constructed three belief function representations for the prior uncertainty about the individual climate model parameters $T_{2x}$, $Q_{\text{S90}}$ and $\kappa_v$. In order to use this information in statistical inferences with the climate model, we need to combine it to a joint imprecise probability model on the full parameter space $\Omega = \Omega(T_{2x}) \times \Omega(\kappa_v) \times \Omega(Q_{\text{S90}})$. This is an important problem, since uncertainty will accumulate from different sources in most assessments of climate change.

In general terms, the problem consists in generating a multivariate uncertainty model from uncertain variables $X = \{X_1, ..., X_n\}$, each of which is described by a structure $\Gamma_{X_i}$ on the real line. If the uncertain quantities $X_i$ were independent and described by marginal probability measures $P_{X_i}$, there would exist a unique way to construct a product measure $P_X$. If they are described by coherent lower probabilities constituting the lower envelopes of the structures $\Gamma_{X_i}$, however, the situation is more complicated. For lower probabilities, the joint lower envelope $\underline{P}_X$ of the independent product depends on the concept of independence that is considered (Walley, 1991).

Couso et al. (2000) have provided an instructive survey of independence concepts in imprecise probability theory and their underlying semantics. The concept of *epistemic independence* (Walley, 1991, Chapter 9) captures the intuitive notion that two uncertain variables are independent if learning about one of them is irrelevant to our knowledge about the other. This mutual irrelevance of variables is encoded in the product rule for combining additive probabilities. However, when combining each $P \in \Gamma_{X_1}(\underline{P}_1)$ with each $Q \in \Gamma_{X_2}(\underline{P}_2)$ by the product rule, we generate a set of joint probabilities $P \times Q \in \mathcal{M}_{X_1 \times X_2}$ that does not conform with the notion of epistemic independence. More precisely, the set $\mathcal{M}_{X_1 \times X_2}$ can be strictly smaller than the set of probabilities that would be desribed by the natural extension of the lower envelopes $\underline{P}_1$ and $\underline{P}_2$ onto the joint space $X_1 \times X_2$ under the assumption of epistemic independence (Walley, 1991, Chapter 9.3). $\mathcal{M}_{X_1 \times X_2}$ has been called *type-1 product* in the literature, and the associated independence concept is often named *strong independence.*

If the uncertainty about the marginal random variables is described by belief functions, there exists an additional concept of independence. Assuming that the uncertainty about the quantities $X_i$ is represented by belief functions $bel_{X_i}$ with associated random sets $(\mathcal{E}_i, \nu_i) = \{(E_{1_i}, \nu_{1_i}),\ ...,\ (E_{k_i}, \nu_{k_i})\}$, $1 \leq i \leq n$. The *random set independent product* is calculated from the marginal random sets $(\mathcal{E}_i, \nu_i)$ by

$$(\mathcal{E}, \nu) := \{\, (E_{l_1 \ldots l_n} = E_{l_1} \times ... \times E_{l_n}, \nu_{l_1 \ldots l_n} = \nu_{l_1} \cdot ... \cdot \nu_{l_n}),\ 1 \leq l_i \leq k_i \,\}\ . \tag{3.22}$$

The random set $(\mathcal{E}, \nu)$ derived from the independent product (3.22) determines a joint belief function $bel_{1 \times \ldots \times n} : \mathcal{R}^n \to [0,1]$ on a field $\mathcal{R}^n$ of the product space $\Omega_{X_1} \times \ldots X_n$. The underlying independence concept associated with Definition (3.22) has been called *random set independence* in the literature. Using the example of random draws from two urns with unknown proportions of red and white balls, Couso et al. (2000) have compared random set independence with the concepts of epistemic and strong independence. Their example showed that a semantics for random set independence cannot be constructed easily. In any case, it would apply only to a very limited set of situations compared with the more natural concept of epistemic independence.

However, here we are less concerned with a semantics of random set independence, but with its relationship to the other independence concepts. Recently, Fetz and Oberguggenberger (2004) and de Cooman (2004) showed that the convex set of joint probabilities encompassed by the belief function $bel_{1\times\ldots\times n}$ always constitutes a superset of the set of probabilities that can be derived from epistemic and strong independence. Hence, random set independence yields the most conservative estimate of the joint uncertainty obtained under the three different independence assumptions. Therefore, it provides us with a simple method to construct a belief function $bel_{1,\ldots,n}$ on the joint uncertainty space that constitutes an outer bound for the convex set of probabilities emerging from the more natural assumption of epistemic independence. It is important to note that the latter cannot be represented by a coherent lower probability in general.

Having assured ourselves that a simple belief function approximation for the independent lower envelope of a set of marginal structures $\Gamma_{bel_1}$, ..., $\Gamma_{bel_n}$ exists, we need to ask whether an independence assumption for the climate model parameters $T_{2x}$, $\kappa_v$ and $Q_{\mathrm{S90}}$ would be justified. We will separate the discussion into a consideration of the physical and the 'informational' dependence of these quantities. Clearly, the factors influencing the radiative properties of sulphate aerosols and their impact on clouds are physically unrelated to the factors influencing climate sensitivity and ocean heat diffusivity. Hence, $Q_{\mathrm{S90}}$ on the one hand, and $\kappa_v$ and $T_{2x}$ on the other hand are physically independent. Moreover, climate sensitivity depends on atmospheric feedbacks related to water vapour and clouds, rather than on mixing schemes in the ocean (e.g. Harvey, 2000, Chapter 9). Therefore, we can assume that $T_{2x}$ does not physically depend on $\kappa_v$. However, the reverse assumption is less clear. Raper et al. (2002) have observed a larger ocean heat uptake per unit temperature increase in models with higher climate sensitivity. The finding could have been an artifact due to the small sample size of models, but could also be related to physical feedbacks of a larger warming onto the ocean. Raper et al. (2002) speculated that the impact of a warmer climate on the stratification of the high latitude ocean and the meridional overturning circulation might provide a possible mechanism for such a feedback. However, these effects are not represented in our simple diffusion ocean energy balance model, since this would require a temperature dependence of the vertical diffusivity parameter $\kappa_v$. Moreover, these feedbacks if existent might have played a minor role for ocean heat diffusivity in the 20th century.

Although the climate model parameters can be assumed to be physically independent, they are linked by our knowledge of, inter alia, the historical temperature record. Comparisons of model results with instrumental temperature data will have a tendency to produce high estimates of $T_{2x}$ for a large negative radiative forcing $Q_{\mathrm{S90}}$ of sulphate aerosols, and vice versa (Forest et al., 2002). The same is true for high climate sensitivities and high ocean heat diffusivities $\kappa_v$. Both types of 'informational' dependencies are clearly visible in the likelihood function that we have derived in Chapter 2 (see Fig. 2.7). These dependencies are also present in the joint probability estimates of Andronova and Schlesinger (2001), Forest et al. (2002) and Knutti et al. (2002), whose marginals were used to construct the belief function representation for our prior uncertainty about $T_{2x}$, $\kappa_v$ and $Q_{\mathrm{S90}}$. As discussed in Section 3.4.1, a dependency structure in the prior uncertainty that matches the shape of the likelihood function would greatly aid the employment of a likelihood-adapted grid. However, we are still lacking the means to construct a belief function representation in the style of the preceding sections directly on a multi-dimensional space. This will require more theoretical work, which is beyond the scope of this analysis.

Therefore, we will take a pragmatic perspective and assume that our prior uncertainty reflects a state of knowledge prior to receiving the instrumental temperature record. Then, 'informational' dependencies are solely introduced by updating the prior uncertainty with the likelihood function $\mathcal{L}(\theta; \hat{T})$ (see Chapter 4). If the emergence of 'informational' dependencies is delegated to the updating process, an independence assumption for the prior uncertainty can be justified on the grounds of physical independence. Hence, we will combine the random sets $(\mathcal{G}, \nu^*)_{T_{2x}}$, $(\mathcal{G}, \nu^*)_{Q_{\mathrm{S90}}}$, and $(\mathcal{G}, \nu^*)_{\kappa_v}$ which represent our prior uncertainty for the parameters $T_{2x}$, $Q_{\mathrm{S90}}$, and $\kappa_v$ by means of the random set independent product (3.22) to a joint random set $(\mathcal{G}, \nu^*)_{\mathrm{CM}}$ on the climate model parameter space $\Omega = \Omega(T_{2x}) \times \Omega(\kappa_v) \times \Omega(Q_{\mathrm{S90}})$. It is associated with a belief function $bel_{\mathrm{CM}} : \mathcal{R}^3 \rightarrow [0, 1]$ that encompasses the convex set of joint probabilities incorporated in our prior uncertainty about the model parameters. By the result of Fetz and Oberguggenberger (2004) and de Cooman (2004) it will include all products of probability measures contained in $\Gamma_{T_{2x}}(\underline{SF}, \overline{SF}, \underline{p})$, $\Gamma_{Q_{\mathrm{S90}}}(\underline{SF}, \overline{SF}, \underline{p})$, and $\Gamma_{\kappa_v}(\underline{SF}, \overline{SF}, \underline{p})$, and more probabilities beyond that. Hence, $bel_{\mathrm{CM}} : \mathcal{R}^3 \rightarrow [0, 1]$ represents a large set of prior probabilities.

The random set $(\mathcal{G}, \nu^*)_{\mathrm{CM}}$ contains $27 \cdot 14 \cdot 16 = 6048$ focal elements. The increase in the number of focal elements by three orders of magnitude mirrors the increase in complexity associated with the "curse of dimensionality". Nevertheless, the Möbius inverse described by $(\mathcal{G}, \nu^*)_{\mathrm{CM}}$ is extremely sparse, because the partition $\{A_1, ..., A_k\}$ of the joint domain $\Omega$ encompasses $k = 20 \cdot 12 \cdot 14 = 3360$ atoms. Hence, the Möbius inverse is defined on a power set with the huge number of $2^{3360}$ elements. In the following, we will use $(\mathcal{G}, \nu^*)_{\mathrm{CM}}$ and the associated belief function as the representation of our prior uncertainty about the climate model parameters.

# Chapter 4

# Updating An Imprecise Probability for the Climate Model

In the preceding chapter we have constructed a belief function for three uncertain climate and forcing model parameters, namely climate sensitivity $T_{2x}$, effective vertical ocean heat diffusivity $\kappa_v$, and the radiative forcing from anthropogenic sulphate aerosols in the year 1990, $Q_{\text{S90}}$. The belief function captures the imprecise information provided by a set of recently published probability density estimates for $T_{2x}$ and $Q_{\text{S90}}$, and by estimates of the heat uptake of the world's oceans in the second half of the 20th century. However, this prior information does not include the *likelihood* $\mathcal{L}(\theta; \hat{T})$ that our energy balance model reproduces the observed temperature record $\hat{T}$ for a given choice $\theta = (T_{2x}, \kappa_v, Q_{\text{S90}})$, which was estimated in Chapter 2.

This chapter is devoted to the derivation of a posterior imprecise probability amalgamating the likelihood and the prior information. We will use *Bayes' rule* in its generalized form for imprecise probabilities to update the prior information with the likelihood of observing the historic temperature record given a model hypothesis $M(\theta)$. As explained in the introduction to Chapter 2, we believe that the Bayesian paradigm suits the requirements of an integrated assessment of climate change well. While the prior information includes the information from the literature about the parameters $\theta$, the likelihood $\mathcal{L}(\theta; \hat{T})$ qualifies our ensemble of model hypotheses $M(\theta)$ in the light of historical temperature observations. Both pieces of information are vital parts that should be included in an assessment of future climate change on the basis of $M(\theta)$.

For given likelihood function $\mathcal{L}(\cdot; \hat{y}) : \Omega \to \mathbb{R}^+$ and prior probability density $\rho : \Omega \to \mathbb{R}^+$, Bayes' rule takes on the form

$$\rho(\theta|\hat{y}) = \frac{\mathcal{L}(\theta; \hat{y})\, \rho(\theta)}{\int_\Omega \mathcal{L}(\theta; \hat{y})\, \rho(\theta)\, d\theta}\,, \tag{4.1}$$

where $\rho(\theta|\hat{y})$ is the posterior probability density for the truth of the hypothesis $M(\theta)$.

The denominator assures that $\rho(\theta|\hat{y})$ is properly normalized. Bayes' rule in its familiar form can be recovered from Equation (4.1) by noting that $\mathcal{L}(\theta;\hat{y}) := \rho(\hat{y}|\theta)$, and $\rho(\hat{y}) := \int_\Omega \mathcal{L}(\theta;\hat{y})\,\rho(\theta)\,d\theta$ constitutes the marginal probability density of observing $\hat{y}$. Then, Bayes' rule is given by $\rho(\theta|\hat{y}) = \rho(\hat{y}|\theta)\,\rho(\theta)\,/\,\rho(\hat{y})$ provided that $\rho(\hat{y}) > 0$. If $\rho(\hat{y}) = 0$, Bayes' rule is not applicable, and the posterior probability density is undefined.

The generalisation of Bayes rule to structures $\Gamma(\underline{P})$ (see Definition C.8) is straightforward. If $\int_\Omega \mathcal{L}(\theta;\hat{y})\,\rho(\theta)\,d\theta > 0$ for all probability densities $\rho \in \Gamma(\underline{P})$, then the posterior convex set of probabilities $\mathcal{M}(\hat{y})$, which is not necessarily a structure any more, is given by

$$\mathcal{M}(\hat{y}) := \left\{ \rho(\cdot|\hat{y}) : \ \exists\, \rho \in \Gamma(\underline{P}) \quad \forall\, \theta \in \Omega \ \ \rho(\theta|\hat{y}) = \frac{\mathcal{L}(\theta;\hat{y})\,\rho(\theta)}{\int_\Omega \mathcal{L}(\theta;\hat{y})\,\rho(\theta)\,d\theta} \right\}. \tag{4.2}$$

Definition (4.2) is the special version for continuous random variables and likelihoods of what has been called *Divisive Conditioning* or *Generalised Bayes Rule* (GBR) in the literature. Walley (1991, Chapter 6.4) used the principle of coherence to establish the GBR in its general form for lower previsions. However, a discussion of this work is beyond the scope of our analysis. Here and in the following, we refer to the special form (4.2) when talking about the GBR.

We discuss the challenges that are posed by an application of the GBR in Section 4.1. Section 4.2 provides an algorithm to update belief functions with a likelihood by means of the GBR and a more restrictive updating rule formulated by Dempster (1968). In Section 4.3 we will apply this algorithm to update the prior belief function for the model parameters with the likelihood function $\mathcal{L}(\theta;\hat{T})$ under both updating rules.

## 4.1 The Generalised Bayes' Rule

Updating convex sets of probabilities with the Generalised Bayes' rule is a thorny issue, in particular when restricting the analysis to the special case of coherent lower probabilities $\underline{P}$ on event spaces (see Definition C.7) or even to belief functions. Two major difficulties arise from the application of the GBR: the *dilation* of imprecise posterior probabilities (Section 4.1.1) and the generation of convex sets of probabilities that cannot be described by a coherent lower probability any more (Section 4.1.2).

### 4.1.1 Dilation

Application of the GBR can generate very large, and therefore uninformative, sets of posterior probabilities $\mathcal{M}(\hat{y})$. This is due to the normalisation of the posterior probability density $\rho(\theta|\hat{y})$ with the marginal density $\rho(\hat{y})$ of observing the data $\hat{y}$. Since $\rho(\hat{y})$ might vary substantially across the set of priors, the bounds on the posterior probability of an event $A$ can be *dilated* by an observation $\hat{y}$, i.e.,

$$\underline{P}(A|\hat{y}) \le \underline{P}(A) \le \overline{P}(A) \le \overline{P}(A|\hat{y}), \tag{4.3}$$

where at least one of the outer inequalities is strict. Consider the example of a universal set $\Omega = \{e_1, e_2, e_3\}$ and a likelihood function $\mathcal{L}(e_1; \hat{y}) = 0$, $\mathcal{L}(e_2; \hat{y}) = 1$, $\mathcal{L}(e_3; \hat{y}) = 10$. Let the imprecise prior probability assessment be given by $P(\{e_1\}) = 1/2$ and $P(\{e_2, e_3\}) = 1/2$. It is fully described by the two extreme points $p_1 = \{1/2, 1/2, 0\}$ and $p_2 = \{1/2, 0, 1/2\}$. We can infer the lower and upper posterior probability for the event $\{e_2\}$ by only updating the two extreme points with the likelihood function (Walley, 1991, Chapter 8.4.8). Application of Bayes' rule yields $\underline{P}(\{e_2\}|\hat{y}) = \underline{P}(\{e_2\}) = 0$, and $\overline{P}(\{e_2\}) = 1/2 < \overline{P}(\{e_2\}|\hat{y}) = 1$. Although the event $e_2$ receives only a low likelihood in the light of the observation $\hat{y}$, its posterior upper probability is twice as large than its prior upper probability. This is due to the fact that the prior includes a probability assessment under which the much more likely event $e_3$ cannot occur at all. Hence, the incorporation of the likelihood will aggravate the imprecision in the prior assessment if it varies strongly over events $A \subset \Omega$, where the distribution of prior probability is (largely) unknown.

We will see below that prior information described by belief functions is particularly prone to this problem. Therefore, we need to take utmost care to limit the imprecision of the prior in areas of large changes of the likelihood function. We have done so by adapting the resolution of the grid that discretises the uncertain parameter space $\Omega = \mathbb{R}(T_{2x}) \times \mathbb{R}(\kappa_v) \times \mathbb{R}(Q_{\mathrm{S90}})$ to the shape of the likelihood function. However, the computational costs pose a natural limit on the improvement to be achieved in this way, in particular for multi-dimensional uncertainty spaces. An alternative would be to use an updating rule that implicitely penalises prior probabilities which imply a low probability (density) of observing the data $\hat{y}$, i.e., for which $\int_\Omega \mathcal{L}(\theta; \hat{y})\, \rho(\theta)\, d\theta \ll \sup_{\rho' \in \mathcal{M}(\hat{y})} \int_\Omega \mathcal{L}(\theta; \hat{y})\, \rho'(\theta)\, d\theta$. Such a candidate is *Dempster's rule of conditioning* that was originally proposed for updating belief functions (Dempster, 1968). It was shown by Gilboa and Schmeidler (1993, Theorem 3.3) that a generalised version of Dempster's rule (to arbitrary coherent lower probabilities $\underline{P}$) coincides with the so-called maximum likelihood classical update rule (if the joint probability on the space $\Omega \times Y$ is 2-monotone). Under this rule, only those prior probabilities in the associated structure $\Gamma(\underline{P})$ are retained and updated via Bayes' rule which maximise the probability (density) of observing $\hat{y}$. Hence, application of Dempster's rule always yields a more informative posterior imprecise probability model than the GBR. If the prior information is described by a precise probability, both rules collapse to the well-known Bayes' rule (4.1). In this analysis, we will calculate the posterior imprecise probabilities under both updating rules and compare them with each other.

It is important to note that the possibility of dilated bounds on the posterior probability $P(A|\hat{y})$ is not confined to the use of simplified imprecise probability models described by coherent lower probabilities, but part and parcel of the application of the GBR in general. If the observation $\hat{y}$ is surprising in the light of the prior information, dilation of the posterior bounds on $P(A|\hat{y})$ will be a reasonable thing to happen. This might also be the case for the example presented above, where a large change of likelihood in an area of large imprecision in the prior information is responsible for the occurrence of dilation. However, Seidenfeld and Wasserman (1993) and Herron et al. (1997) have identified instances of dilation that are much more severe. They showed that the bounds on the posterior probability $P(A|B)$ can be dilated for any event $B$ from a (finite or countable) partition $\mathcal{B}$ of the space of possible observations $Y$. This means that no matter what $B \in \mathcal{B}$ will be observed, we will know less about $A$ than we knew before. As a result, the value of information for the experiment $\mathcal{B}$ is negative, and a Bayesian decision maker should be inclined to a pay a positive price *not to conduct* this experiment. This contradicts the commonly accepted wisdom that new information is always welcome, and any experiment available at no additional costs should be performed. The occurrence of dilation has spurred an ongoing debate about whether and how this counterintuitive consequence of the GBR should be accommodated in a decision-theoretical framework for updating imprecise beliefs. In the context of climate research, a dilation of climate model uncertainty no matter what type of climate observation is received would be clearly unacceptable. However, we speculate that the space of climate observations is sufficiently rich to prevent such an instance of dilation for an entire experiment $\mathcal{B}$.

Seidenfeld and Wasserman (1993) and Herron et al. (1997) have established their dilation result for a variety of frequently used imprecise probability models including the $\epsilon$-contamination model and the atomic lower and upper probability model. Moreover, the probability ratio model is the only model class that has been shown to be dilation immune. Since this model class also allows for a simple implementation of the GBR, it provides an important alternative to the model class considered here. However, since it cannot be captured in terms of a coherent lower probability (see Appendix C), it does not fit our choice of Möbius representation that was motivated in Section 3.1. Therefore, we do not pursue the probability ratio model any further here, but earmark it as an important area of future research. In this analysis, we rely on belief function models, and thus have to watch out for dilation when applying the GBR. Dempster's rule of conditioning provides an alternative that is dilation immune.

### 4.1.2 Application to coherent lower probabilities

As was discussed in Appendix C, coherent lower probabilities $\underline{P}$ and their associated structures $\Gamma(\underline{P})$ constitute only a special class of imprecise probabilities. This lack of generality becomes apparent when updating the structure $\Gamma(\underline{P})$ via the GBR as defined in Equation (4.2). In general, the resulting convex set of posterior probabilities $\mathcal{M}(\hat{y})$ is not a structure any more. Its lower envelope $\underline{P}(\cdot|\hat{y})$ describes a structure $\Gamma(\underline{P}(\cdot|\hat{y}))$ of posterior probabilities that can be strictly larger than $\mathcal{M}(\hat{y})$. Hence, the restriction to the posterior coherent lower probability $\underline{P}(\cdot|\hat{y})$ will include a loss of information. Even if we accept this loss of information, we need to be aware of the consequence that the updating operation on coherent lower probabilities is non-commutative (Chrisman, 1995). The posterior lower probability will depend on whether we update first on observation $\hat{y}_1$ and then on observation $\hat{y}_2$, or vice versa ($\underline{P}(\cdot|\hat{y}_1, \hat{y}_2) \neq \underline{P}(\cdot|\hat{y}_2, \hat{y}_1)$. Since updating is often performed incrementally on a sequence of observations $(\hat{y}_1, \hat{y}_2, ...)$, this can be a severe problem. Chrisman (1995) has proposed a remedy, which we will utilise in Section 4.2.

Notwithstanding these difficulties, the restriction to lower envelopes of convex sets of probabilities can improve the mathematical tractability of the updating task. Wasserman and Kadane (1990) provide lower bounds on the posterior lower probability $\underline{P}(\cdot|\hat{y}) : \mathcal{A}(\Omega) \rightarrow [0,1]$ from updating a prior lower probability $\underline{P} : \mathcal{A}(\Omega) \rightarrow [0,1]$ with a likelihood function $\mathcal{L}(\cdot;\hat{y}) : \Omega \rightarrow \mathbb{R}_0^+$. To simplify the presentation, we adopt their notation in the following. Let $I_A : \Omega \rightarrow \{0,1\}$ be the indicator function of the event $A$, i.e. $I(\theta) = 1$, if $\theta \in A$, and $I(\theta) = 0$ otherwise. Let

$$E(\mathcal{L}_A;\rho) := \int_\Omega I_A(\theta)\,\mathcal{L}(\theta;\hat{y})\,\rho(\theta)\,d\theta = \int_A \mathcal{L}(\theta;\hat{y})\,\rho(\theta)\,d\theta$$

be the expected value of the likelihood function $\mathcal{L}(\cdot;\hat{y})$ restricted to the event $A \in \mathcal{A}(\Omega)$ w.r.t. the prior probability density $\rho : \Omega \rightarrow \mathbb{R}_0^+$. Since $\mathcal{L}(\theta;\hat{y}) := \rho(\hat{y}|\theta)$ is a conditional probability, the expected value $E(\mathcal{L}_A;\rho)$ can be also regarded as the joint probability density of the event $\rho(\hat{y} \times A)$. Likewise, let $E(L;\rho) := E(L_\Omega;\rho)$ be the expectation value of the likelihood (i.e., the marginal probability density of the observation $\hat{y}$, $\rho(\hat{y}) := \rho(\hat{y} \times \Omega)$). According to Bayes' rule (4.1), the posterior probability of the event $A \in \mathcal{A}(\Omega)$ is given by

$$P(A|\hat{y};\rho) = \frac{E(\mathcal{L}_A;\rho)}{E(\mathcal{L}_A;\rho) + E(\mathcal{L}_{A^c};\rho)} \quad \text{if} \quad E(\mathcal{L};\rho) > 0\,. \tag{4.4}$$

Let $\underline{E}_\Gamma(\mathcal{L}_A) := \inf_{\rho \in \Gamma(\underline{P})} E(\mathcal{L}_A;\rho)\,,\quad \overline{E}_\Gamma(\mathcal{L}_A) := \sup_{\rho \in \Gamma(\underline{P})} E(\mathcal{L}_A;\rho)$

be the lower and upper bounds of the expectation $E(\mathcal{L}_A)$ on the structure $\Gamma(\underline{P})$ which constitutes the imprecise prior probability model. Moreover, let

$$\underline{C}_{\underline{P}}(\mathcal{L}_A) := \int_0^\infty \underline{P}(\{\theta | I_A(\theta)\,\mathcal{L}(\theta;\hat{y}) > x\})\,dx\,, \tag{4.5}$$

$$\begin{aligned} \overline{C}_{\overline{P}}(\mathcal{L}_A) &:= \int_0^\infty \overline{P}(\{\theta | I_A(\theta)\,\mathcal{L}(\theta;\hat{y}) > x\})\,dx \qquad (4.6)\\ &= \int_0^\infty (\,1 - \underline{P}(\{\theta | I_A(\theta)\,\mathcal{L}(\theta;\hat{y}) \le x\})\,)\,dx \end{aligned}$$

be the lower and upper *Choquet integral* (see Definition C.11) of the $A$-restricted likelihood function $I_A\,\mathcal{L}(\cdot;\hat{y})$ w.r.t the coherent lower probability $\underline{P}$, and its conjugate upper probability $\overline{P}$, respectively. Given these definitions, Wasserman and Kadane (1990) show that the infimum of the posterior probabilities $P(A|\hat{y};\rho)$ in the set $\mathcal{M}(\hat{y})$, is bounded from below by

$$\underline{P}(A|\hat{y}) := \inf_{\rho_{\hat{y}} \in \mathcal{M}(\hat{y})} P(A|\hat{y};\rho) \ge \frac{\underline{E}_\Gamma(\mathcal{L}_A)}{\underline{E}_\Gamma(\mathcal{L}_A) + \overline{E}_\Gamma(\mathcal{L}_{A^c})} \ge \frac{\underline{C}_{\underline{P}}(\mathcal{L}_A)}{\underline{C}_{\underline{P}}(\mathcal{L}_A) + \overline{C}_{\overline{P}}(\mathcal{L}_{A^c})}\,, \tag{4.7}$$

when the denominators of the ratios are larger than zero. The left inequality is fairly obvious, since no prior probability in the structure $\Gamma(\underline{P})$ can yield expectation values $E(\mathcal{L}_A) < \underline{E}_\Gamma(\mathcal{L}_A)$, or $E(\mathcal{L}_A^c) > \overline{E}_\Gamma(\mathcal{L}_A^c)$. The right inequality stems from the fact that the lower (upper) Choquet integral w.r.t. to a coherent lower (upper) probability provides an outer bound on the lower (upper) expectation value w.r.t. to the associated structure of probabilities.

Wasserman and Kadane (1990) move on to show that the inequalities in Expression (4.7) become equalities if, and only if, $\underline{P}$ is a 2-monotone Choquet capacity (see Definition C.10). The right equality follows from the fact that the lower (upper) expectation equals the lower (upper) Choquet integral if, and only if, the coherent lower probability $\underline{P}$ is 2-monotone (see Appendix C). Establishing the left equality requires to find a prior probability density $\rho \in \Gamma(\underline{P})$ with $E(\mathcal{L}_A) = \underline{E}(\mathcal{L}_A)$ and $E(\mathcal{L}_A^c) = \overline{E}(\mathcal{L}_A^c)$. Such a probability density exists if and only if $\underline{P}$ is 2-monotone.

Hence, Wasserman and Kadane (1990) provide a closed form expression to calculate the lower posterior probability for arbitrary events $A \in \mathcal{A}(\Omega)$ directly from knowledge of the likelihood function $\mathcal{L}(\cdot;\hat{y})$ and the prior lower probability $\underline{P}$. The calculation is further simplified if $\underline{P}$ is characterised by a Möbius inverse (see Definition C.12) with a finite number of focal elements $E_1, ..., E_n$ with non-zero Möbius assignments $\nu_i > 0$ for $1 \le i \le n$. Let $(\mathcal{E}, \nu) := \{\, (E_1, \nu_1 := \nu(E_1)), ..., (E_n, \nu_n := \nu(E_n)) \,\}$ the collection of focal elements and their associated Möbius assignments which fully characterises the Möbius inverse.

Then, the posterior lower probability is given by

$$\underline{P}(A|\hat{y}) \;\ge\; \frac{\sum\limits_{E_i \subseteq A} \inf\limits_{\theta \in E_j} \mathcal{L}(\theta;\hat{y})\;\nu_i}{\sum\limits_{E_i \subseteq A} \inf\limits_{\theta \in E_j} \mathcal{L}(\theta;\hat{y})\;\nu_i + \sum\limits_{E_j \cap A^c \neq \emptyset} \sup\limits_{\theta \in E_j \cap A^c} \mathcal{L}(\theta;\hat{y})\;\nu_j}\,, \tag{4.8}$$

where equality holds if and only if $\underline{P}$ is 2-monotone. Expression (4.8) can be deduced from rewriting the Choquet integral in terms of a summation of Möbius assignments ($\underline{C}_{\underline{P}}(f) = \sum_{E_i} \inf_{\theta \in E_i} f(\theta)\,\nu_i\,,\ \overline{C}_{\overline{P}}(f) = \sum_{E_i} \sup_{\theta \in E_i} f(\theta)\,\nu_i$, see Appendix C).

We note that bounds for the posterior upper probability $\overline{P}(\hat{y})$ can be specified in a similar fashion. Since $\overline{P}(\hat{y})$ is the conjugate set function of the posterior lower probability, i.e. $\overline{P}(A|\hat{y}) = 1 - \underline{P}(A^c|\hat{y})$ for all $A \in \mathcal{A}(\Omega)$, we find immediately

$$\overline{P}(A|\hat{y}) \;\le\; \frac{\overline{C}_{\overline{P}}(\mathcal{L}_A)}{\overline{C}_{\overline{P}}(\mathcal{L}_A) + \underline{C}_{\underline{P}}(\mathcal{L}_{A^c})} = \frac{\sum\limits_{E_i \cap A} \sup\limits_{\theta \in E_i \cap A} \mathcal{L}(\theta;\hat{y})\;\nu_i}{\sum\limits_{E_i \cap A} \sup\limits_{\theta \in E_i \cap A} \mathcal{L}(\theta;\hat{y})\;\nu_i + \sum\limits_{E_j \subseteq A^c} \inf\limits_{\theta \in E_j} \mathcal{L}(\theta;\hat{y})\;\nu_j}\,,$$

where equality holds if and only if $\overline{P}(\hat{y})$ is 2-alternating (see Definition C.10). Indeed, Wasserman and Kadane (1990) have chosen to state their results for the upper posterior probability. Since we have preferred to work with the lower probability in this analysis, we have transferred the expressions accordingly. Due to the conjugacy between lower and upper probability, it suffices to consider just one set function. The choice among the two is a matter of mere convention.

## 4.2 Updating belief functions with a likelihood

We will use Expressions (4.7) and (4.8) as a guideline for updating belief functions $bel : \mathcal{A}(\Omega) \to [0,1]$ with a likelihood function $\mathcal{L}(\cdot;\hat{y}) : \Omega \to \mathbb{R}_0^+$. This problem has received little attention in the literature. Most discussions of statistical inference with belief functions operate on a joint space $\Omega' = \Omega \times Y$, with no explicit distinction between model hypotheses $\theta \in \Omega$ and observations $y \in Y$. Here the problem is to find a conditional lower probability $\underline{P}(\cdot|B) : \mathcal{A}(\Omega'_{|B}) \to [0,1]$ from the unconditional lower probability $\underline{P} : \mathcal{A}(\Omega') \to [0,1]$ after $B \in \Omega'$ has occurred. We will refer to this type of statistical inference as *conditioning* on an event $B$ in order to distinguish it from the more special case of *updating* with an observation $y$.

In the general framework, the application of the Generalised Bayes' rule to a coherent lower probability $\underline{P}$ yields the conditional lower probability

$$\underline{P}(A|B) \geq \frac{\underline{P}(A \cap B)}{\underline{P}(A \cap B) + \overline{P}(A^c \cap B)}\,, \tag{4.9}$$

where equality holds if and only if $\underline{P}$ is 2-monotone (see, e.g., Chrisman, 1995). Hence, equality holds in particular if $\underline{P}$ is a belief function. There exists a different updating rule which is most widely used in the statistical reasoning with belief functions. This rule, called *Dempster's rule of conditioning*, is defined by (Dempster, 1968)

$$\overline{P}(A||^*B) := \frac{\overline{P}(A \cap B)}{\overline{P}(B)} \quad \Rightarrow \quad \underline{P}(A||^*B) = \frac{\underline{P}((A \cap B) \cup B^c) - \underline{P}(B^c)}{1 - \underline{P}(B^c)} \tag{4.10}$$

As discussed above, Gilboa and Schmeidler (1993) have identified Dempster's rule with the maximum likelihood update rule (if the joint probability on the space $\Omega \times Y$ is 2-monotone). Note that we have used different symbols for the conditional lower and upper probabilities that emerge from the two rules to mark their difference.

If $\underline{P} : \mathcal{A}(\Omega) \to [0,1]$ is a belief function, both conditioning rules can be phrased in terms of the Möbius inverse $\nu : \mathcal{A}(\Omega) \to [0,1]$ of $\underline{P}$. Let $(\mathcal{E},\nu) := \{\,(E_1,\nu_1)\,, ..., (E_n,\nu_n)\,\}$ be the random set that characterises the Möbius inverse $\nu$. Then the random set $(\mathcal{E},\nu)_{||^*B}$ of the Dempster-conditioned lower probability $\underline{P}(\cdot||^*B) : \mathcal{A}(\Omega'_{|B}) \to [0,1]$ can be calculated directly from $(\mathcal{E},\nu)$ by (Dempster, 1968)

$$\nu_{||^*B}(A) = \frac{\sum_{E_i \cap B = A} \nu_i}{\sum_{E_{i'} \cap B \neq \emptyset} \nu_{i'}}\,. \tag{4.11}$$

It can be seen that $\nu_{||^*B}(A) \geq 0$ for all $A \in \Omega$, if $(\mathcal{E},\nu)$ contains only positive Möbius assignments. Therefore, $\underline{P}(\cdot||^*B)$ is a belief function if $\underline{P}$ was a belief function.

Jaffray (1992, Theorem 1) presented an analogous relation between the unconditional Möbius assignment $\nu$ and the conditional $\nu_{|B}$ that follows from the application of the GBR described in Equation (4.9). However, the resulting formula is much more intricate than for Dempster conditioning. Although the conditional Möbius inverse $\nu_{|B}$ defines another belief function, the number of focal elements with $\nu_{|B} > 0$ can increase exponentially. Hence, even if we established an imprecise prior probability model with a sparse Möbius inverse, this achievement would be nullified when updating it via the GBR. Moreover, Jaffray (1992) showed that the conditional belief function $bel(\cdot|B) : \mathcal{A}(\Omega'_{|B}) \to [0,1]$ so obtained does not fully describe the updated set of probabilities $\mathcal{M}(B)$ that results from conditioning each probability in the structure $\Gamma(bel)$. As a consequence, the conditional Möbius inverse depends on the order of conditioning events $B_1, B_2, ...$ and therefore, is not applicable to incremental updating.

Chrisman (1995) proposed a remedy for the non-commutativity as well as exponential explosion of focal elements that plague the conditional Möbius assignment $\nu_{|B}$ of Jaffray (1992). He observed that if the lower probability for the occurrence of $B$ is positive ($\underline{P}(B) > 0$), then

$$\frac{\underline{P}(A \cap B)}{\underline{P}(A \cap B) + \overline{P}(A^c \cap B)} = \frac{\underline{P}(A||_* B)\, \underline{P}(B)}{\underline{P}(A||_* B)\, \underline{P}(B) + \overline{P}(A^c||^* B)\, \overline{P}(B)} , \tag{4.12}$$

where $\overline{P}(A^c||^* B)$ is the Dempster-conditioned upper probability from Equation (4.10), and $\underline{P}(A||_* B)$ the lower probability from a dual conditioning rule defined by

$$\underline{P}(A||_* B) = \frac{\underline{P}(A \cap B)}{\underline{P}(B)} . \tag{4.13}$$

As for Dempster's rule of conditioning, the Möbius inverse of $P(A||_* B)$ can be easily derived from the unconditionial Möbius assignment (Chrisman, 1995). It is

$$\nu_{||_* B}(A) = \frac{\sum_{E_i = A} \nu_i}{\sum_{E_{i'} \subseteq B} \nu_{i'}} . \tag{4.14}$$

Hence, knowledge of $\underline{P}(\cdot||_* B)$, $\underline{P}(\cdot||^* B)$, $\underline{P}(B)$ and $\overline{P}(B)$ suffice to determine the conditional lower probability $\underline{P}(\cdot|B)$ for arbitrary $A \in \mathcal{A}(\Omega'_{|B})$ by means of Equation (4.12). Since the Möbius inverses of $\underline{P}(\cdot||_* B)$ and $\underline{P}(\cdot||^* B)$ can be easily computed via Equations (4.11) and (4.14), an exponential explosion of the number of focal elements is avoided. Moreover, Chrisman (1995) shows that both $\underline{P}(\cdot||_* B)$ and $\underline{P}(\cdot||^* B)$ can be updated incrementally. As a consequence, the resulting estimate for the lower bound $\underline{P}(\cdot|B)$ will be independent of the order of conditioning events.

In the following, we will try to adopt the approach of Chrisman (1995) for the case of updating a prior belief function with a likelihood function that was discussed by Wasserman and Kadane (1990). The similarity between Expression (4.7) for the likelihood update and Expression (4.9) for conditioning on the joint probability space is striking. We reformulate Expression (4.7) in analogy to Equation (4.12) by

$$\frac{\underline{C}_{\underline{P}}(\mathcal{L}_A)}{\underline{C}_{\underline{P}}(\mathcal{L}_A)+\overline{C}_{\overline{P}}(\mathcal{L}_{A^c})} \;=\; \frac{\underline{P}(A||_*\hat{y})\,\underline{C}_{\underline{P}}(\mathcal{L})}{\underline{P}(A||_*\hat{y})\,\underline{C}_{\underline{P}}(\mathcal{L})+\overline{P}(A^c||^*\hat{y})\,\overline{C}_{\overline{P}}(\mathcal{L})}\,,$$

where we have defined

$$\underline{P}(A||_*\hat{y}) := \frac{\underline{C}_{\underline{P}}(\mathcal{L}_A)}{\underline{C}_{\underline{P}}(\mathcal{L})}\,, \qquad \overline{P}(A||^*\hat{y}) := \frac{\overline{C}_{\overline{P}}(\mathcal{L}_A)}{\overline{C}_{\overline{P}}(\mathcal{L})}\,, \tag{4.15}$$

with $\overline{C}_{\overline{P}}(\mathcal{L}) := \overline{C}_{\overline{P}}(\mathcal{L}_\Omega)$, and $\underline{C}_{\underline{P}}(\mathcal{L}) := \underline{C}_{\underline{P}}(\mathcal{L}_\Omega)$. Obviously, the ratios in Equation (4.15) are only defined if the denominators are non-zero. In the following, we will always assume $\overline{C}_{\overline{P}}(\mathcal{L}) \geq \underline{C}_{\underline{P}}(\mathcal{L}) > 0$.

By definition, it is $\underline{P}(\Omega||_*\hat{y}) = \overline{P}(\Omega||^*\hat{y}) = 1$. Since $\mathcal{L}(\theta|\hat{y}) \geq 0$ for all $\theta \in \Omega$, we also have $\underline{P}(\cdot||_*\hat{y}) \geq 0$ and $\overline{P}(\cdot||^*\hat{y}) \geq 0$ for all $A \in \mathcal{A}(\Omega)$. By the same token, both set functions are monotone, i.e., $\underline{P}(A||_*\hat{y}) \leq \underline{P}(B||_*\hat{y})$ and $\overline{P}(A||^*\hat{y}) \leq \overline{P}(B||^*\hat{y})$ for $A \subseteq B$. Hence, they are *normalised Choquet capacities* (see Definition C.9).

We assume in the following that the imprecise prior probability $\underline{P}$ to be updated constitutes a belief function $bel : \mathcal{A}(\Omega) \to [0,1]$ associated with a random set $(\mathcal{E},\nu) := \{\,(E_1,\nu_1)\,,...,(E_n,\nu_n)\,\}$. Then, we can show that $\underline{P}(A||_*\hat{y})$ is a belief function, whose random set can be calculated immediately from $(\mathcal{E},\nu)$ and knowledge of

$$\mathcal{L}_{*,i} = \inf_{\theta\in E_i}\ \mathcal{L}(\theta;\hat{y}) \quad \text{for all} \quad 1 \leq i \leq n\,. \tag{4.16}$$

**Lemma 4.1** *Let* $bel : \mathcal{A}(\Omega) \to [0,1]$ *be a belief function with random set* $(\mathcal{E},\nu)$ *as described above. Let* $\mathcal{L}_{*,i}$ *be the infimum likelihood that* $\hat{y}$ *is observed over all hypotheses* $\theta \in E_i$ *(see Definition 4.16). Let* $\underline{C}_{\underline{P}}(\mathcal{L}_A)$ *be the Choquet integral of the A-restricted likelihood function w.r.t. to bel as given in Definition (4.5). Assume that* $\underline{C}_{\underline{P}}(\mathcal{L}) > 0$.

*Let* $\underline{P}(\cdot||_*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ *be the set function defined in Equation (4.15). Let* $bel(\cdot||_*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ *be a belief function associated with the random set*

$$(\mathcal{E},\nu)_{||_*\hat{y}} := \left\{ \left(E_1, \nu_{*,1} := \frac{\mathcal{L}_{*,1}\,\nu_1}{\sum_{i=1}^n \mathcal{L}_{*,i}\,\nu_i}\right), ..., \left(E_n, \nu_{*,n} := \frac{\mathcal{L}_{*,n}\,\nu_n}{\sum_{i=1}^n \mathcal{L}_{*,i}\,\nu_i}\right) \right\}\,. \tag{4.17}$$

*Then, it is* $\underline{P}(\cdot||_*\hat{y}) = bel(\cdot||_*\hat{y})$.

The short proof of Lemma 4.1 is given in Appendix D. Here, we only note that $bel(\cdot||_*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ is indeed a belief function, since $\sum_{i=1}^n \nu_{*,i} = 1$, and $\nu_{*,i} \geq 0$ for all $1 \leq i \leq n$. Comparison of Expression (4.17) with Equation (4.14) reveals that Lemma (4.1) describes the updating equivalent to the dual conditioning rule (4.13). Therefore, we will call the prescription for constructing $bel(\cdot||_*\hat{y})$ specified in Lemma (4.1) 'dual updating rule' in the following.

Likewise, the set function $\overline{P}(\cdot||^*\hat{y})$ constitutes the updating equivalent to the conditional upper probability $\overline{P}(\cdot||^*B)$ that is obtained from Dempster's rule of conditioning (4.10). We now investigate whether we can identify a 'Dempster updating rule' in the style of Expression (4.11) that would allow us to construct $\overline{P}(\cdot||^*\hat{y})$ directly from a combination of the likelihood information with the Möbius inverse of the prior belief function. Obviously, the situation is more complicated than for the case of the dual updating rule, because

$$\overline{P}(A||^*\hat{y}) := \frac{\overline{C}_{\overline{P}}(\mathcal{L}_A)}{\overline{C}_{\overline{P}}(\mathcal{L})} = \frac{\sum\limits_{E_i\cap A\neq\emptyset} \sup\limits_{\theta\in E_i\cap A} \mathcal{L}(\theta;\hat{y})\ \nu_i}{\sum_{i'=1}^{n} \sup\limits_{\theta\in E_{i'}} \mathcal{L}(\theta;\hat{y})\ \nu_{i'}}$$

depends on the supremum of the likelihood over the sets $E_i \cap A$. Since these supremum likelihood values will differ among the events $A$ under consideration, a general rule for combining prescribed likelihood values with the Möbius masses of the focal elements cannot be available. However, we can approximate the likelihood function by an outer step function on a finite partition $\{A_1, ..., A_k\}$ of $\Omega$, whose power set contains the focal elements $E_i \in \mathcal{E}$. Then, it is indeed possible to specify a general rule for constructing an approximation of $\overline{P}(A||^*\hat{y})$ that is based on the supremum likelihood values

$$\mathcal{L}^*_j = \sup_{\theta\in A_j} \mathcal{L}(\theta;\hat{y})\ , \qquad \text{for all }\ 1\le j\le k\ , \tag{4.18}$$

over the $k$ atoms of the partition. We will discuss below that the approximation works the better the less the likelihood function $\mathcal{L}(\cdot;\hat{y})$ varies over the atoms. This property of the updating rule has motivated us to choose a likelihood-adapted partition of the uncertain parameter space $\Omega = \mathbb{R}(T_{2x}) \times \mathbb{R}(\kappa_v) \times \mathbb{R}(Q_{\mathrm{S90}})$ in Section 3.4.1.

Knowledge of $\mathcal{L}^*_j$ for $1 \le j \le k$ and the random set $(\mathcal{E}, \nu)$ of the prior belief function enables us to construct the random set $(\mathcal{E}, \nu)_{||^*\hat{y}}$ of a belief function $bel(\cdot||^*\hat{y}) : \mathcal{A}(\Omega) \to [0, 1]$ by the following algorithm.

**Algorithm 4.1** *1. Order the likelihoods $\mathcal{L}^*_j$ from small to large. Let $\{\mathcal{L}^*_1, ..., \mathcal{L}^*_l\}$ be the ordered set with $\mathcal{L}^*_i < \mathcal{L}^*_j$ for $i < j$. Obviously, $l \le k$.*

*2. Construct all $\mathcal{L}^*_j$-cuts of the likelihood function by generating $C_j = \bigcup_{j'\,|\,\mathcal{L}^*_{j'}\ge\mathcal{L}^*_j} A_{j'}$ for all $1 \le j \le l$. Obviously, the cuts $C_j$ are nested with $C_j \supset C_{j'}$, if $j < j'$. The largest cut is $C_1 = \Omega$, and the smallest cut $C_l$ is constituted by the atom(s) over which the maximum likelihood $\mathcal{L}^*_l$ is attained.*

*3. For each focal element $E_i$ in $\mathcal{E} = \{E_1, ..., E_n\}$, construct the chain of sets $E_{ij} = E_i \cap C_j$ with $1 \le j \le l$. Obviously, the $E_{ij}$ are nested with $E_{ij'} \supseteq E_{ij}$, if $j' < j$. The largest element is $E_{i1} = E_i$. Let $l(i)$ be the largest $j$ for which $E_{ij} \neq \emptyset$.*

*4. The random set $(\mathcal{E}, \nu)_{||^*B}$ contains the sets $E_{ij}$, with $1 \le i \le n$ and $1 \le j \le l(i)$, as focal elements. Their Möbius masses are generated as follows:*

```
For $i = 1$ to $n$
    For $j = 1$ to $l(i)$
        Assign Möbius mass $\nu^*_{ij} = \frac{(\mathcal{L}^*_j - \mathcal{L}^*_{j-1})\,\nu_i}{\sum_{i'=1}^{n} \mathcal{L}^*_{l(i')}\,\nu_{i'}}$, with $\mathcal{L}^*_0 := 0$,
```

```
            to the focal element $E_{ij}$ defined in Step 3.
        end
    end
```

Algorithm 4.1 constructs a random set $(\mathcal{E},\nu)_{||^*\hat{y}}$ with at most $\sum_{i=1}^{n} l(i)$ focal elements. The number of focal elements will be smaller than the sum, if some $E_{ij}$ and $E_{i'j'}$ are identical, and thus can be lumped together by adding their respective Möbius masses. The associated set function $bel(\cdot||^*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ is indeed a belief function, since $\nu^*_{ij} > 0$ for all pairs $(i,j)$, and

$$\sum_{i=1}^{n}\sum_{j=1}^{l(i)} \nu^*_{ij} = \sum_{i=1}^{n} \frac{\left(\sum_{j=1}^{l(i)} \mathcal{L}^*_j - \mathcal{L}^*_{j-1}\right)\nu_i}{\sum_{i'=1}^{n} \mathcal{L}^*_{l(i')}\,\nu_{i'}} = 1\,.$$

**Lemma 4.2** *Let $\{A_1,...,A_k\}$ be a finite partition of a (possibly uncountable) space $\Omega$. Let $bel : \mathcal{A}(\Omega) \to [0,1]$ be a belief function and $pl : \mathcal{A}(\Omega) \to [0,1]$ its conjugate plausibility function, which are associated with a random set $(\mathcal{E},\nu)$. Let $\mathcal{L}'(\cdot|\hat{y}) : \Omega \to \{\mathcal{L}^*_1,...,\mathcal{L}^*_l\}$ be a likelihood function, which takes on constant values $\mathcal{L}^*_j$ (see Definition 4.18) on the atoms $A_j$.*

*Let $\overline{C}_{\overline{P}}(\mathcal{L}'_A)$ be the upper Choquet integral of the A-restricted likelihood function w.r.t. to pl as given in Definitions (4.6). Assume that $\overline{C}_{\overline{P}}(\mathcal{L}') > 0$.*

*Let $\overline{P}(\cdot||^*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ be the set function defined in Equation (4.15). Let $bel(\cdot||^*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ be the belief function, and $pl(\cdot||^*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ the conjugate plausibility function, that are associated with the random set generated by Algorithm 4.1.*

*Then it is $\overline{P}(\cdot||^*\hat{y}) = pl(\cdot||^*\hat{y})$.*

Lemmas (4.1) and (4.2) show that $\underline{P}(\cdot||_*\hat{y})$ is a belief function and $\overline{P}(\cdot||^*\hat{y})$ approximately a plausibility function if the prior lower probability $\underline{P}$ is a belief function. Moreover, we have established formulas to compute the random sets of these set functions from knowledge of the random set of the prior belief function and the infimum or supremum value of the likelihood function on the atoms and focal elements, respectively. This provides the means to (approximately) update the prior $bel : \mathcal{A}(\Omega) \to [0,1]$ with a likelihood function $\mathcal{L}(\cdot|\hat{y}) : \Omega \to \mathbb{R}^+_0$ to the posteriors $bel(\cdot||_*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ and $bel(\cdot||^*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ directly on the level of Möbius assignments. From knowledge of these posterior belief functions, we can approximate the posterior lower probability that results from application of the GBR.

**Theorem 4.1** *Let $bel : \mathcal{A}(\Omega) \to [0,1]$ be a belief function with random set $(\mathcal{E},\nu) = \{\,(E_1,\nu_1)\,,...,(E_n,\nu_n)\,\}$, and $\Gamma(bel)$ its associated structure.*

*Let $\mathcal{L}'(\cdot|\hat{y}) : \Omega \to \{\mathcal{L}^*_1,...,\mathcal{L}^*_l\}$ be a likelihood function with constant values on the elements of a finite partition $\{A_1,...,A_k\}$ of $\Omega$, as described in Lemma 4.2. Let $\underline{C}(\mathcal{L}') = \sum_{i=1}^{n} \inf_{\theta\in E_i} \mathcal{L}'(\theta|\hat{y})\,\nu_i$, and $\overline{C}(\mathcal{L}') = \sum_{i=1}^{n} \sup_{\theta\in E_i} \mathcal{L}'(\theta|\hat{y})\,\nu_i$ be the lower and upper expectation, respectively, for observing $\hat{y}$. Assume $\underline{C}(\mathcal{L}') > 0$.*

*Let $\mathcal{M}(\hat{y})$ be the convex set of posterior probabilities that emerges from updating the structure $\Gamma(bel)$ with the likelihood function $\mathcal{L}'$ by application of the Generalised Bayes Rule (Definition 4.2). Let $\underline{P}(\cdot|\hat{y})$ be the lower envelope of $\mathcal{M}(\hat{y})$, i.e.,*

$$\underline{P}(A|\hat{y}) = \inf_{P \in \mathcal{M}(\hat{y})} P(A) \quad \textit{for all} \quad A \in \mathcal{A}(\Omega) \ .$$

*Let $bel(\cdot||_*\hat{y}) : \mathcal{A}(\Omega) \rightarrow [0,1]$ be the posterior belief function generated from $(\mathcal{E},\nu)$ and $\mathcal{L}'(\cdot|\hat{y})$ as described in Lemma 4.1. Let $bel(\cdot||^*\hat{y}) : \mathcal{A}(\Omega) \rightarrow [0,1]$ be the posterior belief function generated from $(\mathcal{E},\nu)$ and $\mathcal{L}'(\cdot|\hat{y})$ as described in Algorithm 4.1. Let $pl(\cdot||^*\hat{y}) : \mathcal{A}(\Omega) \rightarrow [0,1]$ be the conjugate plausibility function of $bel(\cdot||^*\hat{y})$.*

*Then $\underline{P}(\cdot|\hat{y})$ can be derived from knowledge of $bel(\cdot||_*\hat{y})$, $pl(\cdot||^*\hat{y})$, $\underline{C}(\mathcal{L}')$, and $\overline{C}(\mathcal{L}')$:*

$$\underline{P}(A|\hat{y}) = \frac{bel(A||_*\hat{y})\,\underline{C}(\mathcal{L}')}{bel(A||_*\hat{y})\,\underline{C}(\mathcal{L}') + pl(A^c||^*\hat{y})\,\overline{C}(\mathcal{L}')} \qquad \textit{for all} \ \ A \in \mathcal{A}(\Omega) \ . \tag{4.19}$$

Theorem 4.1 holds for a likelihood function $\mathcal{L}'$ with constant values over the atoms of a partition $\{A_1, ..., A_k\}$ of $\Omega$. Typically, such a step function will only be an approximation of a continuous likelihood function $\mathcal{L}(\cdot|\hat{y}) : \Omega \rightarrow \mathbb{R}_0^+$. Since the constant values $\mathcal{L}_j^*$ were chosen to be the suprema of the continuous likelihood function on the atoms $A_j$ (see Definition 4.18), the step function constitutes an upper envelope of the likelihood, i.e., $\mathcal{L}(\theta|\hat{y}) \leq \mathcal{L}'(\theta|\hat{y})$ for all $\theta \in \Omega$. As a consequence, the upper Choquet integral for the $A$-restricted continuous likelihood function $\mathcal{L}$ will be bounded from above by the estimate that is obtained from the envelope $\mathcal{L}'$, i.e.,

$$\overline{C}_{\overline{P}}(\mathcal{L}_A) \ = \sum_{E_i \cap A \neq \emptyset} \ \sup_{\theta \in E_i \cap A} \ \mathcal{L}(\theta|\hat{y})\,\nu_i \ \leq \sum_{E_i \cap A \neq \emptyset} \ \sup_{j \,|\, A_j \cap (E_i \cap A) \neq \emptyset} \ \mathcal{L}_j^* \ . \tag{4.20}$$

It is shown in detail in the proof of Lemma 4.2 that the right-hand side of the inequality is described by $\overline{C}(\mathcal{L}')\,pl(A||^*\hat{y})$. Hence, the plausibility function generated in Algorithm 4.1 provides an upper bound on the value of $\overline{C}_{\overline{P}}(\mathcal{L}_A)$ that would be obtained from the continuous likelihood function. As a consequence, Equation (4.19) will yield a lower bound on the posterior lower probability $P(\cdot|\hat{y})$. The true value of $P(\cdot|\hat{y})$ will be approximated the better the closer $\overline{C}_{\overline{P}}(\mathcal{L}_A)$ is approximated by $\overline{C}(\mathcal{L}')\,pl(A||^*\hat{y})$. This can be achieved by increasing the resolution of the partition, and/or adapting it to the shape of the likelihood function as we have done in Section 3.4.1.

However, we might also go into the opposite direction and decrease the accuracy of the approximation in exchange for a reduction in computational complexity. The number $\tilde{l} \leq \sum_{i=1}^{n} l(i)$ of focal elements in the random set $(\mathcal{E},\nu)_{||^*\hat{y}}$ can be large if the cardinality $k$ of the partition $\{A_1, ..., A_k\}$ is on the order of hundreds or more atoms. This will usually be the case in multi-dimensional problems as exemplified by our analysis. In order to reduce the number of focal elements, one can decide to work just with a limited number $s$ of $\mathcal{L}_j^*$-cuts of the likelihood function. This would correspond to fixing $s$ likelihood values $\{\mathcal{L}_{\min}^*, \mathcal{L}_2^*, ..., \mathcal{L}_{s-1}^*, \mathcal{L}_{\max}^*\}$ and assigning only these $s$ values to all atoms. Then, the number of focal elements in $(\mathcal{E},\nu)_{||^*\hat{y}}$ will be limited to

$s \cdot n \ll \tilde{l}$, which can reduce the computational costs of processing the posterior imprecise probability greatly. As a consequence, however, the upper approximation of the likelihood is coarsened, and the lower bound on the posterior lower probability from Equation (4.19) will further decrease. We will investigate below in how far the coarsening of the likelihood information leads to overly imprecise bounds on the posterior probability.

The question might arise why we have not been content with Expression (4.8) that follows from the result of Wasserman and Kadane (1990). It already allows to compute the posterior lower probability from knowledge of the likelihood function and the random set $(\mathcal{E}, \nu)$ of the prior belief function. Nevertheless, Approximation (4.19) provides an improvement, because it builds on two posterior belief functions $bel(\cdot||_*\hat{y})$ and $bel(\cdot||^*\hat{y})$ that already include the likelihood information. Application of the explicit formula (4.8) to calculate $\underline{P}(A|\hat{y})$ would require to find the supremum likelihood on all intersections $E_i \cap A^c$, with $E_i \in \mathcal{E}$, for each event $A$ under consideration. Moreover, the posterior belief functions define two structures $\Gamma(bel(\cdot||_*\hat{y}))$ and $\Gamma(bel(\cdot||^*\hat{y}))$ of posterior probabilities. Since we want to process the imprecise probability information in an integrated assessment of climate change, it is important to continue to work with posterior imprecise probability models.

## 4.3 Imprecise posterior probability for the climate model

In this section we will apply the newly established methodology to updating the imprecise prior probability for the climate model parameters from Chapter 3 with the likelihood of reproducing the historical temperature record estimated in Chapter 2. The imprecise prior probability is described by a random set $(\mathcal{G}, \nu^*)_{\mathrm{CM}}$ with 6048 focal elements $G$ (see Section 3.5). In the following, we will drop the superscript $^*$ for the Möbius masses in order to simplify the notation. The focal elements are defined on a partition $\{A_1, ..., A_k\}_{\mathrm{CM}}$ of the joint parameter domain $\Omega = \Omega(T_{2x}) \times \Omega(\kappa_v) \times \Omega(Q_{\mathrm{S90}})$ (see Definition 3.18) with $k = 20 \cdot 12 \cdot 14 = 3360$ atoms $A_j$. Due to the construction of the joint imprecise prior probability from marginal random sets by means of the Möbius independent product (3.22), all focal elements $G \in \mathcal{G}$ constitute cubic product sets $(T_{2x,*}, T_{2x}^*] \times (\kappa_{v,*}, \kappa_v^*] \times (Q_{\mathrm{S90},*}, Q_{\mathrm{S90}}^*] \subseteq \Omega$. Hence, they can be fully described by their lower and upper bound in each dimension of the parameter space. $(\mathcal{G}, \nu)_{\mathrm{CM}}$ defines a belief function $bel_{\mathrm{CM}} : \mathcal{R}^3 \to [0,1]$ that represents the lower envelope of a convex set of prior probabilities $\Gamma(bel_{\mathrm{CM}})$ on the Borel field $\mathcal{R}^3$ of $\Omega$.

For constructing the posterior belief function $bel_{\mathrm{CM}}(\cdot||_*\hat{T})$ and its associated random set $(\mathcal{G}, \nu)_{||_*\hat{T}}$ defined in Equation 4.17, we need to know the infimum likelihood $\mathcal{L}_{*,i}$ on all focal elements $G_i \in \mathcal{G}$. Since the focal elements are included in the power set of the partition $\{A_1, ..., A_k\}_{\mathrm{CM}}$, it suffices to find the infimum likelihood $\mathcal{L}_{*,j}$ for each atom $A_j$. Then, we can calculate immediately

$$\mathcal{L}_{*,i} := \inf_{\theta \in G_i} \mathcal{L}(\theta; \hat{T}) = \min_{A_j \subseteq G_i} \mathcal{L}_{*,j} := \min_{A_j \subseteq G_i} \left( \inf_{\theta \in A_j} \mathcal{L}(\theta; \hat{T}) \right) \tag{4.21}$$

for each $G_i \in \mathcal{G}$. In addition, we need to know the supremum likelihood $\mathcal{L}^*_j := \sup_{\theta \in A_j} \mathcal{L}(\theta; \hat{T})$ for each atom $A_j$ in order to identify the Dempster-conditioned random set $(\mathcal{G}, \nu)_{||^*\hat{T}}$ by means of Algorithm 4.1.

We have pointed out in Section 2.4 that the simple structure of our diffusion ocean energy balance model DOECLIM allows us to employ gradient-based search algorithms for the combination of climate model parameters $\theta^* = (T_{2x}, \kappa_v, Q_{\mathrm{S90}}) \in \Omega$ and nuisance parameters $\theta^*_N = (\beta_{\mathrm{Vol}}, \alpha, \beta_{L,\mathrm{SOI}}, \beta_{S,\mathrm{SOI}}) \in \Omega_N$ which maximises the likelihood of reproducing the instrumental temperature record (see Section 2.4). We use the implementation of DOECLIM in the optimisation environment GAMS (Brooke et al., 1992) which was established for the calculation of the likelihood function in Section 2.4 to identify the maximum likelihood $\mathcal{L}^*_j$ for each atom $A_j$. $\mathcal{L}^*_j$ is calculated by a gradient-based search of the joint parameter space $\Omega \times \Omega_N$ subject to confining $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$ to the area of the respective atom $A_j$. The derivation of the minimum likelihoods $\mathcal{L}_{*,j}$ requires a different approach. Since the definition of the likelihood function $\mathcal{L}(\cdot; \hat{T}) : \Omega \rightarrow \mathbb{R}^+_0$ in Section 2.4 involved a maximal choice of nuisance parameters $\theta_N$ at each point $\theta \in \Omega$, the search for $\mathcal{L}_{*,j}$ on the domain of an atom $A_j$ would constitute a minimax optimisation problem. Fortunately, the minimum likelihoods $\mathcal{L}_{*,j}$ have to be situated at the vertices of the atoms $A_j$, since the likelihood function is approximately quasi-concave (see Section 3.4.1). Hence, we have calculated the likelihood $\mathcal{L}(\theta_{j'} | \hat{T})$ at the $21 \cdot 13 \cdot 15 = 4095$ vertices $\theta_{j'}$ of the atoms that define in effect the partition $\{A_1, ..., A_k\}$. The calculation involved a gradient based search for the optimal choice of nuisance parameters $\theta_{N,j'}$ at each of the points $\theta_{j'}$. The minimum likelihood $\mathcal{L}_{*,j}$ can now be inferred by collecting the 8 vertices of the respective atom $A_j$ and picking the minimum of the corresponding likelihood values. After calculating the minimum and maximum likelihood for all atoms of the partition $\{A_1, ..., A_k\}$, and hence, for all focal elements $G_i \in \mathcal{G}$, we can now turn to the derivation of the posterior random sets $(\mathcal{G}, \nu)_{||^*\hat{T}}$ and $(\mathcal{G}, \nu)_{||_*\hat{T}}$ that follow from Dempster's rule (4.10) and its dual rule (4.13) respectively.

### 4.3.1 Belief functions from Dempster's and its dual rule

The random set $(\mathcal{G}, \nu)_{||_*\hat{T}}$ can be calculated immediately from knowledge of the minimum likelihoods $\mathcal{L}_{*,i}$ and the prior Möbius masses $\nu_i$ of the focal elements $G_i \in \mathcal{G}$ (see Equation 4.17). The resulting posterior Möbius masses $\nu_{||_*,i}$ are normalised by the *lower expectation value* of the likelihood of reproducing the instrumental temperature record, i.e.,

$$\underline{C}(\mathcal{L}) = \sum_{G_i \in \mathcal{G}} \inf_{\theta \in G_i} \mathcal{L}(\theta; \hat{T})\, \nu_i = \sum_{G_i \in \mathcal{G}} \mathcal{L}_{*,i}\, \nu_i = 0.0017 \;, \tag{4.22}$$

in units of the maximum likelihood. The small value of $\underline{C}(\mathcal{L})$ has to be compared with the *upper expectation value* of the likelihood, i.e.,

$$\overline{C}(\mathcal{L}) = \sum_{G_i \in \mathcal{G}} \sup_{\theta \in G_i} \mathcal{L}(\theta; \hat{T})\, \nu_i = \sum_{G_i \in \mathcal{G}} \mathcal{L}^*_i\, \nu_i = 0.856 \;, \tag{4.23}$$

which is approximately 500 times larger. The large ratio between lower and upper expected likelihood already heralds our difficulties to produce meaningful posterior lower probability bounds from the application of the Generalised Bayes' rule (cf. Equation 4.19). We will discuss these difficulties as well as the reasons for the large relative difference between $\underline{C}(\mathcal{L})$ and $\overline{C}(\mathcal{L})$ in the subsequent section.

The small value of $\underline{C}(\mathcal{L})$ indicates that there exists only a small number of focal elements $G_i \in \mathcal{G}$ with non-negligible lower likelihood $\mathcal{L}_i^*$. These focal elements consist primarily of atoms of the partition around the area of maximum likelihood that are included in $\mathcal{G}$. Larger focal elements ranging over several atoms are apparently not confined enough to exclude areas of the parameter space where the likelihood is close to zero. As a consequence, the posterior Möbius assignment $\nu_{||_*}$ that follows from the dual updating rule as described in Equation (4.17) will allocate almost all Möbius mass to the atoms around the point of maximum likelihood. This can be seen clearly in Fig. 4.1 (left panels) which depicts the resulting lower and upper posterior probability distributions on the series of marginal cumulative events in each dimension of the parameter space. The lower and upper distribution functions confine a tight band of distributions that rise steeply at the point of maximum likelihood. It becomes evident that not only most of the imprecision in the prior uncertainty model has been eliminated, but also the possible support of admissible posterior probabilities is strongly constrained. Hence, the fixation on the infimum likelihoods $\mathcal{L}_{*,i}$ of the focal elements $G_i$ would lead to an overvaluation of the maximum likelihood information if the belief function $bel(\cdot||_*\hat{T})$ was considered an eligible posterior probability model as such. In this analysis, however, we do not intend to endow it with this status but use it solely as a means to find the lower and upper probability bounds under the generalised Bayes' rule.

The situation is different for Dempster's rule of conditioning defined in Equation (4.10). In the context of updating a set of prior probabilities with a likelihood, Gilboa and Schmeidler (1993) have shown it to be identical with the so-called maximum likelihood classical update rule (if the joint probability on the space $\Omega \times Y$ is 2-monotone). This rule employs the likelihood information in two ways. First, all prior probabilities yielding a smaller expected likelihood $E(\mathcal{L})$ than the upper expected likelihood $\overline{C}(\mathcal{L})$ are eliminated from the set of prior probabilities. Then all remaining prior probabilities with $E(\mathcal{L}) = \overline{C}(\mathcal{L})$ are updated with the likelihood via Bayes' rule to form the set of posterior probabilities. It is important to note, however, that this probabilistic interpretation of Dempster's rule does not accord with the Generalised Bayes' rule, which follows from the minimum requirement of coherence between conditional and unconditional lower probabilities (Walley, 1991, Chapter 6). Under the GBR, all prior probabilities have to be updated indiscriminately (Walley, 1991, Theorem 6.4.2 and Remark 8.4.8). Dempster's rule of conditioning has been demonstrated to yield unreasonable conditional lower probabilities in some instances, where the combination of unconditional and Dempster-conditioned beliefs would allow an observer to construct a series of bets which are acceptable under the given beliefs but which, nevertheless, will inflict a sure loss (e.g., Walley, 1996b). Therefore, many authors criticised the application of Dempster's rule in the framework of imprecise probabilities. However, to the best of our knowledge all examples of inconsistencies from the application of Dempster's rule are related to the problem of conditioning a lower probability on an event $B$ rather than to the problem of updating a prior assessment with the likelihood of an observation $\hat{y}$. As mentioned above, the latter case has received little scrutiny in the literature. Therefore, we cannot oversee whether similar instances of sure loss under a behavioural interpretation of imprecise probabilities can be identified for a 'Dempster rule of likelihood updating' as formulated in Equation (4.15) and Algorithm 4.1. Intuitively, we regard the application of the maximum likelihood update rule as a reasonable alternative to the GBR in the context of our analysis. Therefore, we will consider the Dempster-updated belief function as a viable model for our posterior uncertainty that amalgamates the prior and the likelihood information.

In order to apply Algorithm (4.1) to generate the random set $(\mathcal{G}, \nu)_{||^*\hat{T}}$ of the Dempster-updated belief function $bel(\cdot||^*\hat{T})$, we need to coarsen the continuous likelihood information to a step function that takes on constant likelihood values $\mathcal{L}'_j \geq \mathcal{L}^*_j$ on the atoms $A_j$ of the partition $\{A_1, ..., A_k\}$. Given the large number of $k = 3360$ atoms and $n = 6048$ prior focal elements in our analysis, Algorithm (4.1) could generate more than a million posterior focal elements if the likelihood function was not coarsened beyond the size of the atoms $A_j$. To ensure computational feasibility, we have reduced the likelihood information to 22 levels $\mathcal{L}'_{22} = \{0.001, 0.01, 0.05, 0.1, ..., 0.95, 1\}$ (in units of the maximum likelihood), on which an outer step function approximation of the likelihood function was constructed. This involved the allocation of the smallest likelihood level $\mathcal{L}'_s \geq \mathcal{L}^*_j$, $1 \leq s \leq 22$ to each atom $A_j$. Algorithm 4.1 can now be used to construct the posterior focal elements $G_{||^*,is}$ constituted by the intersections between the prior focal elements $G_i \in \mathcal{G}$ and the $s$-cuts of the likelihood function. Since the cuts are in general not cubic product subsets of the 3-dimensional parameter space $\Omega$, many of the posterior focal elements will not be cubic product sets as well. This additional complexity considerably slowed the construction of the posterior random set, since we needed to compare each set $E_{ij}$ constructed by Algorithm 4.1 with the list of other sets $E_{i'j'}$ contained in $\mathcal{G}_{||^*}$ in order to eliminate repetitions of focal elements. For the case of 22 likelihood levels, we have found 26910 distinct focal elements $G_{||^*,is} \in \mathcal{G}_{||^*}$.

The resulting posterior lower and upper probability distributions on the series of marginal cumulative events for the climate model parameters are shown in Fig. 4.1 (right panels). A comparison with the marginal prior imprecise probability models $\Gamma_X(\underline{S}F, \overline{S}F, \underline{p})$ from Section 3.4 reveals that the updating procedure has considerably tightened the admissible distribution bands for climate sensitivity and sulphate aerosol forcing, while the impact on the prior uncertainty about ocean heat diffusivity is less pronounced. In particular, the Möbius masses allocated to the low and high end of the prior range for these quantities have been strongly reduced. Inter alia, this reflects the fact that the likelihood function presented in Section 2.5 drops steeply towards small climate sensitivities and large coolings from sulphate aerosols. The reduction of Möbius mass at the high end of climate sensitivities follows from a combined effect of the

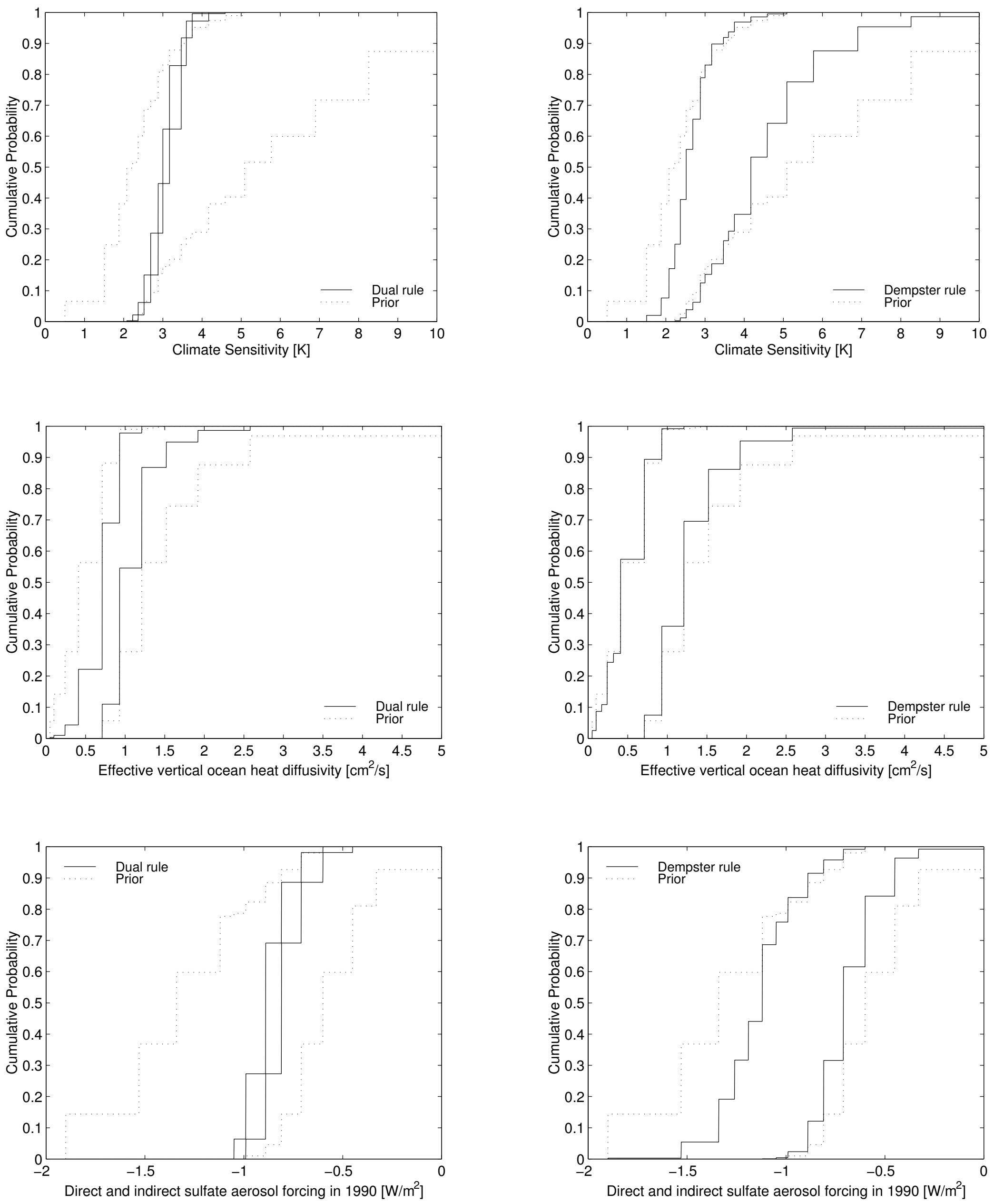


Figure 4.1: Posterior lower and upper distribution functions for the climate model parameters from the application of Dempster's rule (right panels) and its dual updating rule (left panels).

likelihood information and the observationally based estimate for the prior uncertainty about ocean heat diffusivity (see Section 3.4.4). Since the prior allocates only small Möbius mass to diffusivities $\kappa_v \geq 2.5\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, the Möbius mass for large values of $T_{2x}$, which require either large diffusivities $\kappa_v$ or large sulphate coolings $Q_{\mathrm{S90}}$ to obtain a considerable likelihood (see Fig. 2.7), is reduced accordingly.

| # likelihood levels | # distinct focal elements | $\overline{C}(\mathcal{L}')$ |
|---|---|---|
| 7 | 15979 | 0.912 |
| 12 | 20110 | 0.889 |
| 22 | 26910 | 0.876 |

Table 4.1: Results from Algorithm 4.1 for three different approximations of the likelihood function by upper step functions with 7, 12, and 22 likelihood levels.

The accuracy of approximating the likelihood function with 22 likelihood levels $\mathcal{L}'_{22}$ relative to the ideal case of using the full set of supremum likelihoods $\mathcal{L}^*_j$ can be assessed, inter alia, by comparing the resulting value for the upper expected likelihood, $\overline{C}(\mathcal{L}'_{22}) = 0.876$, with $\overline{C}(\mathcal{L}^*) = 0.856$. The approximation $\mathcal{L}'_{22}$ needs to yield a larger value for the upper expected likelihood by definition, since it constitutes an upper envelope of $\mathcal{L}^*$. To put the comparison into perspective, we have also constructed coarser likelihood approximations with 7 levels $\mathcal{L}'_7 = \{0.001, 0.01, 0.2, 0.4, ..., 0.8, 1\}$ and 12 levels $\mathcal{L}'_{12} = \{0.001, 0.01, 0.1, 0.2, ..., 0.9, 1\}$ and calculated the resulting Dempster-updated random sets $(\mathcal{G}, \nu)_{||^*\hat{T},7}$ and $(\mathcal{G}, \nu)_{||^*\hat{T},12}$ by means of Algorithm 4.1. Table 4.1 shows that the accuracy in terms of $\overline{C}(\mathcal{L}') - \overline{C}(\mathcal{L}^*)$ improves by roughly a factor three, while the number of focal elements increases by approximately 70%. We have also assessed the implications of coarsening the likelihood approximation for the Dempster-updated lower and upper probability bounds on the series of marginal cumulative events. Fig. 4.2 shows the result for the case of climate sensitivity. It can be seen that the $\mathcal{L}_{22}$-approximation yields slightly different bounds than the $\mathcal{L}_7$-approximation, but the change due to the higher resolution is surprisingly small. This is underlined by the fact that the lower and upper distribution bounds from the $\mathcal{L}_{12}$-approximation are almost identical with the bounds from the $\mathcal{L}_{22}$-case, for which reason we do not show it explicitly in Fig. 4.2. It is important to note that a Dempster-updated belief function $bel(\cdot||^*\hat{T}, s)$ for a $\mathcal{L}_s$-approximation does not need to completely dominate a posterior belief function $bel(\cdot||^*\hat{T}, s')$ based on a smaller number $s' < s$ of likelihood levels, since it is only required that

$$\overline{C}(\mathcal{L}'_s)\; pl(A||^*\hat{y}, s) \leq \overline{C}(\mathcal{L}'_{s'})\; pl(A||^*\hat{y}, s') \;,$$

as we have discussed at the end of Section 4.2. We conclude that the $\mathcal{L}_{22}$-approximation of the likelihood function already yields a satisfactory approximation of the Dempster-updated random set $(\mathcal{G}, \nu)_{||^*\hat{T}}$ that would be obtained by taking the full information about the supremum likelihoods $\mathcal{L}^*_j$ over the atoms $A_j$ into account. Henceforth, we will always refer to the approximations $(\mathcal{G}, \nu)_{||^*\hat{T},22}$ and $bel(\cdot||^*\hat{T}, 22)$ when talking about the Dempster-updated random set and belief function, respectively. To simplify the notation we will drop the reference to the 22 levels of the likelihood approximation in the following.

On the basis of the Dempster-updated imprecise probability model, we can constrain the radiative forcing from sulphate aerosols in the year 1990 to the interval

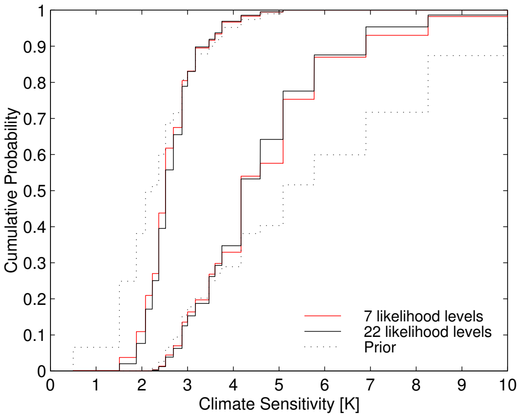


Figure 4.2: Comparison of the lower and upper distribution functions for climate sensitivity from the application of Algorithm 4.1 for two different approximations of the likelihood function by upper step functions with 7 and 22 likelihood levels.

$(-1.53\,\mathrm{W\,m^{-2}}, 0.33\,\mathrm{W\,m^{-2}}]$ with lower probability $bel((-1.53\,\mathrm{W\,m^{-2}}, 0.33\,\mathrm{W\,m^{-2}}]||^*\hat{T}) = 0.99$. This estimate can be compared with the 1%-99% quantile ranges of Andronova and Schlesinger (2001, GTAS scenario) ($\approx -1.5\,\mathrm{W\,m^{-2}}$ to $-0.4\,\mathrm{W\,m^{-2}}$), Knutti et al. (2002) ($\approx -1.8$ to $-0.5\,\mathrm{W\,m^{-2}}$) and Forest et al. (2002) ($\approx -1.0$ to $-0.1\,\mathrm{W\,m^{-2}}$). Since our imprecise posterior probability supports a strictly negative forcing from sulphate aerosols, we do not find any evidence that the heating contribution of black carbon might have neutralised the cooling from sulphate aerosols.

Concerning climate sensitivity, the posterior belief function $bel(\cdot||^*\hat{T})$ virtually excludes values of $T_{2x} < 1.5\,\mathrm{K}$ with lower probability $bel((1.5\,\mathrm{K}, 10\,\mathrm{K}]||^*\hat{T}) = 0.9996$. This strong rejection of low climate sensitivities is in agreement with the results of Murphy et al. (2004) and Andronova and Schlesinger (2001) who allocate probability mass $< 0.005$ to values of $T_{2x} \leq 1.5\,\mathrm{K}$. The probability estimates of Forest et al. (2002) support low climate sensitivities to a greater extent, with their 5% quantiles situated at $T_{2x} = 1.3 - 1.4\,\mathrm{K}$. On the upper end of climate sensitivity, the Dempster-updated belief function cannot provide a strict constraint on high values of $T_{2x}$, albeit it allocates much less Möbius mass to this area compared with the prior belief function (see Fig. 4.1). We find upper probabilities of 0.047 for $T_{2x} > 6.9\,\mathrm{K}$ and 0.014 for $T_{2x} > 8.3\,\mathrm{K}$. The upper 95% quantile of $T_{2x} \approx 6.9\,\mathrm{K}$ can be compared with the 95% quantiles of Andronova and Schlesinger (2001, GTAS scenario) ($\approx 6.2\,\mathrm{K}$), Forest et al. (2002) (7.7 K for uniform prior, 5.3 K for expert prior), Knutti et al. (2002) ($\approx 9.1\,\mathrm{K}$), and Murphy et al. (2004) ($5.3 - 5.4\,\mathrm{K}$).

In contrast to climate sensitivity and sulphate aerosol forcing, the Dempster-updated imprecise probability has added little information to the prior belief function about vertical ocean heat diffusivity $\kappa_v$. This is mainly due to the fact that the construction of the prior uncertainty model was already based on observations of ocean heat uptake in the 20th century (see Section 3.4.4). The only noteworthy effect of the updating

| | $T_{2x} \in [0\,\mathrm{K}, 1.5\,\mathrm{K}]$ | $T_{2x} \in [1.5\,\mathrm{K}, 4.5\,\mathrm{K}]$ | $T_{2x} \in [4.5\,\mathrm{K}, 10.0\,\mathrm{K}]$ |
|---|---|---|---|
| Prior $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$ | [0, 0.07] | [0.31, 0.98] | [0.02, 0.62] |
| Posterior $\Gamma(bel(\cdot \vert\vert^* \hat{T}))$ | [0, 0.00] | [0.53, 0.99] | [0.01, 0.47] |

Table 4.2: Probability intervals for climate sensitivity below, within and above the range of the IPCC estimate for the prior and Dempster-updated imprecise probability models.

procedure consists in a reduction of Möbius masses for high diffusivities $\kappa_v$. We find $\kappa_v$ to be smaller than $2\,\mathrm{cm}^2\,\mathrm{s}^{-1}$ with lower probability $bel((0, 2\,\mathrm{cm}^2\,\mathrm{s}^{-1}]||^*\hat{T}) = 0.953$. Taking the results for all three model parameters $T_{2x}$, $\kappa_v$ and $Q_{\mathrm{S90}}$ together, we conclude that the likelihood information from the comparison of our energy balance model DOECLIM with the instrumental surface temperature record allows us to constrain the prior imprecise probability model under Dempster updating considerably. This is underlined by the change in lower and upper probability for the IPCC estimate that is highlighted in Table 4.2. It is also important to note that the Dempster-updated imprecise probability model tightens not only the probability bounds for the marginal events on the subspaces $\Omega(T_{2x})$, $\Omega(\kappa_v)$ and $\Omega(Q_{\mathrm{S90}})$, but also includes the dependency structure between the three model parameters that is introduced by the likelihood information.

### 4.3.2 Result from the Generalised Bayes' rule

Having calculated the Dempster-updated belief function $bel(\cdot||^*\hat{T})$ (for the $\mathcal{L}'_{22}$ - approximation) and its dual $bel(\cdot||_*\hat{T})$, respectively, we can now apply the GBR to update the entire set of prior probabilities $\Gamma_{\mathrm{CM}}(\underline{SF}, \overline{SF}, \underline{p})$ with the likelihood function $\mathcal{L}(\cdot; \hat{T}) : \Omega \to \mathbb{R}^+$ derived in Chapter 2. Equation (4.19) allows us to derive the resulting posterior lower probability $\underline{P}(A|\hat{T})$ and posterior upper probability $\overline{P}(A|\hat{T})$ for arbitrary subsets $A \in \Omega(T_{2x}) \times \Omega(\kappa_v) \times \Omega(Q_{\mathrm{S90}})$ directly from knowledge of the random sets $(\mathcal{G}, \nu)_{||^*\hat{T}}$ and $(\mathcal{G}, \nu)_{||_*\hat{T}}$, and the lower and upper expected likelihoods $\underline{C}(\mathcal{L}_*)$ and $\overline{C}(\mathcal{L}')$, respectively. The results for the lower and upper distribution functions on the series of marginal cumulative events are shown in Fig. 4.3 for vertical ocean heat diffusivity $\kappa_v$ and total sulphate aerosol forcing $Q_{\mathrm{S90}}$, and in Fig. 4.4 for climate sensitivity. It can be seen that the bounds are extremely wide, approaching a state of complete ignorance with lower bound $\underline{P}((-\infty, x]|\hat{T}) = 0$ and upper bound $\overline{P}((-\infty, x]|\hat{T}) = 1$ over most of the domains $\Omega(T_{2x})$, $\Omega(\kappa_v)$, and $\Omega(Q_{\mathrm{S90}})$. Hence, we have to report a negative result here. Obviously, we were not able to produce a meaningful posterior imprecise probability model under the application of the GBR. Since the marginal prior distribution bands are fully included in the probability bounds of the posterior model, they were *dilated* by the incorporation of the likelihood information. Although the general phenomenon of dilation has been recorded as a potential outcome of statistical inferences with imprecise probabilities (Seidenfeld and Wasserman, 1993), it constitutes an unreasonable result in the context of our analysis. Inspection of the prior imprecise probability model constructed in Chapter 3 and the likelihood function from Chapter 2

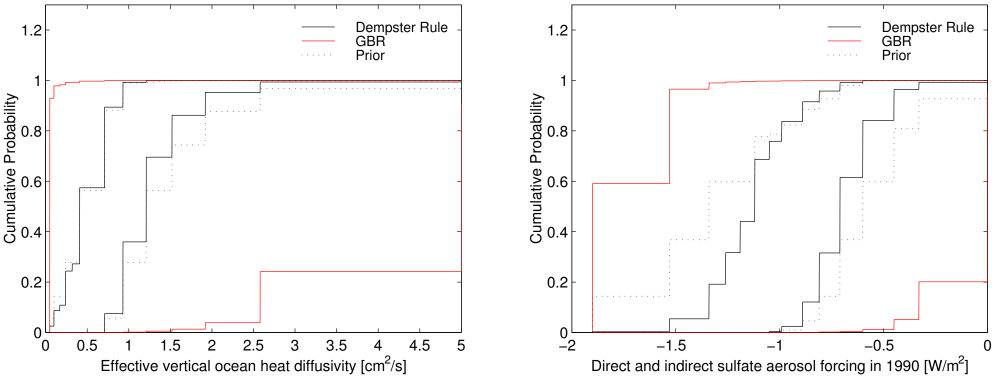


Figure 4.3: Posterior lower and upper distribution for effective vertical ocean heat diffusivity (left panel) and total sulphate aerosol forcing in 1990 (right panel) functions from the application of Dempster's rule and the GBR.

reveals that the likelihood information can hardly be considered a surprise in the light of the prior information, which might have justified an increase of the uncertainty.

Therefore, we need to inquire the steps of our analysis in order to identify the reason for the occurrence of dilation under the GBR. Inspection of Equation (4.19) reveals that the dilated bounds result from the large relative difference between the lower and upper expected likelihoods $\underline{C}(\mathcal{L}_*) = 0.0017$ and $\overline{C}(\mathcal{L}') = 0.876$. Since $\overline{C}(\mathcal{L}')$ is more than 500 times larger than $\underline{C}(\mathcal{L}_*)$, we can only produce posterior lower bounds significantly larger than zero for events $A$ with $bel(A||_*\hat{T}) \gtrapprox 500\, pl(A^c||^*\hat{T})$. This will only happen in the rare cases where $pl(A^c||^*\hat{T})$ is close to zero. Hence, meaningful posterior probability bounds can only be produced with the GBR if $\underline{C}(\mathcal{L}_*)$ and $\overline{C}(\mathcal{L}')$ differ by less than an order of magnitude. Obviously, this is not the case in our analysis. We can exclude as possible reason the approximation of the likelihood function by 22 likelihood levels $\mathcal{L}'_{22}$ that preceded the construction of the Dempster-updated random set $(\mathcal{G}, \nu)_{||^*\hat{T}}$. The corresponding increase in upper expected likelihood from $\overline{C}(\mathcal{L}^*) = 0.856$ to $\overline{C}(\mathcal{L}^*) = 0.876$ is negligible compared with the difference with $\underline{C}(\mathcal{L}_*) = 0.0017$. Instead, the main reason can be found in the large average change of likelihood on the focal elements, because

$$\overline{C}(\mathcal{L}^*) - \underline{C}(\mathcal{L}_*) = \sum_{i=1}^{n} \left(\mathcal{L}^*_i - \mathcal{L}_{*,i}\right) \nu_i \; . \tag{4.24}$$

Since we are working with belief functions, all Möbius masses $\nu_i$ are positive, so that large changes $\mathcal{L}^*_i - \mathcal{L}_{*,i}$ cannot be cancelled out, but will accumulate. Hence, one potential cause for the dilation of posterior bounds might be our choice of prior uncertainty model. However, we have made an effort in Chapter 3 to allocate a portion of Möbius mass to the atoms of the partition $\{A_1, ..., A_k\}$ (see Section 3.3), on which we have tried to limit the change of likelihood by equally distributing it over all atoms. Therefore, another potential cause might be that the construction of the likelihood-adapted grid was not sophisticated enough. It was based solely on projections of the likelihood function on the individual dimensions of the parameter space $\Omega$, and more importantly did only allow for a partition of $\Omega$ into a set of cubic product sets $A_{qrs} = A_q(T_{2x}) \times A_r(\kappa_v) \times A_s(Q_{\mathrm{S90}})$ whose ability to adapt to the shape of the likelihood function is limited.

In order to separate the importance of these two potential causes, we apply the methodology for updating a Möbius inverse under the GBR as developed in Section 4.2 to the special case of probability mass functions. We have pointed out before that probability mass functions constitute a simple case of Möbius inverse, where all probability (Möbius) mass is allocated to the atoms of a partition of the domain $\Omega$. Hence, additive probabilities can be included immediately in the framework of belief functions with a finite random set by coarsening the probabilistic information to a finite partition (which has to be done anyway for conducting numerical analyses). Therefore, Section 4.2 provides also an alternative to Monte Carlo methods for updating additive probabilities with Bayes rule. Having chosen a partition on which the probability mass function $p : \{A_1, ..., A_k\} \to [0, 1]$ is defined, we can calculate the Dempster updated mass function $p_{||^*}$ and its counterpart $p_{||_*}$ under the dual updating rule by

$$p_{||_*,j} = \frac{\mathcal{L}_{*,j}\, p_j}{\sum_{j=1}^{k} \mathcal{L}_{*,j}\, p_j} \ , \quad p_{||^*,j} = \frac{\mathcal{L}_j^*\, p_j}{\sum_{j=1}^{k} \mathcal{L}_j^*\, p_j} \ . \tag{4.25}$$

The latter formula for the Dempster-updated mass function can be seen to emerge directly from Algorithm 4.1 by noting that the intersects of the atoms with the cuts of the likelihood function are either empty or equal to the atoms themselves. Then, Equation (4.19) can be used to generate lower and upper bounds that will include the posterior additive probability emerging from Bayes' rule. As highlighted by Equation (4.24), this will only produce meaningful bounds if and only if the change of likelihood over atoms with significant prior probability mass is small. Hence, we can use the updating of probability mass functions to check whether the construction of the likelihood-adapted grid in Section 3.4.1 allows to generate such meaningful bounds. If this was the case, our failure to avoid dilation of the prior imprecise probability after application of the GBR would need to be attributed entirely to the choice of a prior belief function model.

We used the probability estimates of Andronova and Schlesinger (2001), Forest et al. (2002), Knutti et al. (2002), and Murphy et al. (2004) to generate six mass functions on the likelihood-adapted partition for climate sensitivity and four mass functions for sulphate aerosol forcing (cf. Fig. 3.1 and 3.8). Probability mass functions for vertical ocean heat diffusivity were constructed from the four extreme points of the parametric class of normal distributions specified in Section 3.4.4 (see also Fig. 3.11). In analogy to the combination of marginal random sets to our prior imprecise probability model in Section 3.5, we formed all possible products of these functions, thus generating a set of 96 probability mass functions $p_i : \{A_1, ..., A_k\} \to [0, 1]$, $1 \leq i \leq 96$ on the

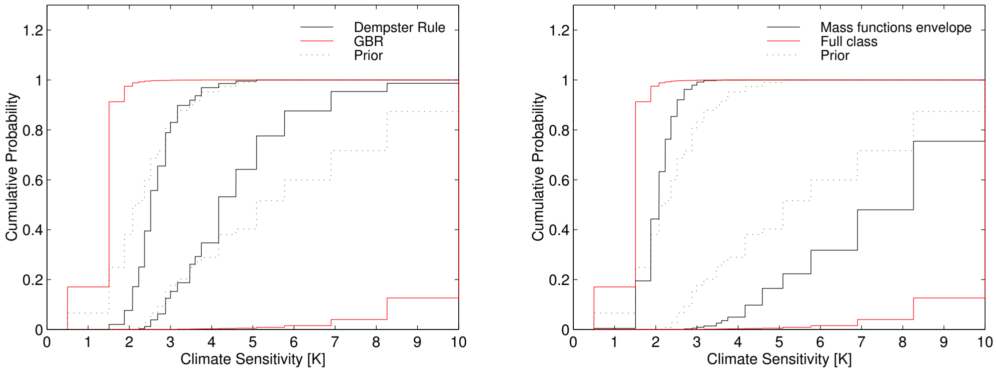


Figure 4.4: Posterior lower and upper distribution functions for climate sensitivity from the application of the GBR. The left panel also shows the Dempster-updated lower and upper distributions functions. The right panel includes the envelope of the posterior bounds from updating a set of 96 probability mass functions with the GBR (see text for further explanation).

partition of the joint climate model parameter space $\Omega$. The upper expected likelihoods $\overline{C}_i(\mathcal{L}^*) = \sum_{i=1}^{n} \mathcal{L}_j^* p_{ij}$ for the 96 probability mass functions were situated in the range 0.08 to 0.33, which is considerably lower than the value of $\overline{C}(\mathcal{L}^*) = 0.856$ for the prior belief function model. Likewise, the lower expected likelihoods $E_i(\mathcal{L}_*)$ were found in the range 0.01 and 0.04, which is significantly higher than $\underline{C}(\mathcal{L}_*) = 0.0017$. As a result, the relative differences between lower and upper expected likelihoods for the mass functions were reduced by almost two orders of magnitude relative to the prior belief function model. $E_i(\mathcal{L}^*)$ exceeded the value of $E_i(\mathcal{L}_*)$ by a factor between 6.6 and 10.3 . Therefore, the major part of our dilation result has to be attributed indeed to our choice of a prior belief function model. However, a smaller part can also be attributed to the insufficiency of the likelihood adapted grid. While we could produce meaningful posterior bounds from Equation (4.19) for some of the 96 mass functions, others were dilated. Fig. 4.4 (right panel) shows the envelope of the resulting 96 lower and upper posterior distribution functions on the series of marginal cumulative events for climate sensitivity (black line). It can be seen that this envelope is still wider than the prior distribution band encompassing the probability mass functions for climate sensitivity from the literature (dotted line), albeit the near-ignorance result for the prior belief function model (red line) has disappeared. We conclude that an improvement of the likelihood-adapted grid could further tighten the posterior bounds from updating the 96 mass functions with the likelihood information. However, it will do little to tighten the posterior bounds from the application of the GBR to the prior belief function. Since the belief function allocates most of the Möbius mass to focal elements extending over several atoms, large changes of likelihood on these focal elements and, hence, a large relative difference between $\underline{C}(\mathcal{L}_*)$ and $\overline{C}(\mathcal{L}^*)$ cannot be avoided. Our attempt to include lower mass functions on the atoms of the partition into the prior model did not succeed, because the lower envelope of the probability estimates from the literature allowed us to do this only to a very limited extent (Lower probability mass 0.32 for $T_{2x}$ and $\kappa_v$, and 0.2 for $Q_{\mathrm{S90}}$, see Section 3.4). We speculate that the application of the GBR to the intersection $\Gamma(\underline{S}F, \overline{S}F, \underline{p})$ of p-box and $\varepsilon$-contamination model will only produce meaningful posterior bounds if the contamination $\epsilon$ is small compared to the lower probability mass.

Our result shows that belief functions will constitute an inappropriate prior uncertainty model in most cases, in which an application of the GBR is attempted. The fundamental reason for this lies in the fact that in the presence of imprecision the focal elements will generally spread over portions of the uncertainty space, where the likelihood changes considerably. Since all the focal elements carry positive Möbius mass in the case of belief functions, such changes in likelihood cannot be compensated for (see Equation 4.24). As a result, the expected likelihood $E_P(\mathcal{L})$ will vary greatly over probabilities $P$ contained in the corresponding structure $\Gamma(bel)$, giving rise to the dilation of posterior probability bounds after application of the GBR (see Equation 4.19). In this sense, belief functions are not informative enough to exclude probabilities with widely different expectations about the likelihood. The situation might be different with lower probability models that allow for negative Möbius masses on the focal elements. The ALUP model is a natural candidate, albeit a general dilation result has been obtained also for this model (Herron et al., 1997). However, such models do not exhibit a sparse Möbius inverse in general, so that their mathematical tractability along the lines of this analysis would be lost. Therefore, the Generalised Bayes' rule points to the limitations of the approach of building on uncertainty representations with a sparse Möbius inverse that we are pursuing here.

An alternative will be to work with updating rules that are more restrictive than the Generalised Bayes rule. Dempster's updating rule provides such an alternative, since it is equivalent to the maximum likelihood updating rule (if the joint probability on the space $\Omega \times Y$ is 2-monotone). The rejection of probabilities that yield a lower than maximum expected likelihood safeguards against dilation even if the prior imprecise probability model was not informative enough to exclude probabilities with widely different expected likelihoods from the beginning. As discussed above, we regard the maximum likelihood update rule as a reasonable choice in the context of our analysis. This is reinforced by the corresponding Dempster-updated belief function for the climate model parameters, which reflects the likelihood information from Chapter 2 in a clearly visible manner (see Section 4.3.1). Therefore, we will employ the Dempster-updated belief function $bel(\cdot||^{*}\hat{T}) : \mathcal{R}^3 \rightarrow [0,1]$ as representation of the posterior imprecise probability for the climate model parameters to conduct an assessment of future climate change in the remaining part of this thesis.

# Chapter 5

# Imprecise Probability of Future Climate Change and Consequences for Climate Policy

In the preceding chapter, we have constructed an imprecise posterior probability for the climate model parameters that amalgamates the likelihood information from Chapter 2 with the imprecise prior probability determined in Chapter 3. We are now in a position to assess the uncertainty about future climate change under several scenarios for future anthropogenic greenhouse gas and aerosol emissions. To estimate the uncertainty about the future temperature increase, we need to project the uncertainty about the model parameters $\theta = (T_{2x}, \kappa_v, Q_{\text{S90}})$ onto the state variables, also called *prognostic* variables, of the energy balance model. In Section 5.1 we investigate a method that achieves this goal for the special case of parameter uncertainty that is quantified by belief functions. Section 5.2 applies the method to generate imprecise probabilities for the temperature increase in the 21st century for several SRES emissions scenarios (Nakićenović and Swart, 2000). The imprecise probabilities for the outcome of climate policy scenarios that aim to stabilise the atmospheric $CO_2$ concentration are assessed in Section 5.3.

An integrated assessment of climate change calls for a policy analysis under uncertainty. Therefore, it is important that we also address this key element of the assessment process. We have pointed out before that a decision analysis under imprecise probabilities cannot simply dwell on the classical expected utility criterion or other probabilistic criteria, because all these measures will become intervals in the presence of sets of probabilities. Hence, criteria that represented a weak ordering $\succeq$ of acts under additive probabilities, will only represent a partial ordering in the presence of imprecision. This formal argument reveals the fundamental difference between a decision under classical probabilistic uncertainty and a decision under imprecise probabilistic uncertainty or *ambiguity*, as it is often called in the economics literature.

We have mentioned the importance of ambiguity aversion for the evaluation of climate policies in our introduction, and a more detailed discussion can be found in

Kriegler et al. (2005). Since the focus of this analysis lies on the evidential foundation of imprecise probabilities for the integrated assessment of climate change, we will only briefly summarise a variety of decision criteria under ambiguity in Section 5.4. To demonstrate their applicability, we will conduct a stylised decision analysis of stabilisation policies under imprecise probability.

## 5.1 Projecting belief functions onto the model output

Consider a dynamical model of some causal relationship, which generates a transfer function $f : \mathbb{R}^n \to \mathbb{R}^m\,,\ y = f(x)$. Here $y$ denotes the state variables of the model, while $x$ constitutes the set of uncertain model inputs comprising model parameters and uncertain initial conditions. Then knowledge of the differential model equations and the remaining initial conditions suffices to determine the transfer function $f$. In our application the initial conditions are set to preindustrial “equilibrium” conditions, and the uncertainty is constituted by imprecise information about the model parameters $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$.

Let the uncertainty in the input variables $x$ be described by $\Gamma_X(bel_{\mathcal{E}}) := \{\, P_X \,|\, \forall\, A \in \mathcal{R}^n \;\; bel_{\mathcal{E}}(A) \leq P_X(A) \,\}$, i.e. a structure with a belief function as lower envelope. The associated random set $(\mathcal{E}, m) = \{(E_1, m_1), ..., (E_k, m_k)\}$ can be transferred to the model output $y$ by applying the so-called extension principle for random sets. It appeared already in Yager (1986) and was later named in analogy to the extension principle for fuzzy sets by Dubois and Prade (1991). The extension principle is defined by

$$f(E_i) := \{\, y \,|\, \exists\, x \in E_i \;\; y = f(x) \,\}\,, \qquad m_f(B) := \sum_{f(E_i)=B} m_i\,. \tag{5.1}$$

We denote the transferred random set by $(f(\mathcal{E}), m_f)$ and the associated belief function by $bel_{f(\mathcal{E})}$.

The question remains whether Extension (5.1) is a useful tool to transfer the convex set of probabilities $\Gamma_X(bel_{\mathcal{E}})$ to the model output space? Let the transfer function $f : \mathbb{R}^n \to \mathbb{R}^m$ be Borel measurable, i.e., $\forall\, B \in \mathcal{R}^m \;\; f^{-1}(B) = \{\, x \in \mathbb{R}^n \,:\, f(x) \in B \,\} \in \mathcal{R}^n$. Then every probability measure $P_X$ on $(\mathbb{R}^n, \mathcal{R}^n)$ is transformed by the mapping $f$ into a probability measure $P_Y$ on $(\mathbb{R}^m, \mathcal{R}^m)$ defined by $\forall\, B \in \mathcal{R}^m \;\; P_Y(B) := P_X(f^{-1}(B))$. Using this definition, we can generate the transformed set of probabilities

$$f(\Gamma_X(bel_{\mathcal{E}})) := \{\, P_Y \,|\, \exists\, P_X \in \Gamma_X(bel_{\mathcal{E}}) \;\; \forall\, B \in \mathcal{R}^m \;\; P_Y(B) = P_X(f^{-1}(B)) \,\}\,.$$

Thus, the question is how $f(\Gamma_X(bel_{\mathcal{E}}))$ relates to the set of probabilities $\Gamma_Y(bel_{f(\mathcal{E})})$ encompassed by the extended belief function $bel_{f(\mathcal{E})}$? To give a satisfactory answer, we need to know how to define a probability measure $P_X$ from a probability $P_Y$ on $\mathcal{R}^m$.

**Lemma 5.1** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a Borel measurable transfer function, whose range is a Borel set, i.e., $Rg(f) \in \mathcal{R}^m$. Let $\mathcal{F}$ be the collection of inverse images of the Borel sets, i.e., $\mathcal{F} := \{f^{-1}(B) \,|\, B \in \mathcal{R}^n \cap Rg(f)\}$. Define for each $P_Y$ on $\mathcal{R}^m \cap Rg(f)$ a set* *function $P_{X|\mathcal{F}}$ on $\mathcal{F}$ by*

$$\forall\, A \in \mathcal{F} \;\; P_{X|\mathcal{F}}(A) := P_Y(f(A))\,. \quad \textit{Then,}$$

*(a) $\mathcal{F}$ is a $\sigma$-field of subsets of $\mathbb{R}^n$ with $\mathcal{F} \subseteq \mathcal{R}^n$,*

*(b) the atoms of $\mathcal{F}$ are constituted by the sets $f^{-1}(y)$ with $y$ in the range of $f$,*

*(c) $P_{X|\mathcal{F}}$ is a countably additive probability measure on $\mathcal{F}$.*

Using the definition of $P_{X|\mathcal{F}}$ presented in Lemma 5.1, we can transform each element of $\Gamma_Y(bel_{f(\mathcal{E})})$ to a probability measure on $\mathcal{F}$, generating

$$f^{-1}(\Gamma_Y(bel_{f(\mathcal{E})})) := \{\, P_{X|\mathcal{F}} \,|\, \exists\, P_Y \in \Gamma_Y(bel_{f(\mathcal{E})}) \;\; \forall\, A \in \mathcal{F} \;\; P_{X|\mathcal{F}}(A) = P_Y(f(A))\,\}\,.$$

The following theorem shows that Extension (5.1) yields indeed a belief function $bel_{f(\mathcal{E})}$, which describes the transformed set of probabilities $f(\Gamma_X(bel_\mathcal{E}))$ in a meaningful manner.

**Theorem 5.1** *Let $\mathcal{R}^n$, $\mathcal{R}^m$ be Borel fields, $f : \mathbb{R}^n \to \mathbb{R}^m$ a Borel measurable transfer function with $Rg(f) \in \mathcal{R}^m$.*

*Let $bel_\mathcal{E}$ be a belief function, encompassing the set of probabilities $\Gamma_X(bel_\mathcal{E})$ and $(\mathcal{E}, m)$ the associated random set. Let*

$$\Gamma_{X|\mathcal{F}}(bel_\mathcal{E}) := \{\, P_{X|\mathcal{F}} \,|\, \forall\, A \in \mathcal{F} \;\; P_{X|\mathcal{F}}(A) \geq bel_\mathcal{E}(A)\,\}$$

*be the projection of $\Gamma_X(bel_\mathcal{E})$ on $\mathcal{F}$.*

*Let $(f(\mathcal{E}), m_f)$ be the $f$-extension of $(\mathcal{E}, m)$ calculated from Equation (5.1) and $bel_{f(\mathcal{E})}$ the associated belief function. Let*

$$\Gamma_Y(bel_{f(\mathcal{E})}) := \{\, P_Y \,|\, \forall\, B \in \mathcal{R}^m \quad bel_{f(\mathcal{E})}(B) \leq P_Y(B)\,\}\,.$$

$$\textit{Then,} \qquad \begin{array}{lrcl} \textit{(a)} & f(\Gamma_X(bel_\mathcal{E})) & \subseteq & \Gamma_Y(bel_{f(\mathcal{E})})\,, \\ \textit{(b)} & \Gamma_{X|\mathcal{F}}(bel_\mathcal{E}) & \supseteq & f^{-1}(\Gamma_Y(bel_{f(\mathcal{E})}))\,. \end{array}$$

Theorem 5.1 shows that by applying the extension principle (5.1), we are not unwittingly adding information by excluding probabilities in $f(\Gamma_X(bel_\mathcal{E}))$ from the set of probabilities $\Gamma_Y(bel_{f(\mathcal{E})})$. Whether we might lose some information, i.e., $\Gamma_Y(bel_{f(\mathcal{E})}) \supset f(\Gamma_X(bel_\mathcal{E}))$, is more difficult to assess. No information will be lost if each probability measure $P_{X|\mathcal{F}} \in \Gamma_{X|\mathcal{F}}(bel_\mathcal{E})$ can be extended onto the larger Borel field $\mathcal{R}^n$ in such a manner that the extended probability $P_X$ is contained in $\Gamma_X(bel_\mathcal{E})$. Whether this is possible will depend on the type of transfer function $f$ as well as belief function $bel_\mathcal{E}$. Fetz and Oberguggenberger (2004) have shown that in any case

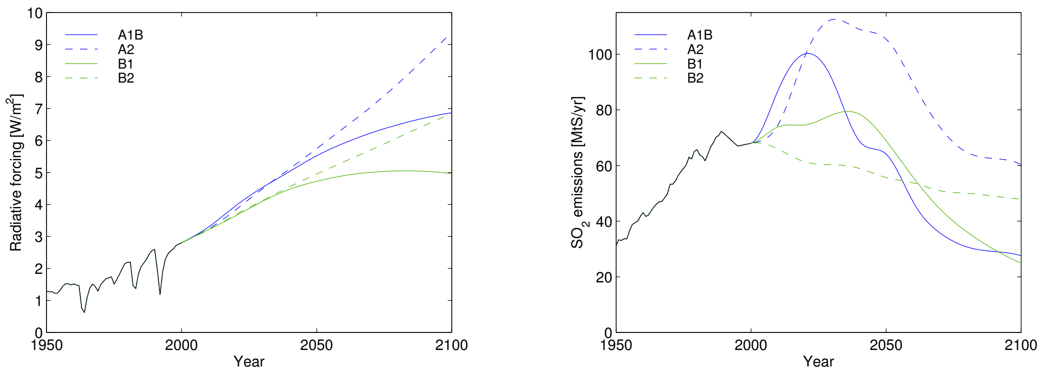
$$\inf_{P_Y \in f(\Gamma_X(bel_\mathcal{E}))} P_Y(B) = \inf_{P_Y \in \Gamma_Y(bel_{f(\mathcal{E})})} P_Y(B)$$

for arbitrary events $B \in \mathcal{R}^n$.

## 5.2 Uncertainty about the warming in the 21st century

In this section we will apply extension (5.1) to project the Dempster-updated random set $(\mathcal{G}, \nu)_{||^*\hat{T}}$ derived in Section 4.3.1 onto a random set $(\mathcal{D}, \nu)$ for global mean temperature change in the 21st century. In line with many other assessments of 21st century temperature change, we will use the four marker scenarios A1B, A2, B1, and B2 of the Special Report on Emissions Scenarios (SRES) of the IPCC (Nakićenović and Swart, 2000) as a proxy for the set of plausible assumptions about anthropogenic emissions of greenhouse gases (GHG) and aerosols in the 21st century. The SRES scenarios specify a host of socio-economic indicators and GHG emissions. Here we are only interested in the resulting radiative forcing trajectory. We have deduced this information from IPCC (2001, Appendix II, Table II.3.11). The tabulated forcing values were employed in the Third Assessment Report (TAR) of the IPCC to estimate the temperature increase in the 21st century (Cubasch and Meehl, 2001, Fig. 9.13.a). Since we want to include the uncertainty about the cooling effect from sulphate aerosols into our projections, we have to separate the direct and indirect sulphate forcing from the remaining part of the radiative forcing trajectory. This was achieved by using Parameterisation (2.9) to convert the sulphur dioxide emissions in the SRES marker scenarios to a radiative forcing, and subtracting it from the total radiative forcing. As a result, we obtained a radiative forcing trajectory for the combined effect of greenhouse gases and carbonaceous aerosols for each SRES marker scenario. This trajectory can now be combined with the cooling effect of sulphate aerosols for different assumptions about the sulphate forcing $Q_{\text{S90}}$ in 1990. Fig. 5.1 depicts the development of combined GHG and carbonaceous aerosol forcing (left panel) and $SO_2$ emissions (right panel) under the 4 marker scenarios considered here. It can be seen that the A2 scenario yields the largest increase in forcing followed by the A1B scenario. Both scenarios describe a fast growing world with little emphasis on sustainable production patterns (Nakićenović and Swart, 2000). Such production patterns are accounted for in the B1 and B2 scenarios, which consequently lead to lower radiative forcing projections.

For a given SRES scenario about the anthropogenic forcing in the 21st century and a given assumption about the climate model parameters $\theta = (T_{2x}, \kappa_v, Q_{\text{S90}})$, we can calculate the global mean sea surface and land air temperature response with our diffusion ocean-energy balance model DOECLIM, which is described by Equations (2.5) and (2.6) and numerically integrated by Equation (A.27). Since the model contains an integro-differential equation inducing a path dependency between future and past temperature change since the onset of the anthropogenic perturbation, we need to include the historical radiative forcing trajectory from anthropogenic and natural sources in our model simulation. Therefore, we have joined the SRES projection for the period 2000-2100 with the historical forcing trajectory from 1750 to 2000 as estimated in Section 2.2. For this purpose, we have fixated the solar forcing after 2000 to an inter-



a. GHGs and carbonaceous aerosols

b. $SO_2$ emissions

Figure 5.1: SRES marker scenarios A1B, A2, B1, and B2 for the anthropogenic interference with the climate system in the 21st century (Nakićenović and Swart, 2000). Panel a shows the combined radiative forcing from greenhouse gases and carbonaceous aerosols (solar forcing fixed, see text). Panel b depicts the anthropogenic $SO_2$ emissions.

mediate value for the solar cycles in the second half of the 20th century (Fröhlich and Lean, 1998; Lean, 2000). Moreover, we do not include a scenario for the volcanic forcing in the 21st century. The value of the volcanic forcing strength $\beta_{\text{Vol}}$ for the historical forcing from volcano eruptions has been chosen separately for each parameter constellation $\theta = (T_{2x}, \kappa_v, Q_{\text{S90}})$ in such a way that it maximises the likelihood $\mathcal{L}(\theta, \beta_{\text{Vol}}; \hat{T})$ of reproducing the instrumental temperature record (see Section 2.4).

Fig. 5.2 shows the resulting global mean temperature projection for the SRES A1B scenario and a selected set of parameter constellations $\theta = (T_{2x}, \kappa_v, Q_{\text{S90}})$. It can be seen that the model response is very sensitive to variations in the model parameters $\theta$. We have generated a huge range of hypotheses for the warming in the 21st century (red and black lines) even for a single radiative forcing trajectory. We can try to narrow down this range by removing all model hypotheses which are rejected by the statistical tests conducted in Section 2.5. The red lines in Fig. 5.2 depict the set of remaining temperature projections. It can be seen that the classical tests have indeed strongly restricted the band of model projections for the 20th century. However, they fail to narrow down the model projections for the 21st century decisively. This result shows that the information provided by the instrumental temperature record does not suffice to exclude low and high warming projections for the future. Therefore, it is important to use a robust Bayesian framework for assigning imprecise probabilities to the possible warming outcomes. This is exactly what we are aiming at in this analysis.

The energy balance model DOECLIM generates a transfer function $f_{t^*} : \Omega \to \mathbb{R}$ that maps the model parameters $\theta = (T_{2x}, \kappa_v, Q_{\text{S90}})$ to an increase in global mean temperature $\Delta T(t^*)$ for an arbitrary time $t^*$ in the 21st century. Following Folland and Karl (2001), we have chosen the late 19th century as baseline from which the deviation of global mean temperature is measured. In order to derive the random set $(\mathcal{D}, \nu)_{\Delta T(t^*)}$

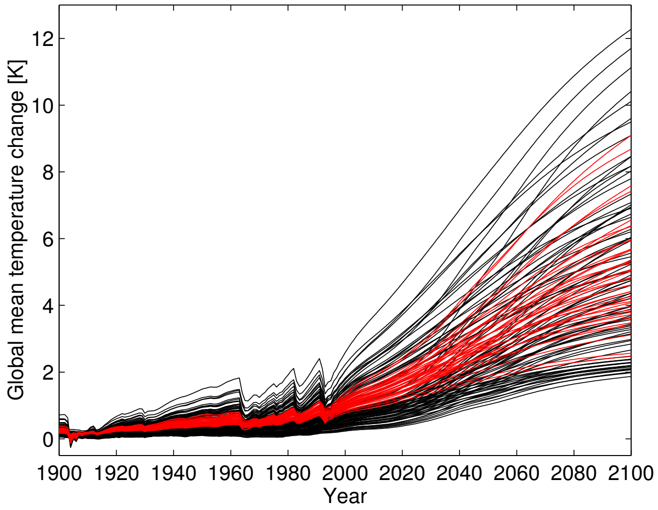


Figure 5.2: Bundle of global mean temperature projections for the SRES A1B scenario and a sample of parameter constellations $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$ (black lines). Red lines show the filtered bundle retaining only those parameter constellations which passed the statistical test in Section 2.5.

for the temperature change $\Delta T(t^*)$, Extension (5.1) requires us to calculate the image $D_{\Delta T(t^*),i} = f_{t^*}(G_{||^*,i})$ of each focal element $G_{||^*,i}$ of parameter constellations that is contained in the Dempster-updated random set $(\mathcal{G}, \nu)_{||^*\hat{T}}$ (see Section 4.3.1). Since the focal elements $G_{||^*,i} \in \mathcal{G}_{||^*\hat{T}}$ constitute connected subsets of the parameter space $\Omega$, and $f_{t^*} : \Omega \to \mathbb{R}$ is a continuous transfer function onto the real line, the resulting images need to be intervals[1], whose lower and upper bound are determined by the minimum and maximum temperature change that can be obtained from the parameter constellations $\theta \in G_{||^*,i}$. Hence, the projections $D_{\Delta T(t^*),i}$ onto global mean temperature change $\Delta T(t^*)$ are given by

$$
\begin{aligned}
D_{\Delta T(t^*),i} \quad &:= \quad [\underline{\Delta T}_i(t^*), \overline{\Delta T}_i(t^*)] \\
&\quad\text{with} \quad \underline{\Delta T}_i(t^*) = \inf_{\theta \in G_{||^*,i}} f_{t^*}(\theta)\ , \qquad \overline{\Delta T}_i(t^*) = \sup_{\theta \in G_{||^*,i}} f_{t^*}(\theta)\ .
\end{aligned}
\tag{5.2}
$$

The random set $(\mathcal{D}, \nu)_{\Delta T(t^*)}$ for $\Delta T(t^*)$ can be constructed immediately from knowledge of $D_{\Delta T(t^*),i}$ for all focal elements $G_{||^*,i} \in \mathcal{G}_{||^*}$ by means of Equation (5.1). Hence, the calculation of $(\mathcal{D}, \nu)_{\Delta T(t^*)}$ requires to find the minimum and maximum temperature change $\Delta T(t^*)$ for each focal element $G_{||^*,i}$. In principle, this could be a difficult task, since the Dempster-updated random set contains 26910 focal elements (for the $\mathcal{L}'_{22}$-

---

[1] The extension of a half-closed connected set through a continuous transfer function can result in an open, half-closed or closed interval. To avoid unnecessary technicalities, we consider the convex hull of the extended focal elements. The clear distinction of open, half-closed and closed intervals was necessary to establish Theorem 3.1, but does not influence the results almost everywhere on the space of measurable events.

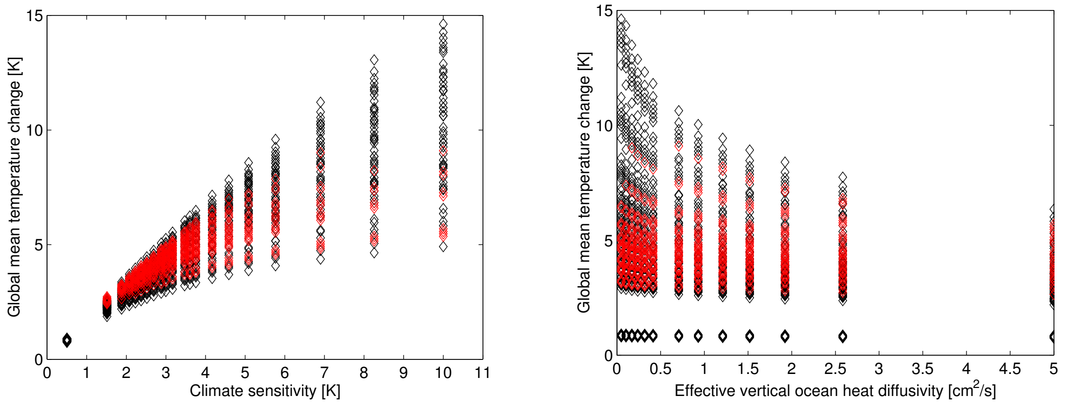
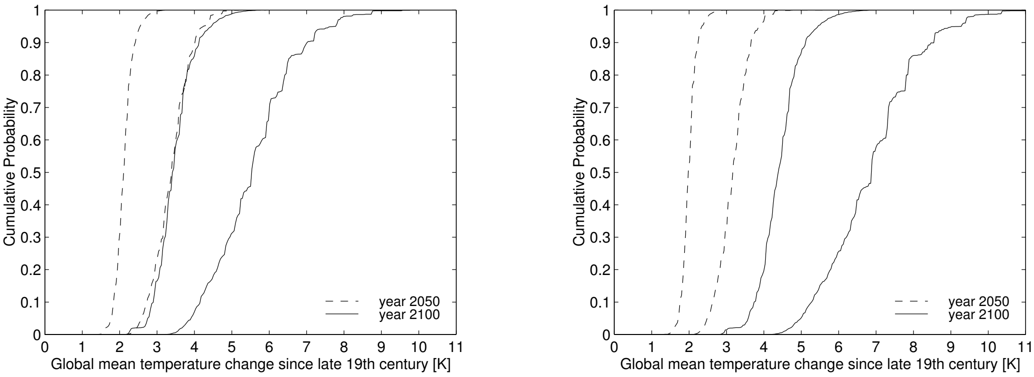


Figure 5.3: Global mean temperature change in the year 2100 (relative to the late 19th century) for the SRES A1B scenario in dependence of climate sensitivity $T_{2x}$ (Left panel, black diamonds) and ocean heat diffusivity $\kappa_v$ (Right panel, black diamonds). The temperature change for fixed $T_{2x}$ and $\kappa_v$, respectively, is smeared out due to variations in the remaining two parameters. Red diamonds show the temperature change for only those parameter constellations which passed the statistical test in Section 2.5.

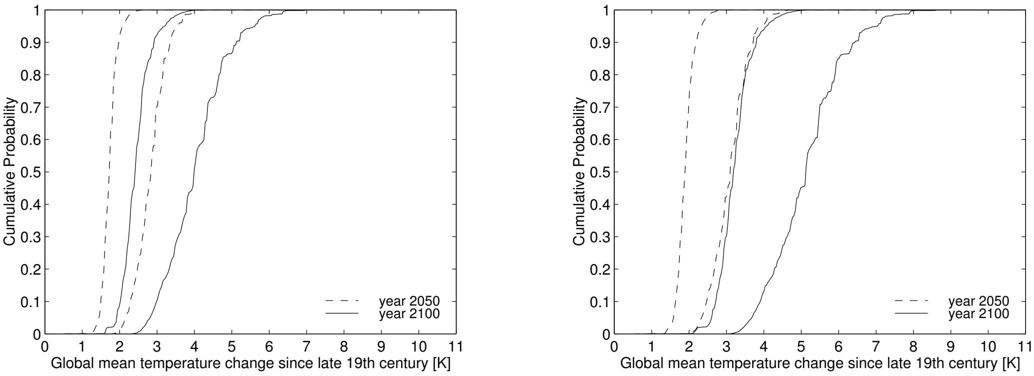
approximation, see Section 4.3.1), where only 30% of them constitute convex cubes in the three-dimensional parameter space $\Omega$. However, since all focal elements are contained in the power set of the partition $\{A_1, ..., A_k\}$ of $\Omega$, it suffices to calculate the minimum and maximum global mean temperature changes that emerge from each individual atom $A_j \in \{A_1, ..., A_k\}$. Given this information, $\underline{\Delta T}_i(t^*)$ and $\overline{\Delta T}_i(t^*)$ can be derived by collecting the atoms contained in the focal element $G_{||^*,i}$ and picking the minimum and maximum of the corresponding temperature changes.

The calculation of the random set $(\mathcal{D}, \nu)_{\Delta T(t^*)}$ can be further simplified by observing that the energy balance model DOECLIM responds monotone to changes in the parameters $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$. This is illustrated in Fig. 5.3 for the case of climate sensitivity (left panel) and ocean heat diffusivity (right panel). It can be seen that the temperature change $\Delta T(2100)$ under the SRES A1B scenario is the larger the larger the value of climate sensitivity. Likewise, the temperature change is reduced if the ocean heat diffusivity is increased. This is due to the fact that larger heat uptake of the ocean delays the surface temperature response. A similar monotone behaviour can be identified for sulphate aerosol forcing. The larger the cooling effect of the sulphate aerosols, the smaller the temperature increase.

Due to the monotone response of DOECLIM to changes in $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$, the minimum and maximum temperature change emerging from an individual atom $A_j$ of the parameter space will be realised at one of the vertices of $A_j$. Hence, all we need to do is to calculate the temperature response at the $21 \cdot 13 \cdot 15 = 4095$ vertices $\theta_{j'}$ of the atoms that define effectively the partition $\{A_1, ..., A_k\}$ for each SRES marker scenario A1B, A2, B1 and B2. Since the model DOECLIM is computationally very efficient, the required 16380 model integrations for the period 1750-2100 could be performed in less than an hour. The minimum and maximum temperature change for each atom $A_j$ can



a. SRES A1B b. SRES A2



c. SRES B1 d. SRES B2

Figure 5.4: Distribution bands for the global mean temperature change in the years 2050 and 2100 that result from the four SRES marker scenarios.

now be inferred by collecting its 8 vertices and picking the minimum and maximum of the corresponding temperature changes. This approach has been called the *vertex method* in the literature (Dong and Shah, 1987). It was advocated by Tonon et al. (2000) as an efficient way to calculate the random set extension from model parameters onto prognostic variables for complex models. Clearly, this approach is only viable as long as the model behaves fairly monotone on the individual atoms of the partition. If the model is not globally monotone, one needs to check carefully if the partition is sufficiently adapted to the model behaviour, so that non-monotonicities on the subgrid scale can be neglected. However, such a check will be particularly difficult to perform for the case of complex models. Moreover, if the number of required grid points to resolve the non-monotonicities becomes too large, the integration of the complex model at all of these points might become infeasible. Obviously, this is an particularly important issue for the case of climate models which are notoriously complex, albeit they behave surprisingly monotone in their global mean temperature response.

As explained above, knowledge of minimum and maximum temperature change emerging from each atom $A_j$ of the parameter space $\Omega$ suffices to calculate the random set $(\mathcal{D}, \nu)_{\Delta T(t^*)}$. We have shown in Theorem 5.1 that the associated belief function $bel_{\Delta T(t^*)} : \mathcal{R} \rightarrow [0,1]$ defines a structure $\Gamma(bel_{\Delta T(t^*)})$ which fully contains the transformations of the probabilities $P \in \Gamma(bel(\cdot||^*\hat{T}))$ onto global mean temperature change $\Delta T(t^*)$. Fig. 5.4 shows the resulting lower and upper probability distribution functions for $\Delta T(t^*)$ in the years 2050 and 2100 under the four SRES marker scenarios A1B, A2, B1 and B2. It is important to note that the underlying belief functions $bel_{\Delta T(t^*)} : \mathcal{R} \rightarrow [0,1]$ contain more information than the distribution bands $\Gamma_{\Delta T(t^*)}(\underline{SF}, \overline{SF})$ depicted in Fig. 5.4. This can be checked by looking at the properties of the random sets $(\mathcal{D}, \nu)_{\Delta T(t^*)}$ associated with $bel_{\Delta T(t^*)}$. They do not fulfil the requirements of Lemma 3.1, and therefore it follows from Corollary 3.1 that the belief functions $bel_{\Delta T(t^*)}$ strictly dominate the lower envelopes (see Definition 3.7) of the distributions bands $\Gamma_{\Delta T(t^*)}(\underline{SF}, \overline{SF})$. Fig. 5.4 shows that the lower and upper bounds of the distribution bands do not exhibit the step function characteristics any more that were markedly visible in the prior and Dempster-updated imprecise probability models (see Fig. 3.7 and Fig. 4.1, upper right panel). This is due to the fact that the random sets $(\mathcal{D}, \nu)_{\Delta T(t^*)}$ contain a large number of focal elements on the order of 10000 which emerged as distinct images of the 26910 focal elements $G_{||^*,i} \in \mathcal{G}_{||^*}$.

The distribution bands span a large range of possible values for the global mean warming until the end of the 21st century, no matter what scenario is considered. The corresponding lower bounds $\Delta T_*(t^*) = \overline{F}_{t^*}^{-1}(0.05)$ and upper bounds $\Delta T^*(t^*) = \underline{F}_{t^*}^{-1}(0.95)$ for the range of 90% confidence intervals supported by the probabilities $P \in \Gamma(bel_{\Delta T(t^*)})$ are listed in Table 5.1. We observe that the upper bound $\Delta T^*(2100)$ for any of the SRES scenarios except of B1 is significantly higher than the estimate of the IPCC for the global mean warming until the end of the 21st century (upper bound $\Delta T(2100) = 5.8\,\mathrm{K}$ relative to the year 1990 plus $\approx 0.4\,\mathrm{K}$ until 1990, see Cubasch and Meehl 2001). This is due to the fact that the Dempster-updated belief function $bel(\cdot||^*\hat{T})$ allows for high values of climate sensitivity up to 10 K (see Table 4.2) which have not been considered by the IPCC (upper bound $T_{2x} = 4.5\,\mathrm{K}$, Cubasch and Meehl 2001). Such high values of climate sensitivity could not be fully excluded on the basis of the likelihood information (see Section 2.5). They are responsible for the long tail towards large warmings in the 21st century that can be seen in the distribution bands.

It is interesting to note that a different result is obtained for the lower bound on the warming in the 21st century. Although we have initially included values of climate sensitivity as low as 0.5 K, they were not supported by the comparison with the instrumental temperature record (see Section 2.5). As a consequence, the Dempster-updated belief function $bel(\cdot||^*\hat{T})$ did only allow for values of climate sensitivity as low as 1.5 K (see Table 4.2), which agrees with the low estimate of the IPCC (Cubasch and Meehl, 2001). Therefore, we find a similar lower bound for the warming until the end of the 21st century as the IPCC ($\Delta T^*(2100) = 1.4\,\mathrm{K}$ relative to the year 1990 plus $\approx 0.4\,\mathrm{K}$ until 1990, see Cubasch and Meehl 2001). As a consequence, we find only a small upper probability $\overline{P}_{B1}(\Delta T(2100) \leq 2\,\mathrm{K}) = 0.076$ that the lowest SRES scenario considered here, i.e., SRES B1, will lead to a warming of less than 2 Kelvin until 2100.

| SRES scenario | A1B | A2 | B1 | B2 |
|---|---|---|---|---|
| $[\Delta T_*(2050), \Delta T^*(2050)]$ | [1.7 K,4.3 K] | [1.7 K,3.9 K] | [1.4 K,3.6 K] | [1.5 K,4.0 K] |
| $[\Delta T_*(2100), \Delta T^*(2100)]$ | [2.7 K,7.8 K] | [3.5 K,9.3 K] | [1.9 K,5.7 K] | [2.5 K,7.1 K] |

Table 5.1: Lower and upper bound for the range of 90% confidence intervals for global mean temperature increase in the years 2050 and 2100 (relative to the late 19th century) that are supported by the structure $\Gamma(bel_{\Delta T(t^*)})$.

Since the B1 scenario constitutes one of the most moderate scenarios for anthropogenic interference with the climate system, it is very unlikely that the warming until the end of the 21st century will remain below 2 Kelvin in the absence of policy intervention.

## 5.3 Uncertainty about the outcome of stabilisation policies

The temperature projections for the 21st century highlight the challenge that is faced by climate policy makers. Article 2 of the United Nations Framework Convention on Climate Change calls for a "*stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system. Such a level should be achieved within a time-frame sufficient to allow ecosystems to adapt naturally to climate change, to ensure that food production is not threatened and to enable economic development to proceed in a sustainable manner.*" (United Nations, 1992). In the aftermath of the Framework Convention, several stabilisation targets for atmospheric $CO_2$ have been investigated with respect to their consequences for the climate (e.g., Cubasch and Meehl, 2001) and for the economy (e.g., Wigley et al., 1996). The discussion about the level of greenhouse gas concentrations in the atmosphere "that would prevent dangerous anthropogenic interference with the climate system" has gained new vigour at a time where the entry into force of the Kyoto Protocol on February 16, 2005, is imminent, and the focus shifts to long-term climate protection targets to guide the implementation of the Post-Kyoto commitment period. Therefore, the investigation of the uncertainty about long-term temperature change that results from different stabilisation targets for $CO_2$ and other GHGs in the atmosphere is an important contribution to the climate policy debate which can be provided by an application of imprecise probability theory.

In this analysis, we consider four policies that aim to stabilise the atmospheric $CO_2$ concentration at 400 ppm, 450 ppm, 550 ppm and 650 ppm, respectively. The latter three stabilisation policies were adopted directly from Wigley et al. (1996). The 400 ppm policy was constructed as an intermediate case between the 350 ppm and 450 ppm policies investigated by the same authors. Fig. 5.5 shows the corresponding concentration trajectories of atmospheric $CO_2$ (solid lines). Since the other GHGs of anthropogenic origin like methane, nitrous oxide, and tropospheric ozone amplified the $CO_2$ forcing by approximately a factor of two third in the year 2000, it is important to include them in

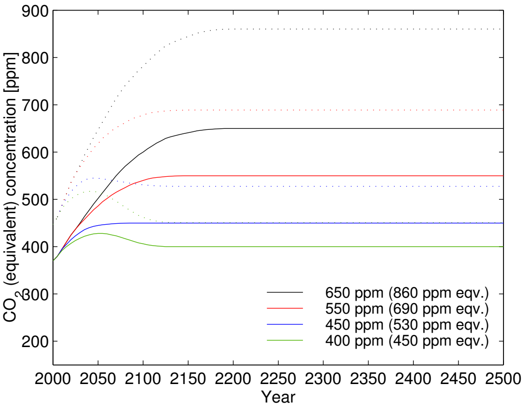
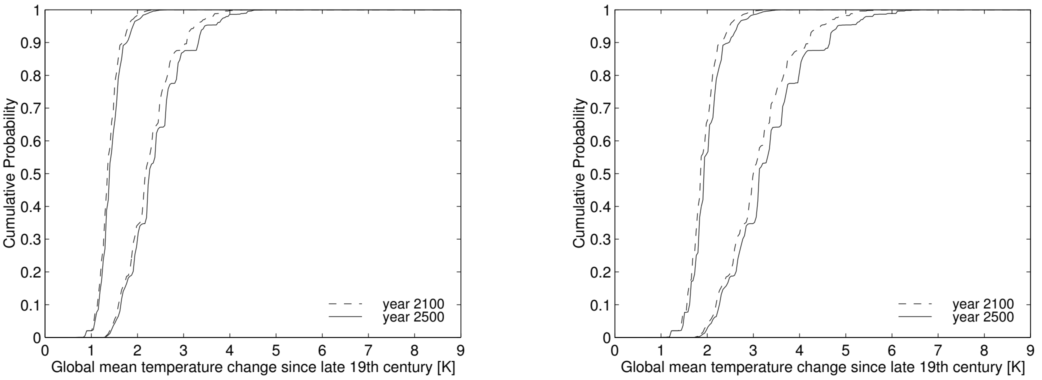


Figure 5.5: Climate policies for stabilising the atmospheric $CO_2$ content at 400, 450, 550, and 650 ppm (solid lines). The corresponding trajectories of the $CO_2$ equivalent concentration are shown by dotted lines with the same colour.

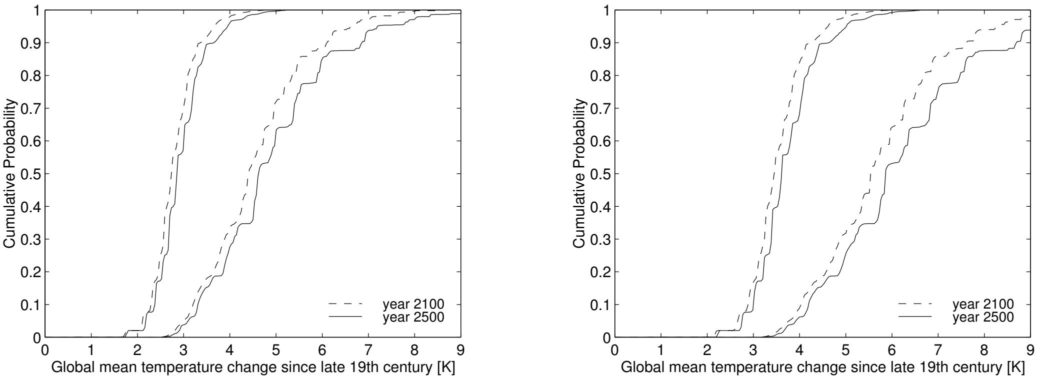
the specification of the stabilisation policy. For this investigation, we assumed that the forcing contribution of the other GHGs was reduced from 67% of the concurrent $CO_2$ forcing to 33% in a time frame of 150 years. The resulting combined radiative forcing of $CO_2$ and the other GHGs can be expressed in terms of the *$CO_2$ equivalent concentration* (in units ppm eqv.) that would lead to the same amount of radiative forcing. The trajectories for the $CO_2$ equivalent concentrations are shown in Fig. 5.5 (solid lines). It can be seen that the inclusion of the other GHGs increases the stabilisation levels considerably, in particular for the leaner stabilisation policies. The larger increase at higher levels of atmospheric $CO_2$ is due to the fact that the $CO_2$ forcing saturates as the $CO_2$ absorption bands in the atmosphere are filled with molecules. Therefore, the forcing effect of the other GHGs which occupy different absorption bands has to be mimicked by a larger portion of additional $CO_2$. Under our assumption about the forcing contribution of the other GHGs, $CO_2$ levels of 400 ppm, 450 ppm, 550 ppm and 650 ppm correspond to $CO_2$ equivalent concentrations of approx. 450 ppm eqv., 530 ppm eqv., 690 ppm eqv., and 860 ppm eqv., respectively. Finally, we have also added the cooling effect of aerosols to the forcing scenarios for the various stabilisation policies, because we want to include the uncertainty about $Q_{\mathrm{S90}}$ in our assessment. Following Cubasch and Meehl (2001), we have simply prescribed the SRES A1B $SO_2$ emissions scenario for the 21st century (see Fig. 5.1, right panel) and fixated the emissions beyond this period at the value of $E_{SO_2}(2100) = 27.6$ MtS emitted in the year 2100.

The construction of the random sets $(\mathcal{D}, \nu)_{\Delta T}$ for the global mean temperature change since the late 19th century that results from the four stabilisation policies proceeds in complete analogy to the case of the SRES marker scenarios. Therefore, we do not comment the derivation process any further here. We only mention that we have integrated the energy balance model DOECLIM up to the year 2500 in order to approach a new equilibrium of the climate system after the stabilisation of the GHG



a. 400 ppm (≈ 450 ppm eqv.)

b. 450 ppm (≈ 530 ppm eqv.)



c. 550 ppm (≈ 690 ppm eqv.)

d. 650 ppm (≈ 860 ppm eqv.)

Figure 5.6: Distribution bands for the global mean temperature change in the years 2100 and 2500 that result from the four different stabilisation policies.

concentrations in the atmosphere. However, the model continues to warm after the year 2500, albeit with very small rates. Fig. 5.6 depicts the resulting lower and upper distribution functions for the temperature change in the years 2100 and 2500. It can be seen that the step function characteristics of the Dempster-updated lower and upper distribution functions for climate sensitivity (Fig. 4.1, upper right panel) begin to re-emerge in the distribution bands for $\Delta T(2500)$. This indicates a close approach to the equilibrium temperature change which is determined predominantly by climate sensitivity, since the aerosol cooling effect is small.

Fig. 5.6 delivers a disturbing perspective on the prospects of the stabilisation policies to protect the climate. The uncertainty is large enough that even for the most stringent stabilisation level of 400 ppm $CO_2$ a warming in excess of 4 Kelvin cannot be excluded. The situation is even less favourable for the higher stabilisation levels. As discussed above, the large values of climate sensitivity up to 10 K that could not be discarded from the Dempster-updated belief function $bel(\cdot||^*\hat{T})$ are responsible for the long tails of the distribution bands reaching out to high warming values. Therefore, it will be of

| Stabilisation policies | $\Delta T \leq 2\,\mathrm{K}$ | $2\,\mathrm{K} < \Delta T \leq 4\,\mathrm{K}$ | $\Delta T > 4\,\mathrm{K}$ |
|---|---|---|---|
| 400 ppm (≈ 450 ppm eqv.) | [0.30, 0.97] | [0.03, 0.70] | [0, 0.02] |
| 450 ppm (≈ 530 ppm eqv.) | [0.03, 0.57] | [0.41, 0.97] | [0, 0.19] |
| 550 ppm (≈ 690 ppm eqv.) | [0, 0.02 ] | [0.27, 0.96] | [0.04, 0.73] |
| 650 ppm (≈ 860 ppm eqv.) | 0 | [0.06, 0.70] | [0.32, 0.94] |

Table 5.2: Probability intervals for three complementary outcomes of the stabilisation policies.

utmost importance for the climate policy debate to better constrain climate sensitivity from above. We have listed the probability intervals for the events that the stabilisation policies limit the global mean temperature change to 2 Kelvin, to a value between 2 and 4 Kelvin or fail to avoid a larger warming, in Table 5.2. It can be seen that the prospects of a warming below 2 Kelvin decrease rapidly when moving to leaner stabilisation levels, while the chances of obtaining a warming above 4 Kelvin increase in a similarly rapid manner. It is also important to note from Fig 5.6 that the imprecision in the distribution bands, i.e., the area between the lower and upper bound, increases with increasing stabilisation level. Hence, the leaner a stabilisation policy the more uncertain its consequence for global mean temperature change.

A warming of 2 Kelvin since preindustrial times constitutes a benchmark in the climate policy debate, since it has been frequently proposed as a climate protection target in its own right to prevent dangerous anthropogenic interference with the climate system (WBGU, 1995). Recently, the European Union has adopted this goal of limiting warming to 2 Kelvin above the preindustrial level. Table 5.2 shows that such a climate protection goal cannot be easily identified with a stabilisation target for the concentration of greenhouse gases in the atmosphere. The numbers tell us, however, that it requires a very stringent stabilisation level of approximately 450 ppm $CO_2$ equivalent to obtain a significant lower probability of meeting the 2 Kelvin target.

## 5.4 Climate policy analysis under imprecise probability

As a final step of our analysis we want to evaluate the stabilisation policies considered above with a stylised decision analysis under imprecise probability. Formally, each stabilisation policy can be considered as an act $f : S \to X$ with outcome $X$ (in whatever measure), which depends on a complex set $S$ of possible states of the world. In our context $S$ is constituted by the set of possible hypotheses $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}}) \in \Omega$ for the climate model parameters. The outcome $X$ shall be quantified in terms of the global mean temperature change $\Delta T$ since the late 19th century. In order to conduct a climate policy analysis, we also would need to include the socio-economic consequences of a stabilisation policy into the outcome vector $X$. This would require to account for the uncertainty about pertinent socio-economic factors in the set $S$ of possible states of the world. Obviously, a serious effort in this direction is beyond the scope

| Stabilisation policies | Objectives Article 2 | | |
|---|---|---|---|
| | Climate ($\Delta T \leq 2\,\mathrm{K}$) | Socio-economic | All |
| 400 ppm ($\approx$ 450 ppm eqv.) | [0.30, 0.97] | [0,$\overline{p}_{400}$] | [0, 0.97$\overline{p}_{400}$] |
| 450 ppm ($\approx$ 530 ppm eqv.) | [0.03, 0.57] | [0,1] | [0, 0.57] |
| 550 ppm ($\approx$ 690 ppm eqv.) | [0, 0.02 ] | [$\underline{p}_{550}$,1] | [0, 0.02 ] |
| 650 ppm ($\approx$ 860 ppm eqv.) | 0 | 1 | 0 |

Table 5.3: Stylised probability intervals for meeting the climate, socio-economic and combined objective of Article 2 of the UNFCCC. See text for further explanation.

of this thesis. However, it is our long term goal to combine the imprecise probability for the climate model parameters presented here with an imprecise probability for the key parameters of an economic growth model to the development of which we have contributed (Edenhofer et al., 2005). Such key parameters are constituted by, inter alia, the elasticity of substitution between energy, labour and capital, the impact of research and development investments on labour and energy efficiency, the learning rate in the renewable energy sector, and the remaining fossil resource base.

In the stylised decision analysis presented here we will short-circuit the socio-economic dimension of the climate change problem by simply considering two possible socio-economic outcomes of a stabilisation policy: it either meets the related objective of Article 2 of the United Nations Framework Convention on Climate Change to "*ensure that food production is not threatened and to enable economic development to proceed in a sustainable manner*", or it does not. For the sake of demonstration, we simply stipulate imprecise probabilities for these two possible outcomes of each stabilisation policy (see Table 5.3, second column). Our stylised probability bounds reflect an increase in socio-economic viability with an increase in stabilisation level (which, of course, can be questioned) as well as the large imprecision about the socio-economic outcome of the stabilisation policies (with the exception of the 650 ppm policy).

The imprecise probability for meeting the socio-economic objective of Article 2 has to be contrasted with the imprecise probability for observing its climate protection goal to "*prevent dangerous anthropogenic interference with the climate system*". Here, we will adopt the climate policy goal of the European Union to limit warming since preindustrial times to 2 Kelvin as a possible specification of the climate protection objective of Article 2. The imprecise probability of a temperature change $\Delta T \leq 2\,\mathrm{K}$ is a direct outcome of our analysis and has already been looked at in the preceding section. The resulting probability intervals for the 4 stabilisation scenarios are listed in Table 5.3 (first column).

The overall imprecise probability for fulfiling the objectives of Article 2 can now be calculated from the imprecise probabilities of the socio-economic and climatic component. If we assume independence between global mean temperature change $\Delta T$ and the economic consequences of a given stabilisation policy, such a calculation is straightforward. Let $A \times X_{\mathrm{SE}} \subset X$ be the event of observing the 2 Kelvin target, and $X_{\mathrm{CL}} \times B \subset X$

the event of fulfiling the socio-economic requirements of Article 2. Then the joint fulfilment of all requirements, i.e., the event $A \times X_{\text{SE}} \cap X_{\text{CL}} \times B = A \times B$ exhibits the imprecise probability

$$[\,\underline{P}(A \times B),\ \overline{P}(A \times B)\,] = [\,\underline{P}(A) \cdot \underline{P}(B),\ \overline{P}(A) \cdot \overline{P}(B)\,]\ . \tag{5.3}$$

Table 5.3 (third column) shows the resulting imprecise probabilities for the complete fulfilment of Article 2 by the 4 stabilisation policies. It is important to note that Equation (5.3) holds unanimously whether epistemic, strong or random set independence is assumed (Couso et al., 2000, see also Section 3.5), because all three independence concepts lead to the same result on product sets $A \times B$ (Walley, 1991, 9.3.5). Assuming independence between temperature change and the economic outcome of a given stabilisation policy can be justified on the grounds that the uncertainty about these quantities arises from factors located either in the realm of the natural system or in the realm of the socio-economic system.

The situation listed in Table 5.3 constitutes a decision problem under imprecise probability. What stabilisation policies are superior to meet the objectives of Article 2 of the UNFCCC? In the following, we discuss briefly a number of decision criteria under imprecise probability that have been proposed and axiomatically justified in the literature. All these criteria are based on the notion of a lower and upper expectation, or *prevision*, of *gambles* $G_f = u \circ f : S \to \mathbb{R}$ constituting the combination of an act $f : S \to X$ and a real-valued utility function $u : X \to \mathbb{R}$ for the set of possible outcomes. This framework is general enough to accommodate the special case listed in Table 5.3 where we want to evaluate the acts according to their implication for the lower and upper probability of a specific event $A \times B \subset X$. This case can be described by the *indicator gamble* $I_{C(f)} : S \to \{0, 1\}$ for the set $C(f) = \{s \in S | f(s) \in A \times B\}$ of all 'good' states of the world in which the act $f$ can meet the objective of Article 2. $I_{C(f)}$ pays one unit of utility if $s \in C(f)$ and nothing in the remaining case. As briefly mentioned in Appendix C, the lower prevision (expectation) of the indicator gamble $\underline{Pr}(I_{C(f)})$ indeed equals the lower probability $\underline{P}(C(f)) = \underline{P}_f(A \times B)$[2]. Therefore, decision criteria that establish a (partial) ordering of gambles will also induce a (partial) ordering of sets $C(f)$ of 'good' states of the world, in which the objective of the UNFCCC can be met for a given stabilisation policy $f$. Since the sets $C(f)$ depend directly on the policies, this piece of information allows to rank the policies themselves.

**Maximality:** The criterion of maximality was proposed by Walley (1991, Chapter 3.9) and received an axiomatic foundation by Seidenfeld et al. (1995). A gamble $G : S \to \mathbb{R}$ is called *maximal* if it is not dominated by any other available gamble $F \in \mathcal{K}$

---

[2] It is important to note that the lower probability for the event $A \times B$ of meeting the UNFCCC objective with a stabilisation policy $f$ is determined by the lower probability of the set $C(f)$ of 'good' states of the world, i.e. $\underline{P}_X(A \times B) := \underline{P}_S(f^{-1}(A \times B)) = \underline{P}_S(C(f))$. We need to assume that $f$ is measurable for this definition to make sense (see Section 5.1).

in terms of

$$\forall\, F \in \mathcal{K} \qquad \underline{Pr}(F - G) \le 0\,. \tag{5.4}$$

Walley (1991) motivated the criterion on the basis of a lower and upper betting rate interpretation of imprecise probabilities. If your infimum buying price for all gambles $F - G$ is smaller or equal to zero, then you are not willing to pay a positive price for exchanging the gamble $G$ with any other gamble $F$. It is important to note that maximality defines only a partial ordering of gambles. In general, there will exist more than one maximal gamble. This will also happen frequently in practice, since maximality is the weakest criterion that is discussed here.

In our special case of meeting Article 2 of the UNFCCC, a stabilisation policy $g$ is maximal in the set of policies $\mathcal{F}$, if

$$\forall\, f \in \mathcal{F} \quad \underline{Pr}(I_{C(f)} - I_{C(g)}) \overset{\text{2 monotone}}{=} \underline{P}(C(f) \cap C(g)^c) - \overline{P}(C(g) \cap C(f)^c) \le 0\,, \tag{5.5}$$

where the equality holds if and only if the imprecise probability about the possible states of the world can be described by a *2-monotone* lower probability $\underline{P} : \mathcal{A}(S) \to [0,1]$ (see Definition C.10). In this case, the lower prevision $\underline{Pr}(I_{C(f)} - I_{C(g)})$ equals the *Choquet integral* of the indicator gamble $I_{C(f)} - I_{C(g)}$, which can be evaluated immediately to yield the expression on the right-hand side of the equality (see Definition C.11). Since the uncertainty about the climate model parameters is described by a belief function, we can assume 2-monotonicity of $\underline{P}$ in our stylised decision analysis.

Hence, we have to evaluate the expression on the right-hand side of Equation 5.5 for each combination of stabilisation policies at 400, 450, 550, and 650 ppm $CO_2$. The situation is complicated by the fact that we have not specified the socio-economic constituents of the possible states of the world, and therefore cannot determine the sets $C(f)$ of 'good' states explicitly. However, this is not necessary in this specific decision situation because the lower probability $\underline{P}_f(A \times B) = \underline{P}(C(f)) \ge \underline{P}(C(f) \cap C(g)^c)$ is zero for all stabilisation policies under consideration (see Table 5.3). As a consequence, $\underline{Pr}(I_{C(f)} - I_{C(g)}) \le 0$ for all combinations $g, f$ of stabilisation policies, and therefore all of them are maximal under Definition (5.4). Obviously, the maximality criterion is too weak to discriminate between the decision alternatives. In our particular case, it even assigns the status of maximality to the 650 ppm stabilisation policy which will fail to meet the Article 2 objective with certainty. This is an unreasonable result. We can remove the 650 ppm policy from the set of maximal policies by requiring a stronger version of maximality, i.e.,

$$\forall\, F \in \mathcal{K} \qquad \underline{Pr}(F - G) < 0 \quad \text{or} \quad \underline{Pr}(G - F) \ge 0\,.$$

The behavioural interpretation is that for any gamble $F$ you have either to be paid to exchange $G$ in return for $F$, or you would always be willing to exchange $F$ in return for $G$ without payment. Since $\overline{P}(C(g) \cap C(f)^c) \le \overline{P}(C(g) = 0$ for the $g = 650\,\text{ppm}$ stabilisation policy independently of the choice of $f$, we find $\underline{Pr}(I_{C(f)} - I_{C(g)}) = 0$ for all $f \in \mathcal{F}$. Since there will exist some policies $f \in \mathcal{F}$ with $\underline{Pr}(G - F) < 0$, the 650 ppm stabilisation policy does not fulfil the strengthened condition of maximality.

**$E$-admissibility:** This criterion was formulated for decisions under uncertainty that is described by a (not necesarily convex) set of probabilities $\mathcal{M}$ (Levi, 1974, 1980). A gamble $G \in \mathcal{K}$ is *$E$-admissible* if there exists a classical probability $P \in \mathcal{M}$, for which the expected utility $E_P(G)$ reaches a maximum on the set of all gambles $\mathcal{K}$, i.e.,

$$\exists\, P \in \mathcal{M} \quad \forall F \in \mathcal{K} \;\; E_P(G) \geq E_P(F)\ . \tag{5.6}$$

An underlying rationale for $E$-admissibility can be constructed, inter alia, in the context of group decisions. A group of Savage-type decision makers is imagined that hold subjective probabilities, which are compatible with the ambiguous state of information. For each $E$-admissible gamble, a decision maker in this group might be found who supports this act. Thus, the size of the $E$-admissible choice set indicates the scope for controversy that might arise in a group of expected utility maximizers due to the presence of ambiguity. There exists a clear relationship between $E$-admissibility and maximality discussed above. Every $E$-admissible gamble is also maximal, but the reverse is only true if the set of probabilities $\mathcal{M}$ is convex (Walley, 1991, Theorem 3.9.5). Schervish et al. (2003) provide a nice example of a discrete set of probabilities, for which the choice set of $E$-admissible gambles is a true subset of the maximal gambles.

In our special case, $E$-admissibility of a stabilisation policy $g \in \mathcal{F}$ boils down to the existence of a probability measure $P \in \mathcal{M}$ under which $g$ yields the maximal probability of meeting the objectives of Article 2, i.e., $P(C(g)) \geq P(C(f))$ for all $f \in \mathcal{F}$. Hence, for identifying the $E$-admissible stabilisation policies, we need to reconstruct the set of probability measures $\mathcal{M}$ that gave rise to the imprecise probabilities of meeting Article 2 listed in Table 5.3 (third column). This can be done even without having specified the socio-economic dimension of the possible states of the world explicitly. Since we are only interested in two different outcomes of the stabilisation policies (to meet or not to meet the objective of Article 2), we can base the reconstruction of $\mathcal{M}$ on a reduced event space that is generated by the sets $C(f) \subset S$ of 'good' states of the world associated with the stabilisation policies $f \in \mathcal{F}$. The field of sets generated from $\{C(400\,\text{ppm}), C(450\,\text{ppm}), C(550\,\text{ppm}), C(650\,\text{ppm})\}$ by union and complementation contains $2^{16}$ elements that can be derived from a partition of $S$ with 16 elementary events $\{C_1, ..., C_{16}\}$. Then, the set of probabilities $\mathcal{M}$ contains all probability mass functions that distribute their masses over the 16 elementary events in a way which respects the lower and upper probability constraints of meeting Article 2 (see Table 5.3). To identify a stabilisation policy as $E$-admissible, it suffices to construct a probability mass function compatible with the constraints, which maximises the probability of meeting Article 2 for the given policy. For the sake of brevity, we do not describe the tedious calculation explicitly here. We can find probability mass functions in $\mathcal{M}$, under which the 400 ppm, 450 ppm and 550 ppm stabilisation policies maximise the probability of meeting Article 2. Obviously, it is impossible to identify such a mass function for the 650 ppm policy, since it will fail to meet the objectives of Article 2 with certainty. Hence, only the lower three stabilisation policies are $E$-admissible. Nevertheless, the criterion of $E$-admissibility is too weak to narrow down the choice set of policy options decisively, which highlights the scope for controversy in the climate policy arena.

**Γ-Maximin:** The Γ-Maximin criterion is maybe the most prominent decision criterion under ambiguity that can be found in the economic literature. It was axiomatically justified by Gilboa and Schmeidler in a series of seminal papers (Gilboa, 1987; Gilboa and Schmeidler, 1989; Schmeidler, 1989), in which they constructed various decision theories under uncertainty that accommodate the phenomenon of ambiguity aversion as identified in the famous experiment of Ellsberg (1961) (see introduction to this thesis). Their decision theories motivated both convex sets of probabilities as well as 2-monotone lower probabilities as eligible models of subjective belief (using the classical frameworks of Savage 1954 and Anscombe and Aumann 1963; see Casadesus-Masanell et al. 2000 for the completion of the cycle of theories).

Unlike the aforementioned decision criteria, Γ-Maximin induces a weak ordering $\succeq$ of available gambles $G \in \mathcal{K}$ in terms of their lower expected value, i.e.,

$$g \succeq f \quad \Leftrightarrow \quad \underline{Pr}(G) \geq \underline{Pr}(F) \ . \tag{5.7}$$

Hence, the set of optimal gambles includes all $G \in \mathcal{K}$ which maximise the lower expected outcome. It is interesting to note that there can be Γ-maximin optimal gambles that are not $E$-admissible (see Schervish et al. 2003 who give an example).

In the special case of our application, the optimal stabilisation policy under Γ-Maximin will be the one which maximises the lower probability of meeting Article 2. However, Table 5.3 shows that the lower probability for this event is zero across all stabilisation policies. Hence, a Γ-Maximin decision maker will be indifferent among them. This type of indifference might be best described as the 'agony of the desperate'.

**Hurwicz Criterion:** The failure of Γ-Maximin to yield a useful decision support in our example indicates that it is sometimes important to include information on the upper expected outcome in the decision criterion. Such a criterion has been already proposed by Hurwicz (1951) for the case of complete ignorance within the range of outcomes of the acts under consideration. The proposal is referred to as Hurwicz criterion by Luce and Raiffa (1957) and was axiomatically justified by Arrow and Hurwicz (1972). Later on, Jaffray (1989) and Strat (1990) adopted the Hurwicz criterion in the context of belief functions. An axiomatic foundation can be found in Jaffray (1994).

In its general form for imprecise probabilities, the Hurwicz-criterion ranks the gambles $G \in \mathcal{K}$ on the basis of a linear combination of lower and upper expected outcomes,

$$\begin{aligned} G \succeq F \qquad \Leftrightarrow \qquad & \alpha\, \underline{Pr}(G) + (1-\alpha)\, \overline{Pr}(G) > \\ & \qquad \alpha\, \underline{Pr}(F) + (1-\alpha)\, \overline{Pr}(F) \ , \qquad \text{with} \quad \alpha \in [0,1] \ . \end{aligned} \tag{5.8}$$

$\alpha$ is a security index that determines the proportion in which considerations of security, related to the lower expected outcome $\underline{Pr}(G)$ and opportunity, related to the upper expected outcome $\overline{Pr}(G)$ influence the decision criterion. The security index adds an additional degree of freedom to the policy analysis. It allows to explore the continuum between completely security-based ($\alpha = 1$) and completely opportunity-based decision rules ($\alpha = 0$). Note that the $\Gamma$-Maximin rule is recovered for $\alpha = 1$. In an interesting paper, Chevé and Congar (2000) used this property of the Hurwicz criterion to identify precautionary behaviour in a consumption-pollution problem. Like $\Gamma$-Maximin, the Hurwicz-Criterion establishes a weak ordering of available gambles. The optimal choice is constituted by the gambles which maximise the $\alpha$-mixture between lower and upper probability in Equation 5.8.

In our special application, the Hurwicz criterion is easy to implement. For a prescribed $\alpha \in [0, 1]$, we have to search for the optimal stabilisation policy $g^*(\alpha) \in \mathcal{F}$ which maximises

$$\alpha \underline{Pr}(I_{C(g)}) + (1-\alpha)\overline{Pr}(I_{C(g)}) \quad = \quad \alpha \underline{P}_g(A \times B) + (1-\alpha)\overline{P}_g(A \times B)\ . \tag{5.9}$$

Inspection of the imprecise probabilities for meeting Article 2 (see Table 5.3) reveals that their exists only one stabilisation policy that will be endorsed unanimously for all values $\alpha < 1$. The optimal choice is constituted either by the 400 ppm or the 450 ppm stabilisation policy. It will depend on the upper probability $\overline{p}_{400}$ for the economic viability of the 400 ppm stabilisation policy under Article 2, which of the two are preferable. If $\overline{p}_{400} > 0.58$, the 400 ppm policy should be chosen. If the upper probability of economic viability is smaller than 0.58, then the 450 ppm policy is preferable.

This concludes our discussion of decision theory under imprecise probability. The list of decision criteria presented above is by no means exhaustive. However, it suffices to point out that imprecise probability theory can be embedded in a rich set of well-founded decision theories, most of which have been derived with the same rigour as expected utility theory (von Neumann and Morgenstern, 1944; Savage, 1954; Anscombe and Aumann, 1963). Moreover, the decision criteria under imprecise probability can accommodate a variety of behaviour like ambiguity aversion - or ambiguity seeking - that cannot be represented in the framework of an expected utility theory. This aspect is particularly relevant to climate policy analysis since the ambiguity about climate change and the potential mitigation options is large. We conclude that the application of imprecise probabilities in the integrated assessment of climate change is reinforced rather then discouraged by the need to conduct a viable climate policy analysis.

We have applied the various decision criteria to an important climate policy question: at which level should the atmospheric $CO_2$ concentration be stabilised in order to meet the objectives of Article 2 of the UNFCCC? This is a typical example of a policy question, which does not easily fit the expected utility concept (see Schellnhuber 1998, 1999 for a discussion of various policy paradigms in the context of climate change). It can be addressed, inter alia, in the framework of the tolerable windows approach (TWA;

Tóth et al., 1997; Petschel-Held et al., 1999; Bruckner et al., 1999). The TWA tries to identify a set of admissible policies that meet predefined constraints in the outcome space (such as the requirements of Article 2). We have shown in earlier work (Kriegler and Bruckner, 2002, 2004), which is not included in this thesis, that the admissibility of policies is very sensitive to the uncertain states of the world in which they might be realised. Hence, it is crucial to account for uncertainty in the identification of admissible policies. By now, a probabilistic extension of the tolerable windows approach has been presented, which takes an important step into this direction (Kleinen, 2005). We have shown above that the concept of indicator gambles provides another way which extends the results of conventional and non-conventional decision theories under uncertainty to the realm of the TWA. In this sense, we have pointed to a set of potential decision criteria for a tolerable windows approach under imprecise probability.

However, our main objective in evaluating a set of stabilisation policies in the light of Article 2 was to demonstrate that the introduction of imprecision does not necessarily lead to a weaker policy recommendation. As can be seen from Table 5.3 we have introduced large imprecision by making only weak assumptions about the economic viability of the stabilisation policies (with the exception of the 650 ppm policy). Nevertheless, we could successively narrow down the choice set of optimal stabilisation policies by moving - in a lexicographic manner - from weaker to stronger decision criteria. In this way, we have identified the policies with a stringent stabilisation target at either 400 ppm or 450 ppm as the best choice in order to meet the objectives of Article 2 of the Framework Convention on Climate Change. This policy recommendation reflects the information about future global mean temperature change that we have collected from the analysis in the preceding chapters. Although we have allowed for large imprecision in the specification of the prior uncertainty (see Chapter 3) and smoothed the likelihood function by optimising the nuisance parameters at each point of the uncertainty space (see Chapter 2), we have found little ambiguity about the prospects of the 550 ppm and 650 ppm stabilisation policies to limit the temperature change since preindustrial times to 2 Kelvin. The 650 ppm policy will fail to meet this objective with certainty, while the 550 ppm policy exhibits only a negligible upper probability for success. This shows clearly how imprecise probabilities can help to make a policy recommendation more robust. In order to challenge our recommendation of a stringent stabilisation target, one would need to point to important information that is discarded by the imprecise probability model. In the context of our prototypical analysis this can, of course, be done. We will discuss at length in the following summary that there is much room for improvement. However, we want to make the point here that a clear-cut policy recommendation emerging from an assessment that accounts for large uncertainty will be more trustworthy than this would be the case if part of the uncertainty was neglected.

# Chapter 6

# Summary and Conclusion

The analysis presented here has demonstrated an application of imprecise probability theory to the integrated assessment of climate change. To the best of our knowledge, it constitutes the first such application that attempts to cover the entire assessment chain from generating the imprecise probability on the basis of available evidence to a climate policy analysis. Even in a more limited context, applications of imprecise probabilities in climate change assessments have been rare (Luo and Caselton, 1997; Lange, 2000; Ha Duong, 2003; Kriegler and Held, 2005; Hall et al., 2005). This might be attributed to the fact that imprecise probability theory matured only in the last decade, a process that was particularly fostered by the seminal work of Walley (1991). Moreover, many theoretical questions relating to, e.g., dynamic decision frameworks, problems of natural extension and efficient representations of imprecise probability models, still await major advancement. This is rough terrain for the practitioner, as she cannot rely on textbook knowledge in many instances. Therefore, we have also established several theoretical results that in some cases substantiated the steps of our analysis and in other cases where necessary in order to proceed.

## 6.1 Summary of the imprecise probability analysis

In the light of the complexity of a serious climate change assessment, we have decided to rely on imprecise probability models that allow for an efficient representation in terms of a coherent lower probability with *sparse Möbius inverse* (see Section 3.1). Sparse Möbius inverses are characterised by only a small number of focal elements carrying non-zero Möbius mass. The sparseness of a Möbius inverse can be capitalised on if methods are available to process coherent lower probabilities directly on the level of focal elements.

Probability boxes (p-boxes) constitute an important example of an imprecise probability model that allows for a sparse Möbius representation (Ferson et al., 2002). We have relied on them as a starting point for constructing a prior imprecise probability for the climate model parameters, since they can be generated easily in most instances of imprecise information, including diverse probability estimates in the literature and expert elicitations (see Section 3.2.4). The algorithm to construct a random set from a p-box (see Algorithm 3.1) is well known from the literature. Adding to this, we have shown that it conforms indeed with a 'convex set of probability'-interpretation of the p-box even on continuous spaces (see Theorem 3.1). Since p-boxes contain a large class of probabilities including many that might be considered an unreasonable choice in the light of the prior information, we have attempted to narrow down the set of admissible probabilities without losing the virtue of a sparse Möbius representation. This has been achieved by including lower bounds for eligible probability mass functions on a predefined partition of the uncertainty space into the model. We provided an algorithm to construct the associated random set for this more informative imprecise probability model, and showed that it exhibits indeed the favourable properties of sparseness and non-negativity (see Algorithm 3.2, Theorem 3.2 and the preceding lemmas). To the best of our knowledge, this is a new result which might also be of interest for applications in other fields that try to process imprecise probability information on the basis of random sets.

We demonstrated the applicability of this methodological framework to an assessment of climate change by constructing random sets for the prior uncertainty about climate sensitivity, ocean heat diffusivity and sulphate aerosol forcing from information that is found in the literature. A key step in this process was the discretisation of the continuous parameter space in order to arrive at a finite uncertainty model. This step is not specific to an imprecise probability analysis as it surfaces also in conventional probabilistic applications, e.g., when processing histogrammed data. However, the partitioning of a continuous space might be guided by different objectives in the imprecise probability context. We can identify at least three objectives, which are closely linked to the three steps of a Bayesian analysis from constructing a prior model to updating the prior model and to projecting the posterior information on the prognostic quantity in question.

1. Approximation of the continuous *prior uncertainty model* as nearly as possible in order to generate the most informative imprecise prior with a predefined level of finite complexity. For the case of p-boxes, this objective was recast in an optimisation problem for partitioning a continuous space by Kriegler and Held (2005).

2. Approximation of the continuous *likelihood function* as close as possible in order to avoid overly imprecise results in the updating process. We have adopted this objective in the present analysis. However, as discussed in Section 4.3.2 and taken up below, our attempt did not suffice to avoid overly imprecise posterior probability bounds from the application of the Generalised Bayes' rule. This result depended more on the choice of prior model than on the choice of partition of the continuous parameter space. Hence, a likelihood adapted partition is a necessary, but by no means sufficient condition for obtaining meaningful results under the Generalised Bayes' rule.

3. Partitioning of the model parameter space in a manner that reflects the monotonicity properties of the *model transfer function* onto the prognostic variables. If this, admittedly very difficult, task can be achieved, methods like the vertex method might become applicable to the problem of random set extension. This would greatly simplify the extension problem because the vertex method allows to reduce an optimisation problem to a finite number of forward integrations (see Section 5.1 and 5.2). So far, we have not seen any work in this direction. Nevertheless, we believe that advancements in the computability of the random set extension are strongly needed, in particular for its application to more complex climate models.

Despite the importance of discretising the continuous parameter space, we have devoted little attention to it in this analysis. We have neither tried to identify a systematic way to reconcile the conflicting requirements of the first two objectives (the third was not in issue here, since the transfer function was globally monotone), nor have we searched the literature for advanced methods concerning adaptive grids. Therefore, we think that there is much room for improvement here.

Having established an imprecise prior probability for the climate model parameters in terms of a *belief function* and its associated random set, we went on to the problem of updating the prior information with a likelihood function. As pointed out in Section 4.1, the application of the Generalised Bayes' rule (GBR) is a thorny issue for many imprecise probability models (with the notable exception of the probability ratio class) due to the computational complexity and the presence of dilation. We have established a methodology for updating a belief function with a likelihood function via the GBR by adapting the framework of Chrisman (1995) to this particularly important case (see Theorem 4.1 and the preceding lemmas). En route, we have also derived an algorithm to apply an alternative updating rule, i.e., Dempster's rule of conditioning, to the problem of updating with likelihood information (see Algorithm 4.1). The algorithm allows us to perform the updating operation directly on the associated (sparse) random set of the belief function. This is an important result of its own, since Dempster's rule is widely used in other contexts. Moreover, Dempster's rule has been shown to be equivalent to the maximum likelihood update rule (if the joint probability on the parameter × observation space is 2-monotone; Gilboa and Schmeidler 1993) which gives its application to the lower envelope of a convex set of probabilities a clear interpretation. To the best of our knowledge, the established methodology has not appeared in the literature so far. It constitutes an achievement that might also be of interest for applications in other fields.

We have demonstrated the applicability of our version of Dempster's rule of conditioning to the prior belief function for the climate model parameters. However, we have failed to produce meaningful imprecise posterior probability bounds from the application of the GBR. We have been able to attribute this result primarily to our choice of prior model (see Section 4.3.2). If the likelihood function changes strongly on the focal elements of a belief function, it will dilate the imprecise probability model under the application of the GBR. This is an important, albeit less comforting result. Since it will be the rule rather than the exception that the prior imprecision ranges over areas of large likelihood changes, belief functions do not constitute a viable model for updating via the GBR. This highlights a critical branching point in our analysis. Either other more restrictive updating rules can be justified, or other prior models have to be constructed. The latter would impede many of the steps that we have taken in this analysis to process imprecise probabilities. Consequently, we have dwelled on the former option by employing Dempster's rule of conditioning for generating a posterior belief function. Whether this is a reasonable choice will have to be subject to close scrutiny in future research. Hence, the updating issue continues to await theoretical advancements that will help the practitioner to calculate imprecise posterior probabilities on the entire event space in a computationally efficient way (for the calculation of posterior bounds for single events or expectation values see Cozman 1999b).

The final steps of our analysis consisted in projecting the posterior belief function for the climate model parameters onto global mean temperature change and then in conducting a stylised climate policy analysis on the basis of the imprecise temperature information. Belief functions can be projected onto prognostic model variables by means of random set extension, which is a long established concept in the literature. We have added an investigation of the relationship between the convex set of projected probabilities and the convex set of probabilities defined by the extended belief function (Theorem 5.1, see also Fetz and Oberguggenberger 2004 for establishing equality of the lower envelopes). In the particular case of our analysis, the random set extension was easy to calculate due to the monotonicity properties and the computational efficiency of the energy balance model. Hence, we have spared ourselves the difficulties that will be encountered with more complex models. As mentioned above, an improvement of methods for calculating the random set extension will be an important prerequisite for the applicability of random set based imprecise probability analysis to assessments of climate change with complex climate models. Our stylised policy analysis provided an overview of several decision criteria under imprecise probability. By drawing on the concept of indicator gambles we illustrated that the decision theories are not only able to deal with imprecise expected utilities, but also apt to assess imprecise probabilities for sets of admissible outcomes. This points the way to their application to constraint-based decision making frameworks like the tolerable windows approach, which play a significant role in climate policy analysis.

## 6.2 Summary of the integrated assessment of climate change

Having discussed the methodological advances and weaknesses in our application, we now turn to the 'climate side of things'. We have tried to avoid a common fallacy when applying a new methodology to a subject, i.e., the fancier the method the less serious one can be about the subject itself. Therefore, we have made a serious effort to account for the uncertainty about the global mean climate response to a secular trend in the radiative forcing. For this purpose, we have constructed an energy balance model resolving land and ocean boxes with a 1-D diffusive interior ocean (DOECLIM). The model formulation was fairly standard besides the fact that we have solved the heat diffusion in the interior ocean analytically. As a consequence, the model could be closed by adding a memory to the differential equation for sea surface temperature. We calibrated DOECLIM against the seasonal cycle and an ensemble of $CO_2$ doubling experiments with the climate system model of intermediate complexity CLIMBER-2. Due to the calibration process, the number of free parameters could be reduced to just two key quantities for the climate response to secular changes in the radiative forcing: global climate sensitivity $T_{2x}$, and effective vertical ocean heat diffusivity $\kappa_v$. The model compared well against $CO_2$-multiplication experiments with CLIMBER-2 and the coupled atmosphere-ocean general circulation model HadCM3. It was also able to reproduce the instrumental temperature record in the 20th century, and the total heat uptake of the world ocean after 1950. Therefore, we conclude that the model provides a satisfactory description of the global mean temperature response to secular trends in radiative forcing.

In the following, we have interpreted DOECLIM in a statistical sense as representing a set of hypotheses about the global mean climate response to secular forcing trends. The space of possible hypotheses was spanned by the model parameters $T_{2x}$, $\kappa_v$, and the direct and indirect sulphate aerosol forcing in 1990, $Q_{\mathrm{S90}}$. This set of parameters constitutes a frequent choice in a series of widely received studies on the uncertainty about climate change (e.g., Andronova and Schlesinger, 2001; Knutti et al., 2002; Forest et al., 2002). By investigating the uncertainty about three parameters simultaneously, we have not spared ourselves the additional complexity of working on a multi-dimensional uncertainty space. This is a very important point, since most integrated assessments of climate change will exhibit several uncertain key quantities. Therefore, an application of imprecise probability theory that works nicely in one dimension, but can hardly be extended to higher dimensions, would have demonstrated little. In this analysis, we have suffered from the 'curse of dimensionality' in two ways. First, the number of focal elements increased from an average of 20 on the marginal spaces to more than 6000 on the joint uncertainty space, to more than 26000 after updating with the likelihood information, to approx. 10000 for the projection on global mean temperature change. Second, the Dempster-updated focal elements did not inherit the simple cubic structure of their prior predecessors. Despite these complications, we have shown that the imprecise probability analysis remained feasible. We could have even processed a larger amount of focal elements if we had invested more in the improvement of computational efficiency.

The space of model hypotheses about the climate response was scanned for a given scenario about the radiative forcing trajectory in the industrial era (excluding sulphate which was part of the hypotheses). We have compared the resulting set of hypotheses about the global mean sea surface temperature (SST) and land surface air temperature (LAT) change in the 20th century with the instrumental temperature record in order to generate a likelihood function for the model parameters. The likelihood function was derived, inter alia, from the conditional probability that the residual between the simulated SST response and the stochastic SST data constitutes an AR(1) process after removal of the ENSO signal. The inclusion of the ENSO signal in the likelihood definition was crucial, since it could be easily identified as a significant source of deviations from a residual AR(1) process that emerges from weather noise. Compared to ENSO, two other modes of interannual climate variability, the NAO and the PDO, had a much smaller impact on the residual, so that a removal of these signals was not justified. Moreover, a statistical Portmanteau test could not reject the AR(1)-assumption for the ENSO adjusted residual, and higher order AR-processes (up to fourth order) did also not provide a significantly better explanation. We conclude that the AR(1) assumption for the SST part of the likelihood function is a reasonable choice. A similar conclusion can be drawn for the remaining part involving the difference between LAT and SST response (see Section 2.4). We note as potential caveats that statistical tests are not very powerful for short time series, and that we have not considered all potentially influential modes of interannual climate variability, nor did we check for long-range correlations in the residual.

In the process of constructing the likelihood function, we have capitalised on the simple structure of the model DOECLIM by directly implementing it in an optimisation framework. Obviously, this is an infeasible task for complex general circulation models. As a consequence, we did not need to fit the model response in terms of a linear superposition of response patterns to the data, but could subject the full model to the fitting procedure. The trade-off is, of course, that we can only estimate the likelihood from global mean temperature data. The implementation of DOECLIM in an optimisation framework allowed us to treat the remaining unknowns in our assumption about the radiative forcing scenario and the temperature residual (amplitude of ENSO signal, propagator of AR(1) process, strength of volcanic forcing) as nuisance parameters that have to be chosen in an optimal way to provide the best fit to the data at each point $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$ of the set of hypotheses. This particular treatment of the nuisance parameters gives less weight to the point of maximum likelihood (see Section 2.5). The resulting likelihood function for the parameters $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$ reveals a strong positive correlation between climate sensitivity and effective vertical heat diffusivity, and between climate sensitivity and strength of the sulphate cooling effect. Due to this correlation, high values of climate sensitivity cannot be excluded, as long as the upper range of values for sulphate cooling and/or ocean heat diffusivity is not further constrained. This result has been identified by many other studies before (e.g., Forest et al., 2002; Knutti et al., 2002). If we apply a series of classical statistical tests to the residuals between model response and instrumental temperature record on the parameter space $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$, we can only exclude values of climate sensitivity below 1.1 K and sulphate aerosol coolings that are stronger than $1.75\,\mathrm{W\,m^{-2}}$ at the 5% significance level.

Along the lines of the robust Bayesian approach, we have used the likelihood function to update a set of prior probabilities for the climate model parameters including a large class of prior assumptions. The prior uncertainty assessment for $T_{2x}$ and $Q_{\mathrm{S90}}$ was based on a set of probability estimates in the literature and allowed for very low ($T_{2x} \approx 0.5\,\mathrm{K}$) as well as very high values ($T_{2x} \approx 10\,\mathrm{K}$) of climate sensitivity. It also included the extreme case of a vanishing sulphate cooling ($Q_{\mathrm{S90}} \approx 0\,\mathrm{W\,m^{-2}}$) due to a neutralising effect of black carbon. The imprecise prior for ocean heat diffusivity $\kappa_v$ was constrained directly by observations of ocean heat uptake in the second half of the 20th century and strongly discarded large values of $\kappa_v$. As a caveat, the relationship between heat uptake and heat diffusivity was established by neglecting the uncertainty in the instrumental SST record, albeit it can be sensitive to it depending on the (unknown) serial correlation of the measurement errors. As a consequence of discarding large $\kappa_v$, the Dempster-updated imprecise probability allocated much smaller weight to high values of climate sensitivity (upper 95% quantile: $T_{2x} \approx 6.9\,\mathrm{K}$), but did not fully exclude values as high as $T_{2x} = 10\,\mathrm{K}$ (upper probability $\overline{P}_{T_{2x}}(10\,\mathrm{K}) = 0.014$). This result agrees qualitatively with the finding of Forest et al. (2002). In contrast, the imprecise posterior probability excluded low values of climate sensitivity $T_{2x} < 1.5\,\mathrm{K}$ with almost certainty, offering strong support for the lower bound of the IPCC estimate of $T_{2x} \in [1.5\,\mathrm{K}, 4.5\,\mathrm{K}]$. Overall, the lower probability of the IPCC estimate increased from 0.31 in the prior assessment to 0.53 after inclusion of the likelihood information. Moreover, the radiative forcing from sulphate aerosols in the year 1990 could be constrained to $Q_{\mathrm{S90}} \in [-1.53\,\mathrm{W\,m^{-2}}, -0.33\,\mathrm{W\,m^{-2}}]$ with 99% lower confidence, giving little support to the idea of a complete compensation of the sulphate cooling by black carbon.

We have projected the imprecise posterior probability for the climate model parameters onto estimates of future global mean temperature change under several emissions scenarios for the 21st century and several long-term stabilization policies. The uncertainty in the estimates for the end of the century and beyond was dominated by the uncertainty about climate sensitivity. Due to the admission of high values of $T_{2x}$, the upper end of the warming estimates for the late 21st century is more than 30% higher than the already high upper estimate of the IPCC. This points to the strain that a large climate sensitivity would inflict on humankind in the late 21st century and beyond. Moreover, in this case stabilisation policies as low as 450 ppm $CO_2$ equivalent concentration in the atmosphere will generally fail to limit the warming to 2 or even 3 Kelvin. Therefore, it is of utmost importance for the discussion about long-term climate policies to better constrain climate sensitivity from above than this could be achieved in the present analysis and in most preceding analyses of the uncertainty about climate sensitivity (for a noteworthy exception see Schneider von Deimling et al. 2005). We can identify tighter limits to the low end of warming estimates. It is very unlikely that the warming in the 21st century will remain below 2 Kelvin in the absence of policy intervention. Moreover, we find that it will require a very stringent stabilisation level of around 450 ppm $CO_2$ equivalent in the atmosphere to obtain a non-negligible value for the lower probability of limiting the warming to 2 Kelvin. This highlights the serious challenge for climate policy.

In the light of the preceding paragraphs we claim to have made indeed a serious effort to capture and constrain the uncertainty about the climate response to secular trends in the radiative forcing. What then would prevent us from regarding our results as robust statements about the plausible range of climate sensitivity, sulphate cooling effect and future global mean temperature change? We can identify several critical points in our analysis that offer room for improvement. First and foremost, we have specified a scenario for the radiative forcing in the industrial era, which only allowed uncertainty about the contribution of sulphate and volcanic aerosols. However, the overall uncertainty about the historical forcing trajectory is much larger due to the poorly understood contribution of black carbon, biomass burning aerosols, solar activity and, to a lesser extent, tropospheric and stratospheric ozone (Ramaswamy, 2001). Although the sum of these forcings is likely to be small compared to the contribution of well-mixed greenhouse gases and sulphate aerosols, they nevertheless can influence the shape of the likelihood function considerably. This is due to the unfortunate fact that the assumptions about these forcings determine to a large extent whether the forcing trajectory in the 20th century reflects the marked shaped of the 20th century warming with a switch between warming and cooling periods around 1950 and 1980, or whether it does not. Therefore, it will be important to include the uncertainty about these forcings into the likelihood estimation in a next step. Second, the likelihood function is also sensitive to the assumptions about the residual between model simulation and data. Although the AR(1) assumption has passed several tests after removal of the ENSO signal, we have by no means compared it against an exhaustive list of alternative hypotheses. Third, we have allowed for some uncertainty in our assumptions by introducing nuisance parameters to the estimation of the likelihood function. In principle, this could be done for any of the neglected uncertainties mentioned above. However, we have resolved the uncertainty about these parameters by adjusting them at each point $\theta = (T_{2x}, \kappa_v, Q_{\mathrm{S90}})$ in a way that achieved the best fit to the data. While this is better than simply fixing them, it would be much better to allow for a set of likelihood functions $\mathcal{L}_{\theta_N}(\theta; \hat{y}) : \Omega \to \mathbb{R}^+$, one for each constellation of nuisance parameters $\theta_N$. In this case, we would need to update an imprecise prior probability with an *imprecise likelihood function*. We believe that the methodological framework presented in Chapter 4 can be extended in this direction. However, it will require utmost care to limit the imprecision in the likelihood information to a degree that allows the derivation of meaningful posterior probability bounds.

## 6.3 Conclusion

In summarising the entire scope of this analysis, we want to point out that we have attached great importance to providing an imprecise probability analysis of climate change that covers all pertinent aspects of an integrated assessment. We believe this to be an important prerequisite for demonstrating that imprecise probability theory provides indeed a viable uncertainty framework for an integrated assessment of climate change. Any such framework must be able to cope with all steps of the assessment in order to produce meaningful results that can be fed into the climate policy arena. Therefore, we have gone through the entire assessment chain from generating the prior uncertainty representation for the climate model parameters to updating it with likelihood information, to projecting the posterior uncertainty onto estimates of future climate change, to finally conducting a climate policy analysis under imprecise probability.

Moreover, we think that this analysis has not only demonstrated the applicability of imprecise probabilities to the entire assessment chain, but also has revealed their strength from treating uncertainty in more general terms than classical probability. We want to point out the following aspects of this strength that played out in the particular application presented here.

- We have utilised the ability of imprecise probability concepts to capture poor states of information. We were able to construct prior uncertainty models for the climate model parameters that contained a large class of plausible prior probabilities. In doing so, it became evident that the construction of imprecise probability models resembles a constraint-based approach where probability bounds on a limited set of events are chosen, and only those probabilities are excluded from the analysis which violate these bounds.

- By the same token, we could construct necessary approximations for, e.g., converting a continuous uncertainty model into a finite one by simply giving up information. Therefore, the approximation always fully contained the approximated set of probabilities. This assures that our results are robust in the sense that anything not supported by the approximation would also not have been supported by the original information. On the other hand, the approximation might support a host of statements that would not have been supported by the original information. Hence, if we give up too much information in the approximation process, we will end up with meaningless results close to the state of complete ignorance. We want to emphasise that we consider this a virtue rather than a deficiency of the imprecise probability framework. It prevents the analyst from gaining a false sense of precision and points her into the direction where the informational basis needs improvement.

- We have also drawn on the flexibility of imprecise probability theory to incorporate very different types of evidence. For the construction of the prior belief function, we have combined bounds on probability distribution functions with a lower bound on probability mass functions. Most importantly, we have pointed out at several instances that any classical probability can be directly included in our methodological framework by means of its mass functions. As a consequence, the updating framework presented in Chapter 4 is directly applicable to calculate lower and upper posterior bounds for a classical probability. It is also noteworthy that we could have included interval uncertainty with similar ease into our analysis. We have not done so because the introduction of interval uncertainty induces much imprecision, which quickly reduces the analysis to a problem of interval propagation.

- Finally, we have demonstrated that a larger uncertainty as captured by imprecise probabilities does not necessarily need to impede a useful decision support. In contrast, it can help to make the policy advice more robust since it emerged from a much larger set of eventualities under consideration. Moreover, we have highlighted the rich structure of decision criteria under imprecise probability which are much more likely to capture (at least parts of) the complexity of climate policy making than the conventional expected utility frameworks.

We conclude that our analysis has demonstrated imprecise probability theory to be indeed a promising framework for processing uncertainty in the integrated assessment of climate change. It goes without saying that a final verdict on the usefulness of imprecise probabilities in this field will require more experience with applications and more methodological work in response to it than we could produce here. As already pointed out above, improvements are particularly needed for discretising continuous uncertainty spaces, for identifying reasonable and efficient updating procedures, for tackling the problem of random set extension even with complex models and for including imprecise likelihood information in the analysis. We also need investigations of the usefulness of other imprecise probability models that do not dwell on sparse Möbius representations. This includes, but is not limited to, probability ratio models, density bands and ALUP models. Definitely, there is a strong need to generate prior imprecise probability models about climate change directly from expert elicitations or constraint-based reasoning with evidence. What is ultimately needed are applications to real-world climate policy analysis to assess how it is impacted by the presence of imprecision which might lead to more robust choices in some instances and might diffuse taken for granted choices in other instances. Hence, much more work is required until imprecise probabilities can be used as a standard tool for an improved treatment of uncertainty in the integrated assessment of climate change. With this analysis, we have started to build the bridge between imprecise probability theory and climate (change) research. We hope this work to be appealing enough that researchers from either side will feel invited to join the construction. Whether they will, is ultimately in their and the future's hands. We can only await the resolution of yet another uncertainty.

# Appendix A

# Construction of a diffusion ocean energy balance model

In this appendix, we will describe in detail the construction of the Diffusion Ocean Energy balance CLIMate model (DOECLIM) that is used throughout the analysis. Section A.1 provides a deeper discussion of the assumptions underlying the linearised treatment of the climate response to perturbations in the Earth's energy budget than could be provided in Section 2.1.1. Section A.2 motivates the dynamic core of the model. Section A.3 outlines the treatment of the interior ocean which dominates the transient model response. In Section A.4, the set of free model parameters is narrowed to climate sensitivity $T_{2x}$ and effective vertical diffusivity $\kappa_v$ of heat in the ocean. Finally, Section A.5 describes the numerical integration of DOECLIM.

## A.1 Separation of radiative forcing and temperature feedback

The earth surface receives almost all of its energy from the sun.[1] The incoming solar insolation exhibits the spectral density of a black body with a temperature $T \approx 6000\,\mathrm{K}$, and thus contains radiation centered around the spectral range $\lambda = 0.4\,\mu\mathrm{m} - 0.7\,\mu\mathrm{m}$ of visible light. The total amount of solar radiation at the average distance sun-earth is specified by the solar "constant" $S_0$. It depends on the 23-kyr precession of the earth's axis, its obliquity (tilt with respect to the orbital plane), changing with 41-kyr periodicity, and the eccentricity of the earth orbit, exhibiting 100-kyr, 400-kyr and 2-Myr periodicities (Milankovitch, 1941). In addition, $S_0$ depends on the solar activity which fluctuates with a period of approximately 11 years. Currently, the solar constant varies between $S_0 = 1366\,\mathrm{W\,m^{-2}}$ and $S_0 = 1367\,\mathrm{W\,m^{-2}}$ (Fröhlich, 2000). The earth receives a solar insolation of $S_0/4$, where one factor of 1/2 stems from the fact that only half of the earth surface receives daylight. The other factor of 1/2 accounts for the

---

[1] The contribution of geothermal processes is estimated to be less than 0.03% of the solar radiation absorbed by the earth, and therefore can be neglected (Peixoto and Oort, 1992, Chap. 6.8.1).

fact that the earth surface is approximately a sphere. About $\alpha \approx 0.3$ of the incoming solar insolation is reflected back to space, and the remaining fraction is either absorbed by the surface ($a_S \approx 0.5$) or by the atmosphere ($a_A \approx 0.2$) (Peixoto and Oort, 1992, Chap. 6.1). $\alpha$ is called the planetary albedo.

In contrast to the short-wave radiation $\lambda < 4\,\mu$m emitted by the sun, the radiation from the earth surface, with temperature $T_S \approx 287\,$K, lies in the infrared spectral band $\lambda = 4 - 60\,\mu$m. Due to the negligible overlap of the spectral bands, the atmosphere exhibits very different absorption properties for the long-wave radiation from the surface and for the short-wave radiation from the sun. Only a small fraction $t_A \approx 0.06$ of the surface radiation can escape to space, while the largest portion is absorbed by water vapour and GHGs like carbon dioxide ($CO_2$), methane ($CH_4$), and nitrous oxide ($N_2O$). In addition, energy is exchanged by physical transport of latent heat (evaporation) and sensible heat (convection of air parcels) from the surface to the atmosphere. This physical heat transport, whose energy content amounts to approximately 30% of the solar insolation, induces a strong coupling between surface and troposphere. Thus, a perturbation of the radiation balance at the surface-troposphere interface has only a small influence on surface temperature compared to a perturbation at the top of the atmosphere (Harvey, 2000, pp. 41–43).

Given the situation described above, the energy balance at the surface and the top of the atmosphere (TOA) can be cast in the following two equations (compare to Equation 2.2),

$$\text{Surface} \qquad \sigma T_S^4 + L = \sigma \left(T_{A,\text{eff}} + \delta T_A\right)^4 + a_S \frac{S_0}{4} \ , \tag{A.1}$$

$$\text{TOA} \qquad t_A \sigma T_S^4 + \sigma T_{A,\text{eff}}^4 = (1-\alpha)\,\frac{S_0}{4} \ , \tag{A.2}$$

with $T_S$ Earth surface temperature ,

$T_{A,\text{eff}}$ Effective radiating temperature of the atmosphere to space ,

$T_{A,\text{eff}} + \delta T_A$ Effective rad. temperature of the atmosphere to the surface ,

$L$ latent and sensible heat fluxes .

Since the actual temperature in the atmosphere decreases with height, the effective radiating temperature $T_{A,\text{eff}}$ for upward radiation is lower than the effective radiating temperature $T_{A,\text{eff}} + \delta T_A$ as seen from the surface. $T_{A,\text{eff}} + \delta T_A$ incorporates the effect of warmer lower layers absorbing and re-emitting radiation from colder upper layers. We define $R = (T_{A,\text{eff}} + \delta T_A)/T_{A,\text{eff}}$ as the ratio of the two effective temperatures.

The sensible and latent heat fluxes can be parameterised by $L = c_L\,(T_S - T_{A,\text{eff}})$, with $c_L \approx 2.5\,\text{W}\,\text{m}^{-2}\text{K}^{-1}$ a constant. Under this assumption, Equations (A.1) and (A.2) can be solved numerically for surface temperature $T_S$ and atmospheric effective radiating temperature $T_{A,\text{eff}}$ given $R$, $t_A$, $a_S$, $\alpha$ and $S_0$. With $R \approx 1.1$, $S_0 = 1366\,\text{W}\,\text{m}^{-2}$ and the values of the other quantities as specified above, we find $T_S \approx 287\,$K and $T_{A,\text{eff}} \approx 248\,$K. This amounts to a surface radiation $F_S \uparrow := \sigma T_S^4 \approx 390\,\text{W}\,\text{m}^{-2}$. As pointed out in Section 2.1.1, $F_S \uparrow$ is considerably larger than the spacebound radiation at the TOA, which equals the absorbed incoming solar radiation. The greenhouse effect $G = F_S \uparrow - (1-\alpha)\, S_0/4$ specifies the amount of energy (per second and square meter) that contributes to the warming of the earth surface due to absorption of infrared radiation in the atmosphere. As can be seen from Equations (A.1) and (A.2), $G = G(t_A, R, a_S, \alpha, S_0)$ is a function of the incoming solar radiation absorbed by the surface and the earth system as a whole, of the atmospheric transmissivity $t_A$ for infrared radiation, and of the ratio $R$ of effective temperature for downward and upward radiation, which depends on the temperature profile and the distribution of clouds, water vapour and GHGs in the atmosphere.

The sensitivity of the energy balance to small deviations of the radiation parameters $t_A$, $R$, $a_S$, $\alpha$ can be assessed by a Taylor approximation. Let $E = G + (1-\alpha)\,\frac{S_0}{4}$ denote the total energy received by the surface (per second and square meter). Then, a perturbation $\Delta E$ of the incident energy at the surface is approximated by

$$\Delta E \approx \left.\frac{\partial E}{\partial R}\right|_{\mathrm{eq}} \Delta R + \left.\frac{\partial E}{\partial t_A}\right|_{\mathrm{eq}} \Delta t_A + \left.\frac{\partial E}{\partial \alpha}\right|_{\mathrm{eq}} \Delta \alpha + \left.\frac{\partial E}{\partial a_S}\right|_{\mathrm{eq}} \Delta a_S + \left.\frac{\partial E}{\partial S_0}\right|_{\mathrm{eq}} \Delta S_0\,. \tag{A.3}$$

The resulting perturbation of the radiative equilibrium induces a heat flux $\dot{H}(t)$ at the surface that can be described in a good approximation by (see Section 2.1.1)

$$\dot{H}(t) \quad \approx \quad \Delta E(t) \quad - \quad 4\,\sigma\, T_{S,\mathrm{eq}}^3\, \Delta T_S(t)\,. \tag{A.4}$$

We can simplify Equation (A.4) by approximating $\Delta E(t)$ with Equation (A.3). The changes $\Delta R$, $\Delta t_A$, $\Delta\alpha$, and $\Delta a_S$ depend on surface temperature $T_S$ and various other climate system properties $y$ (not necessarily independent of $T_S$), some of which might be altered due to human and natural interferences. Obviously, the solar constant $S_0$ is independent of any earth system properties. In a second approximation, we assume that the quantities $x \in \{\, R,\, t_A,\, \alpha,\, a_S \,\}$ vary linearly around their equilibrium values,

$$\begin{aligned} \dot{x}(y, T_s) \quad &= \quad \frac{\partial x}{\partial y}(y, T_S)\, \dot{y}(t) + \frac{\partial x}{\partial T_S}(y, T_S)\, \dot{T}_S(t) \\ &\approx \quad \frac{\partial x}{\partial y}(y_{\mathrm{eq}}, T_{S,\mathrm{eq}})\, \dot{y}(t) + \frac{\partial x}{\partial T_S}(y_{\mathrm{eq}}, T_{S,\mathrm{eq}})\, \dot{T}_S(t)\,. \end{aligned} \tag{A.5}$$

Approximation (A.5) is valid, if the sensitivity of $x$ to small interferences $\frac{\partial x}{\partial y}$ as well as the magnitude of the instantaneous temperature feedback $\frac{\partial x}{\partial T_S}$ neither saturates nor is exacerbated significantly with increasing (or decreasing) surface temperature in the range of $T_S$, $y$ and $x$ under consideration. These assumptions need close scrutiny. It can be expected, for example, that the sensitivity of surface albedo to a change in surface temperature differs considerably for different extents of continental ice sheets. Thus, the validity of Approximation (A.5) will be limited to small temperature changes $\Delta T_S$, and time scales of a few centuries. Keeping this in mind, the heat flux into (or out of) the earth system can be approximated by inserting Equations (A.3) and (A.5) into Equation (A.4). We have

$$
\begin{aligned}
\dot{H}(t) \;=\;& \underbrace{\left.\frac{\partial E}{\partial R}\right|_{\mathrm{eq}} \Delta R(t, T_{S,\mathrm{eq}}) + \left.\frac{\partial E}{\partial t_A}\right|_{\mathrm{eq}} \Delta t_A(t, T_{S,\mathrm{eq}})}_{Q_{\mathrm{Atm}}} \;+ \\
& \underbrace{\left.\frac{\partial E}{\partial \alpha}\right|_{\mathrm{eq}} \Delta \alpha(t, T_{S,\mathrm{eq}}) + \left.\frac{\partial E}{\partial a_S}\right|_{\mathrm{eq}} \Delta a_S(t, T_{S,\mathrm{eq}})}_{Q_{\mathrm{Alb}}} \;+\; \underbrace{\left.\frac{\partial E}{\partial S_0}\right|_{\mathrm{eq}} \Delta S_0(t)}_{Q_{\mathrm{Sol}}} \;- \\
& \underbrace{\left(4\,\sigma\,T_{S,\mathrm{eq}}^3 - \left.\frac{\partial E}{\partial t_A}\,\frac{\partial t_A}{\partial T_S}\right|_{\mathrm{eq}} - \left.\frac{\partial E}{\partial R}\,\frac{\partial R}{\partial T_S}\right|_{\mathrm{eq}} - \left.\frac{\partial E}{\partial a_S}\,\frac{\partial a_S}{\partial T_S}\right|_{\mathrm{eq}} - \left.\frac{\partial E}{\partial \alpha}\,\frac{\partial \alpha}{\partial T_S}\right|_{\mathrm{eq}}\right)}_{\lambda} \Delta T_S(t) \\
\;=\;& Q_{\mathrm{Atm}}(t) + Q_{\mathrm{Alb}}(t) + Q_{\mathrm{Sol}}(t) - \lambda\, \Delta T_S(t)\;, \qquad \text{(A.6)}
\end{aligned}
$$

with $Q_{\mathrm{Atm}}$ radiative forcing due to changes in atmospheric parameters,
$Q_{\mathrm{Alb}}$ changes in albedo parameters, $Q_{Sol}$ changes in solar activity,
$\lambda$ climate feedback strength .

Hence, the approximations made so far have allowed us to recover the linear energy balance model (2.3) postulated in Section 2.1.1. As a direct consequence of separating $\Delta E(t)$ into a radiative forcing and a linear *temperature feedback* on the radiation balance, the equilibrium temperature response to a radiative perturbation is predicted to be proportional to the amount of radiative forcing independently of the forcing mechanism. We will briefly discuss in how far this strong assumption is justified.

The radiative perturbation at the tropopause before any temperatures are allowed to adjust is called the *instantaneous radiative forcing.* Since the stratosphere and troposphere are only weakly coupled, while the surface and troposphere are tightly coupled due to the presence of large physical heat fluxes, the stratospheric temperatures will adjust independently and much faster (within months) than the troposphere-surface temperatures. This alters in turn the net radiation at the tropopause, which will equal, in stratospheric equilibrium, the net radiation at the TOA. The remaining radiative perturbation after stratospheric adjustments, but before any surface-troposphere temperatures are allowed to adjust, is called the *adjusted radiative forcing* (Harvey, 2000, pp. 42). Studies with atmospheric general circulation models (AGCMs) have confirmed that the equilibrium surface temperature response was indeed proportional to the adjusted radiative forcing (e.g. Marshall et al., 1994). It is also rather independent, to within 20%, of the type of forcing mechanism whether constituted by solar variations, well mixed GHGs or scattering aerosols (mostly sulphate) (Hansen et al., 1997; Forster et al., 2000; Rotstayn and Penner, 2001; Stuber et al., 2001; Joshi et al., 2003).

However, these studies also indicated that the temperature response to the adjusted radiative forcing of tropospheric and stratospheric ozone and absorbing aerosols (mostly soot) was markedly different. For some species of absorbing aerosols, the adjusted radiative forcing does not even predict the sign of the consequent temperature change (Cook and Highwood, 2004). The markedly different response is due to the strong vertical variation of ozone and aerosol concentrations in the atmospheric column. This leads to an inhomogeneous heating of the atmospheric layers due to absorption of infrared radiation by ozone and absorption of solar radiation by soot. Since cloud cover tends to decrease in the atmospheric layers having the greatest direct heating, ozone and absorbing aerosols induce a direct cloud feedback on the ratio of effective radiating temperatures $R$. In the case of absorbing aerosols, the direct cloud feedback counteracts the net negative radiative forcing after stratospheric adjustments.

To address the issue, Shine et al. (2003) recently proposed the concept of *adjusted troposphere and stratosphere forcing (ATS forcing)*, which specifies the radiative perturbation at the TOA after stratosphere *and* troposphere were allowed to adjust for fixed surface temperature. They showed with an intermediate AGCM that the ATS forcing differs considerably from the adjusted (stratosphere only) forcing for ozone and absorbing aerosols, and that the equilibrium temperature response was indeed proportional to the ATS forcing induced by these substances. Moreover, the relative strength of the temperature feedback to the ATS forcing, captured by the parameter $\lambda$ in Equation (A.6), depends much less on the forcing mechanism (within 10%) than it is the case with the adjusted (stratosphere only) forcing. Thus, the ATS forcing seems to be the proper concept to use in simple energy balance models based on Equation (A.6). Unfortunately, the proposal of Shine et al. (2003) is too recent to be included in this analysis, since there exists no other estimates for the ATS forcing of GHGs and aerosol species, and the corresponding feedback strength to date. Therefore, we will continue the common practice of using the adjusted (stratosphere only) forcing.

Given the complexity of the climate feedback mechanisms discussed in Section 2.1.1, it is not a trivial result that the equilibrium surface temperature change depends fairly linear on the adjusted radiative forcing in the range of the anthropogenic perturbations considered here ($Q < 10\,\mathrm{W\,m^{-2}}$). Another question concerns the *transient behaviour* of the climate feedback strength $\lambda$. Since the transient temperature response is dominated by the ocean, an investigation of the time-dependence of $\lambda$ requires experiments with atmosphere-ocean general circulation models (AOGCMs). The feedback strength at any given point in time, for a fixed radiative forcing Q, can be calculated by (cf. approximation A.6)

$$\lambda(t) = \frac{Q - \dot{H}(t)}{\Delta T_S(t)} \ . \tag{A.7}$$

$\lambda(t)$ is generally expressed in terms of *effective climate sensitivity* $T_{2x,\mathrm{eff}}(t) := Q_{2x}/\lambda(t)$, which constitutes the climate sensitivity due to a doubling of $CO_2$ that would occur at equilibrium, if the feedback strength was held fixed at its value at time $t$ (Murphy, 1995). Several experiments with different versions of the Hadley Centre Coupled Model (HadCM2 and HadCM3) have shown an increase of effective climate sensitivity with time (Murphy, 1995; Senior and Mitchell, 2000; Gregory et al., 2004). This time-dependence was associated with differences in cloud feedback arising from inter-hemispheric temperature differences due to the slower warming rate of the southern ocean. However, other AOGCMs show little dependence of effective climate sensitivity on time. Watterson (2000), for instance, found with the Commonwealth Scientific and Industrial Research Organisation (CSIRO) coupled model that $T_{2x,\mathrm{eff}}$ does not vary significantly during the experiments. Moreover, a suite of transient runs with various AOGCMs have been successfully emulated by an 1-D upwelling-diffusion energy balance model (Cubasch and Meehl, 2001) and a 2-D model of intermediate complexity (Sokolov and Stone, 1998; Sokolov et al., 2003), which both assumed constant climate sensitivity. Thus, besides its limitations, Approximation (A.6) provides a useful basis to separate an anthropogenic (or natural) perturbation of the earth's energy balance, captured in terms of radiative forcing, from an assumed linear surface temperature response. This allows to work with the concept of an overall climate sensitivity to radiative perturbations in a meaningful manner.

## A.2 Atmosphere-ocean-land model

In its initial form, the energy balance model (EBM) of the atmosphere-ocean-land system is constituted by four stylized boxes: land $L$, troposphere over land $AL$, troposphere over the sea $AS$, and ocean mixed layer $S$. The model does not include the stratosphere, since it adjusts to a radiative perturbation within months (Harvey, 2000). Hence, we will assume an equilibrated stratosphere throughout the analysis, so that the radiative perturbation at the tropopause equals the radiative perturbation at the TOA. The radiative heating is distributed among the boxes, before diffusing into the interior ocean. Such a 4-box energy balance model is described by (cf. Murphy, 1995):

$$\begin{array}{rl}
\text{Troposphere (land)} & C_A\,\dot{T}_{AL} = \tilde{Q}_{AL} - \tilde{\lambda}_{AL}\,T_{AL} - \dfrac{\tilde{k}}{f_L}\,(T_{AL} - T_{AS}) - \tilde{k}_L\,(T_{AL} - T_L) \\
\text{Land} & C_L\,\dot{T}_L = \tilde{Q}_L - \tilde{\lambda}_L\,T_L - \tilde{k}_L\,(T_L - T_{AL}) \\
\text{Troposphere (sea)} & C_A\,\dot{T}_{AS} = \tilde{Q}_{AS} - \tilde{\lambda}_{AS}\,T_{AS} - \dfrac{\tilde{k}}{1 - f_L}\,(T_{AS} - T_{AL}) - \tilde{k}_S\,(T_{AS} - T_S) \\
\text{Ocean mixed layer} & c_V\,z_S\,\dot{T}_S = \tilde{Q}_S - \tilde{\lambda}_S\,T_S - \tilde{k}_S\,(T_S - T_{AS}) - F_O\,,
\end{array}$$

| | |
|---|---|
| with $\tilde{\lambda}_{AL,L,AS,S}$ | climate feedback parameter in the respective boxes (in $[\mathrm{W\,m^{-2}K^{-1}}]$), |
| $\tilde{k}$ | atmospheric land-sea heat exchange coefficient (in $[\mathrm{W\,m^{-2}K^{-1}}]$), |
| $\tilde{k}_L$, $\tilde{k}_S$ | atmosphere-land/sea heat exchange coefficients (in $[\mathrm{W\,m^{-2}K^{-1}}]$), |
| $F_O$ | heat flux into the interior ocean (in $[\mathrm{W\,m^{-2}}]$), |
| $C_A, C_L$ | heat capacity of atmosphere, land (in $[\mathrm{J\,m^{-2}K^{-1}}]$), |
| $c_V$ | heat capacity of cubic metre of seawater (in $[\mathrm{J\,m^{-3}K^{-1}}]$), |
| $z_S$ | depth of ocean mixed layer (in [m]), |
| $f_L$ | land fraction of earth surface area. |

$T_{AL}$, $T_L$, $T_{AS}$, $T_S$ are the effective temperature anomalies in the respective boxes relative to their equilibrium temperatures before the radiative perturbation is applied. They have replaced the single state variable $T_S$ in Equation (2.3). Likewise, the globally aggregated radiative perturbation $Q$ at the TOA, after stratospheric adjustments, is replaced by radiative forcings $\tilde{Q}_{AL}$ and $\tilde{Q}_{AS}$ of the troposphere over land and sea, respectively, and radiative forcings $\tilde{Q}_L$, $\tilde{Q}_S$ at the troposphere-land/sea interface. Hence, the differential equations are based on the linear ansatz (2.3), and include in addition the anomalous heat transfer to the neighbouring boxes (no direct heat transfer is assumed between land and sea box). In an often used approximation, the heat flux anomaly is assumed to be proportional to the temperature anomaly gradient between the boxes. Since we express the heat flux per unit area, we need to weigh the heat transfer coefficient between troposphere over land and sea by the land fraction of the earth surface.

The 4-Box model exhibits a variety of parameters, of which in particular the climate feedback parameters of the individual boxes will be difficult to estimate. Moreover, the partitioning of the radiative forcing onto the four boxes will depend on the forcing agent. Hence, the separation of troposphere and surface introduces a detail, that complicates the assessment of radiative forcing and climate feedback strength. However, there exists a strong coupling between surface and troposphere due to large physical heat fluxes of latent and sensible heat (see Equation A.1). A perturbation of the radiation balance at the surface-troposphere interface equilibrates much faster due to readjustments of the surface-troposphere temperature gradient than a perturbation at the TOA due to a change of the effective radiating temperature of the entire earth system (Harvey, 2000, Chap. 3). Hence, for the sake of modelling the secular climate response to a radiative perturbation, we can assume that the tightly coupled surface and troposphere temperatures increase in proportion to each other ($\dot{T}_{AL} \approx a_{\Gamma,L}\dot{T}_L$, $\dot{T}_{AS} \approx a_{\Gamma,S}\dot{T}_S$). It is generally expected that the troposphere warms faster than the surface ($a_{\Gamma,L} > 1$, $a_{\Gamma,S} > 1$). This follows from the fact that the lapse rate is smaller in moist air than in dry air. If the atmosphere warms, its moisture content will increase on the global average and, therefore, the lapse rate decreases. The enhancement of effective troposphere warming relative to surface warming has been investigated in several AOGCM experiments, suggesting a globally averaged value of $a_\Gamma = 1.2$ (Hansen et al., 2002). Satellite-borne measurements have seemingly indicated the opposite behavior with surface warming outpacing troposphere warming (Mears et al., 2003). Recently, Fu et al. (2004) proposed an explanation to reconcile model projections with observations by correcting the satellite data for the stratosphere cooling.

Since a large portion of the ocean covers the humid low latitudes while much of the land mass is situated at the drier high latitudes, the enhancement of troposphere warming will be more pronounced over the ocean ($a_{\Gamma,S} > a_{\Gamma,L}$). The sea ice feedback provides an additional, strong mechanism that warms the marine atmosphere faster than the sea surface. In areas with sea ice, the marine air is in contact with the cold ice, and therefore much colder than the underlying water temperature. If the sea ice cover retreats in a warming climate, the marine atmosphere over a significant portion of the high latitudes will warm rapidly as the cold ice is replaced by the warmer water surface. On the global average, this can lead to a 20% stronger increase in marine air temperature than in sea surface temperature (Raper and Cubasch, 1996).

By drawing on the proportionality between surface and troposphere warming we can attempt to simplify the 4-box model. EBMs that lump together atmosphere and surface boxes have been constructed before for the assessment of anthropogenic climate change (e.g., Rowntree, 1998). For our modelling purpose, we express the increase in effective atmospheric temperatures $T_{AL}$ and $T_{AS}$ in terms of the corresponding increase in near surface air temperature (SAT), i.e. $\dot{T}_{AL} = a_{\Gamma,L}\,\dot{T}_L^{SAT}$, and $\dot{T}_{AS} = a_{\Gamma,S}\,\dot{T}_S^{SAT}$. Over land, the increase in SAT is taken to be equal to the increase in land surface temperature $T_L$. Over the sea, the increase in SAT being enhanced by a reduction in sea ice cover is taken proportional to the increase in ocean mixed layer temperature $T_S$, i.e. $\dot{T}_S^{SAT} = b_{SI}\,\dot{T}_S$. Under these assumptions, we can amalgamate the differential equations for the land and surface boxes, and the model simplifies to

$$\text{Land + Tropos.} \qquad (\,a_{\Gamma,L}\,C_A + C_L\,)\,\dot{T}_L = Q_L - \lambda_L\,T_L - \frac{k}{f_L}\,(T_L - b_{SI}\,\frac{a_{\Gamma,S}}{a_{\Gamma,L}}\,T_S) \tag{A.8}$$

$$\text{Sea + Tropos.} \qquad (\,a_{\Gamma,S}\,b_{SI}\,C_A + c_V z_S\,)\,\dot{T}_S = Q_S - \lambda_S\,T_S - \frac{k}{1-f_L}\,(b_{SI}\,\frac{a_{\Gamma,S}}{a_{\Gamma,L}}\,T_S - T_L) - F_O\,, \tag{A.9}$$

where $Q_L = \tilde{Q}_{AL} + \tilde{Q}_L$, $Q_S = \tilde{Q}_{AS} + \tilde{Q}_S$ are the radiative forcings at the TOA after stratospheric adjustments, $\lambda_L = a_{\Gamma,L}\,\tilde{\lambda}_{AL} + \tilde{\lambda}_L$, $\lambda_S = a_{\Gamma,S}\,b_{SI}\,\tilde{\lambda}_{AS} + \tilde{\lambda}_S$ the climate feedback parameters of the surface-troposphere system over land (relative to land surface temperature) and sea (relative to SST), respectively, and $k = a_{\Gamma,L}\,\tilde{k}$ the effective land-sea heat exchange coefficient relative to the temperature gradient in the surface air. In the following we will neglect the fraction $a_{\Gamma,S}/a_{\Gamma,L}$, so that the dynamic core of our energy balance DOECLIM is represented by Equation (2.5) and (2.6) in Section 2.1.2.

## A.3 Interior ocean model

The transient behavior of the EBM is dominated by the heat uptake of the ocean. Therefore, we need a reasonable model to calculate the heat flux $F_O$ into the interior ocean. Most EBMs for the assessment of anthropogenic climate change utilize a 1-D upwelling-diffusion ocean model that describes the transfer of heat in the oceanic water column (Hoffert et al., 1980; Dickinson and Schaudt, 1998). The model is described by the so-called diffusion-convection equation of heat in fluid or gaseous media (see Appendix B),

$$\frac{\partial}{\partial t}T_O(z,t) = \frac{\partial}{\partial z}\left(\kappa_v(z)\,\frac{\partial}{\partial z}T_O(z,t)\right) + w\,\frac{\partial}{\partial z}T_O(z,t)\,, \tag{A.10}$$

where $z > 0$ denotes the depth of the interior ocean below the mixed layer, $T_O(z,t)$ the ocean temperature at depth $z$ and time $t$, $\kappa_v$ the *effective vertical diffusivity* of heat (in units of $\mathrm{m}^2\,\mathrm{s}^{-1}$) and $w$ the *upwelling velocity* in the water column, which transports heat from the bottom to the surface. In the context of climate modelling, Equation (A.10) is called upwelling-diffusion equation, and we will adopt this terminology in order to avoid confusion. It gained interest as a model of the globally averaged ocean column, because its equilibrium solution,

$$\frac{\partial}{\partial t} T_O(z,t) = 0 \quad \Rightarrow \quad T_{O,\mathrm{eq}}(z) = T_{S,\mathrm{eq}}\, e^{-\int_0^z w/\kappa(z')\,dz'} \ ,$$

provides a good explanation for the observed global mean temperature profile in the ocean for plausible values of upwelling velocity $w \approx 4\,\mathrm{m\,yr}^{-1}$ and uniform vertical diffusivity $\kappa_v(z) = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$ (Munk, 1966).

Consequently, the 1-D upwelling-diffusion model has been employed to model the penetration of a heat anomaly into the ocean due to rising surface temperatures (Hoffert et al., 1980; Wigley and Raper, 1992; Schlesinger et al., 1997; Harvey and Huang, 2001). Since the upwelling-diffusion equation is linear, the anomalous heat fluxes are governed by Equation (A.10) with initial equilibrium condition $T_O(z,0) = 0$. Then $T_O(z,t)$ represents the temperature anomaly at depth $z$ and time $t$ relative to its equilibrium value at depth $z$ and time $t = 0$. Typical upwelling-diffusion ocean models for the anomalous heat fluxes are discussed in detail in Appendix B. They usually include a heat source at the ocean floor which is motivated by downwelling polar water entering the main ocean column at the bottom. Here, we only mention that a "bucket" ocean model with depth $z_B = 4000\,\mathrm{m}$, and a uniform cross-section and vertical diffusivity of the water column yields a good approximation of a more realistic upwelling-diffusion model with depth-dependent cross section and diffusivity profiles. The heat accumulation that arises from the decrease of ocean cross-section with depth is counteracted to a large degree by the increase of vertical diffusivity with depth (see Fig. B.2). It is also discussed in Appendix B that upwelling-diffusion models have difficulties to describe the warming of the ocean column as it is seen in $CO_2$ doubling experiments with coupled atmosphere-ocean general circulation models (AOGCMs). In contrast, the temperature anomaly profiles seem to be well described by a pure diffusion ocean with vanishing upwelling velocity $w = 0\,\mathrm{m\,yr}^{-1}$ (Raper et al., 2001). Since our intention is to construct a model which is as simple as possible, and exhibits a minimum of uncertain parameters, we restrict ourselves to a pure diffusion ocean model.

Hence, we consider an ocean model defined by the following heat diffusion problem:

$$\text{for } 0 < z < z_B : \qquad \frac{\partial}{\partial t} T(z,t) = \kappa_v \frac{\partial^2}{\partial z^2} T_O(z,t) \ , \tag{A.11}$$

$$\text{B.C.:} \qquad T_O(0,t) = T_S(t) \ , \quad \frac{\partial}{\partial z} T_O(z_B,t) = 0 \ ,$$

$$\text{I.C.:} \qquad T_O(z,0) = 0 \ .$$

The boundary conditions ensure that the interior ocean temperature at $z = 0$, i.e., the boundary to the mixed layer, equals the mixed layer temperature $T_S$, and that the heat flux into the ocean floor at $z = z_B$ vanishes. As shown in Appendix B, Problem (A.11) can be solved analytically. The solution consists in an infinite series of solutions for the heat diffusion problem on the half line, i.e.,

$$\begin{aligned} T_O(z,t) &= T_S(t) - \int_0^t \dot{T}_S(t')\,\mathrm{Erf}\left(\frac{z}{2\sqrt{\kappa_v(t-t')}}\right) dt' \\ &\quad + \sum_{n=1}^{+\infty} (-1)^n \int_0^t \dot{T}_S(t') \left(\mathrm{Erf}\left(\frac{2nz_B - z}{2\sqrt{\kappa_v(t-t')}}\right) - \mathrm{Erf}\left(\frac{2nz_B + z}{2\sqrt{\kappa_v(t-t')}}\right)\right) dt' \,. \end{aligned} \tag{A.12}$$

Expression (A.12) allows us to calculate the heat flux into the interior ocean as a function of the mixed layer temperature $T_S$, which then can be inserted into Equation (A.9) to close the energy balance model (A.8)-(A.9) (see Section 2.1.2). Once Equations (2.5)-(2.6) are solved, we can recover the development of the entire temperature anomaly profile in the interior ocean by use of Equation (A.12).

## A.4 Model parameterisation

Equations (2.5) and (2.6) constitute the dynamic core of DOECLIM, which is complemented by Expression (2.7) for the heat flux in the interior ocean. The model parameters are summarised in Table 2.1. For the purpose of the imprecise probability analysis conducted in the main body of this thesis, we want to limit the free model parameters to the global climate sensitivity $T_{2x}$, and the effective vertical diffusivity $\kappa_v$ of heat in the ocean. Hence, we have to determine the remaining parameter values from a comparison with complementary data that are not directly related to the global mean temperature response on secular time scales.

We have used seasonal data to estimate the land-sea heat exchange coefficient $k$ and the thermal inertia of the land-troposphere box $C_{AL}$ (see Section A.4.1). Admittedly, this constitutes a difficult choice since our highly aggregated energy balance model is better suited to simulate secular trends in response to a much smaller forcing modulation than is represented by the seasonal cycle. On secular time scales, the interior ocean rather than the mixed layer dominates the inertia of the earth system, so that a detailed representation of the atmosphere-ocean interaction becomes less important. Nevertheless, simple energy balance models have been successfully calibrated with seasonal data (Schlesinger et al., 1997), or even been used to deduce estimates for climate sensitivity from seasonal variations (Dutton, 1995). Therefore, we consider it a viable approach to calibrate those model parameters with seasonal data that influence the model response on monthly time scales ($k$, $C_{AL}$).

Concerning the remaining parameters, we have estimated the marine surface air warming enhancement $b_{SI}$ (see Section A.4.2) and the climate feeback parameters $\lambda_L$

and $\lambda_S$ (see Section A.4.3) from an ensemble of model runs with the climate model of intermediate complexity CLIMBER-2 (Schneider von Deimling et al., 2005). The geographical parameters $f_L$ and $f_{SO}$ can be derived from topographical data of the earth surface. The choice of a "bucket" ocean with depth $z_B = 4000\,\mathrm{m}$ is motivated in Appendix B. It is the usual choice for 1-D upwelling-diffusion ocean models with uniform water column (Wigley and Raper, 1992; Schlesinger et al., 1997). The specific heat capacity $c_V$ of seawater is well known from the literature (e.g., Dickinson, 1981).

### A.4.1 Estimation of thermal inertia and land-sea heat flux

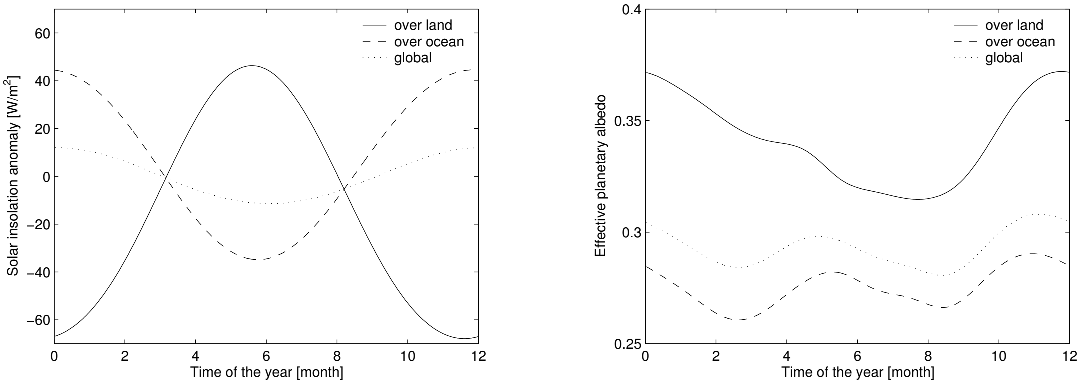
We follow the approach by Schlesinger et al. (1997), who used the model response to the seasonal cycle of solar insolation to calibrate the effective heat capacities and the land-sea heat exchange coefficient $k$. Fig. A.1.a shows the seasonal cycle of solar insolation averaged over land, ocean and global surface area around the mean value of $S_0/4 \approx 341.5\,\mathrm{W\,m^{-2}}$. It was calculated by weighing a latitudinally resolved map of daily insolation with fractional land, ocean, and global surface area, respectively. The average solar insolation at a given latitude and day of the year can be derived from the seasonal variation of solar declination (latitude where sun is overhead at noon) and sun-earth distance (Peixoto and Oort, 1992, Chap. 6.3.2). Due to the concentration of land masses on the Northern hemisphere (NH), the seasonal cycle of solar insolation over land is quite pronounced. It peaks during NH summer, and attains its minimum in NH winter. Since the larger part of the ocean area lies in the Southern hemisphere, the solar insolation over the ocean shows the opposite, albeit less pronounced behavior. The global average insolation reflects the fact that the earth reaches its aphelion during NH summer.

For estimating the seasonal cycle of solar radiation that is absorbed by the earth system, we also need to know the albedo of land and ocean areas. More precisely, the global daily radiation $SQ_i(t)$ taken up by surface type $i = L, O$ (land/ocean) is determined by the solar constant $S_0$, the modulation of solar insolation $DQ(\theta, t)$ at a given latitude and day of the year, the fractional longitudinal and latitudinal distribution $f_i(\phi, \theta)$ of a particular surface type, and the spatial and seasonal variation of planetary albedo $\alpha(\phi, \theta, t)$, i.e.,

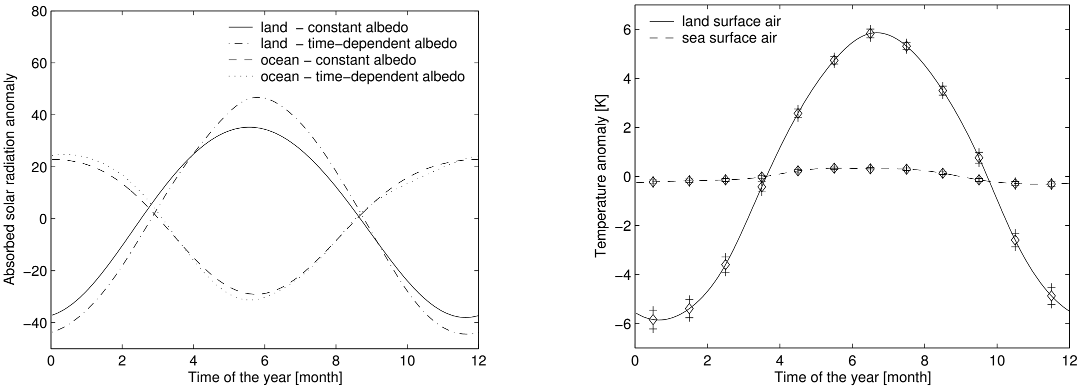
$$\begin{aligned} SQ_i(t) &= \frac{1}{f_i}\frac{S_0}{4}\frac{1}{4\pi}\int_{-\pi/2}^{\pi/2}\int_{-\pi}^{\pi} DQ(\theta,t)\, f_i(\phi,\theta)\ (1-\alpha(\phi,\theta,t))\ d\phi\ \cos(\theta)\, d\theta \\ &:= \frac{S_0}{4}\,\overline{DQ}_i(t)\ (1-\overline{\alpha}_i(t)) \qquad \text{(A.13)} \end{aligned}$$

$$\Rightarrow \overline{DQ}_i(t) = \frac{1}{f_i}\frac{1}{4\pi}\int_{-\pi/2}^{\pi/2}\int_{-\pi}^{\pi} DQ(\theta,t)\, f_i(\phi,\theta)\, d\phi\ \cos(\theta)\, d\theta \qquad \text{(A.14)}$$

$$\Rightarrow \quad \overline{\alpha}_i(t) = \frac{\int_{-\pi/2}^{\pi/2}\int_{-\pi}^{\pi} DQ(\theta,t)\, f_i(\phi,\theta)\, \alpha(\phi,\theta,t)\, d\phi\ \cos(\theta)\, d\theta}{\int_{-\pi/2}^{\pi/2}\int_{-\pi}^{\pi} DQ(\theta,t)\, f_i(\phi,\theta)\, d\phi\ \cos(\theta)\, d\theta}, \qquad \text{(A.15)}$$



a. Seasonal cycle of solar insolation b. Seasonal variations of planetary albedo



c. Seasonal anomalies in absorbed solar radiation d. Seasonal temperature anomalies

Figure A.1: Seasonal cycles of solar insolation (a), effective planetary albedo (b), solar radiation absorbed by the earth system (c), and corresponding surface air temperatures (d). Shown are the averages over land and ocean surfaces. The monthly temperature data (indicated by diamonds with error bars delineated by '+') was interpolated with a cubic spline.

where $\overline{DQ}_i(t)$ is the modulation of solar insolation at a given day of the year averaged over surface type $i$, and $\overline{\alpha}_i(t)$ the daily average effective planetary albedo over surface type $i$. On the annual average, the land surface receives only a fraction $\overline{DQ_L} = 0.96$ of the solar insolation $S_0/4$, while the ocean receives a fraction $\overline{DQ_O} = 1.02$.

We use the 2.5° × 2.5° climatology of monthly mean planetary albedo provided by the NASA Earth Radiation Budget Experiment (ERBE, data publicly available at iridl.ldeo.columbia.edu/SOURCES/.NASA/) to derive the seasonal variation of albedo over land, ocean and global surface areas. Missing data for high latitude NH albedo during NH winter months, and high latitude SH albedo during NH summer months were filled by inserting the maximum measured albedo at the respective grid points. We generated a daily time series by interpolating the monthly data at each grid point with a cubic spline. Fig A.1.b shows the seasonal variations of the effective planetary albedo $\overline{\alpha}_i(t)$ over land, ocean, and on the global average. The land albedo is clearly dominated by the NH seasonal cycle, reaching its maximum during NH winter and its minimum during NH summer. The situation is different for the albedo of the oceans. It peaks both during NH winter and SH winter, when the sea ice extent in the Arctic ocean and around Antarctica, respectively, reaches its maximum. In the annual average, the albedo $\overline{\alpha_L} = 0.35$ over land is larger than the albedo $\overline{\alpha_O} = 0.28$ over the oceans.

Fig. A.1.c shows the seasonal anomalies of the solar radiation absorbed over land and over oceans from the respective annual mean values (dashed-dotted line). They are compared with the anomalies that would be derived under the assumption of a time-invariant albedo fixed at the annual average value $\overline{\alpha_i}$ (solid line). It can be seen that the seasonal variation of albedo enhances the seasonal cycle of absorbed solar radiation. While the effect is small over the oceans, it is considerable over land. In particular, the distinct minimum of land albedo during NH summer enhances the positive anomaly of absorbed radiation in NH summer by approx. 30%, and delays its maximum by approx. 1 to 2 weeks.

The radiation anomalies can be compared with the seasonal cycle of temperature anomalies over land and oceans as shown in Fig. A.1.d. The data were derived from the monthly global mean land and sea surface temperature time series for the period 1880-2003 provided by the US National Climatic Data Center (NCDC; data publicly available at www.ncdc.noaa.gov/oa/climate/research/anomalies/anomalies.html). The time series was based on the monthly mean 1961-90 climatology of Jones et al. (1999). Seasonal anomalies showed no significant trend over the last 120 years, so that the mean anomaly averaged over the period 1880-2003 and its standard deviation could be directly estimated for each month. The much larger amplitude of the temperature anomaly over land than over the oceans indicates the much smaller thermal inertia of the land-troposphere system compared to the ocean mixed layer. Since the seasonal signal over land is more pronounced, we will try to determine the land-sea heat exchange coefficient $k$ by fitting Equation (2.5) to the observed land temperature anomalies.

The energy balance model was initially constructed to simulate annual mean temperature changes in response to secular perturbations of the radiation balance. We need to check carefully if the same modelling framework can be used to simulate daily mean temperature anomalies in response to the seasonal cycle of absorbed solar radiation. Therefore, we repeat the perturbation analysis around the radiative equilibrium as outlined in Appendix A for the case of seasonal radiation changes over land. By neglecting any decadal or slower perturbations to the radiation balance, we assume that the outgoing longwave radiation from the land surface ($\sigma T_L^4$) plus the heat flux to the sea equals, on the annual average, the solar radiation absorbed over land plus the additional radiation energy $G$ provided by the greenhouse effect (cf. Equation 2.2), i.e.,

$$E := G + (1 - \overline{\alpha_L})\,\overline{DQ_L}\,\frac{S_0}{4} = \sigma T_L^4 + \frac{k}{f_L}\,(T_L - \tilde{T}_S)\,.$$

Here, $\tilde{T}_S$ denotes the sea surface temperature as depicted in Fig. A.1.d. We neglect warming differentials between sea surface and marine air temperature due to changes in sea ice cover, since seasonal changes in Arctic and Antarctic sea ice roughly counterbalance each other (Peixoto and Oort, 1992, Fig. 9.3).

A seasonal perturbation $\Delta E(t)$ of the radiation balance leads to a heat flux $C_{AL}\dot{T}_L(t)$ into the land-troposphere system, and consequently to a change $\Delta T(t)$ of land temperature, i.e.

$$C_{AL}\dot{T}_L(t) = E + \Delta E(t) - \sigma\left(T_L + \Delta T_L(t)\right)^4 - \frac{k}{f_L}\left(T_L + \Delta T_L(t) - \tilde{T}_S + \Delta\tilde{T}_S(t)\right).$$

Neglecting second order perturbation terms, we can approximate the radiative damping by a linear term, and separate the radiative perturbation into a greenhouse effect anomaly, a seasonal anomaly of solar insolation over land (for annual mean albedo), and a seasonal perturbation of land albedo (for annual mean insolation), i.e.

$$\begin{aligned} C_{AL}\dot{T}_L(t) \quad \approx \quad & \Delta G(t) + (1-\overline{\alpha_L})\,\Delta\overline{DQ}_L(t)\,\frac{S_0}{4} - \Delta\overline{\alpha}_L(t)\,\overline{DQ_L}\,\frac{S_0}{4} \\ & -4\sigma T_L^3 \Delta T_L(t) - \frac{k}{f_L}\left(\Delta T_L(t) - \Delta\tilde{T}_S(t)\right), \end{aligned}$$

where $\Delta\overline{DQ}_L(t) := \overline{DQ}_L(t) - \overline{DQ_L}$, and $\Delta\overline{\alpha}_L(t) := \overline{\alpha}_L(t) - \overline{\alpha_L}$. While the change of solar insolation constitutes the radiative forcing that drives the seasonal cycle, the land albedo and the greenhouse effect respond to changes of the land temperature. If we assume that the temperature feedback on the radiation balance varies linearly with temperature change around its equilibrium value, we can add up the greenhouse, albedo and direct radiative feedback, so that

$$C_{AL}\dot{T}_L(t) = (1-\overline{\alpha_L})\,\Delta\overline{DQ}_L(t)\,\frac{S_0}{4} - \lambda_L\,\Delta T_L - \frac{k}{f_L}\left(\Delta T_L - \Delta\tilde{T}_S\right), \tag{A.16}$$

where $\lambda_L$ denotes the climate feedback parameter discussed above. The radiative forcing $Q(t) = (1-\overline{\alpha_L})\,\Delta\overline{DQ}_L(t)\,S_0/4$ in Equation (A.16) is the daily solar insolation that would be absorbed by a land mass with the annual mean albedo $\overline{\alpha_L}$ (solid line in Fig. 5c).

The climate feedback parameter $\lambda_L$ will be of similar magnitude on seasonal and secular time scales, if the feedback processes over land respond very fast, i.e., on the order of hours and days, to a temperature change. This is particularly true for the greenhouse feedback, which involves changes of atmospheric water vapour content, lapse rate and cloud cover. Analyses of ERBE data have shown that the seasonal change of the natural greenhouse effect scales indeed fairly linear with seasonal surface temperature anomalies (e.g., Dutton, 1995). The albedo feedback is more critical, since the response of continental snow cover might lag surface temperature up to several weeks. Fig. A.2 clearly shows the delay of the albedo response to rising temperatures in spring (MAM), which leads to deviations from a linear relationship between the two quantities. Since Model (A.16) neglects the inertia in the albedo feedback, a comparison with the temperature data depicted in Fig. A.1.d might give rise to a bias in the estimates for the

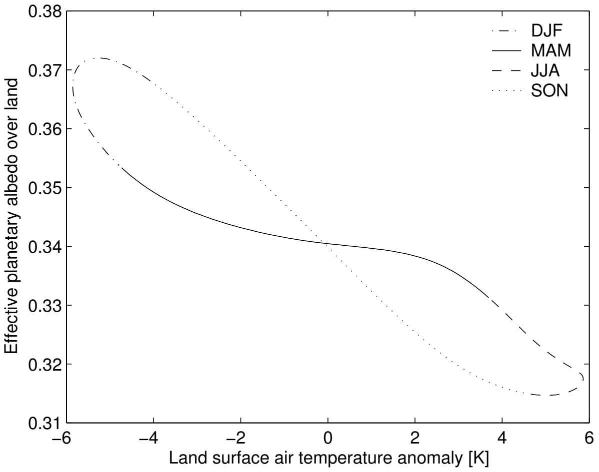


Figure A.2: Relationship between the seasonal cycle of land surface air temperature and the seasonal variation of effective planetary albedo over land.

land-sea heat exchange coefficient $k$ and the effective heat capacity $C_{AL}$ of the land-troposphere system. We check for such a possible bias by comparing the observations with a second model, which treats the albedo feedback separately, i.e.,

$$C_{AL}\dot{T}_L(t) = (1-\overline{\alpha_L})\,\Delta\overline{DQ}_L(t)\,\frac{S_0}{4} - \Delta\overline{\alpha}_L(t)\,\overline{DQ_L}\,\frac{S_0}{4} - \lambda_L'\,\Delta T_L - \frac{k}{f_L}\,(\Delta T_L - \Delta\tilde{T}_S)\,. \tag{A.17}$$

The first two terms on the right-hand side of Equation (A.17) summarize, in a first order approximation, the seasonal anomaly of absorbed solar radiation over land including the albedo feedback (dashed-dotted line in Fig. A.1.c). The second order term proportional to $\Delta\overline{DQ}_L(t)\,\Delta\overline{\alpha}_L(t)$ contributes only approx. $1\,\mathrm{W\,m^{-2}}$ to the anomaly, and therefore can be neglected. The climate feedback parameter $\lambda_L'$ in Model (A.17) explicitely excludes the albedo feedback, and therefore overestimates the negative temperature feedback on the radiation balance. Comparing with Model (A.16), it is $\lambda_L' > \lambda_L$.

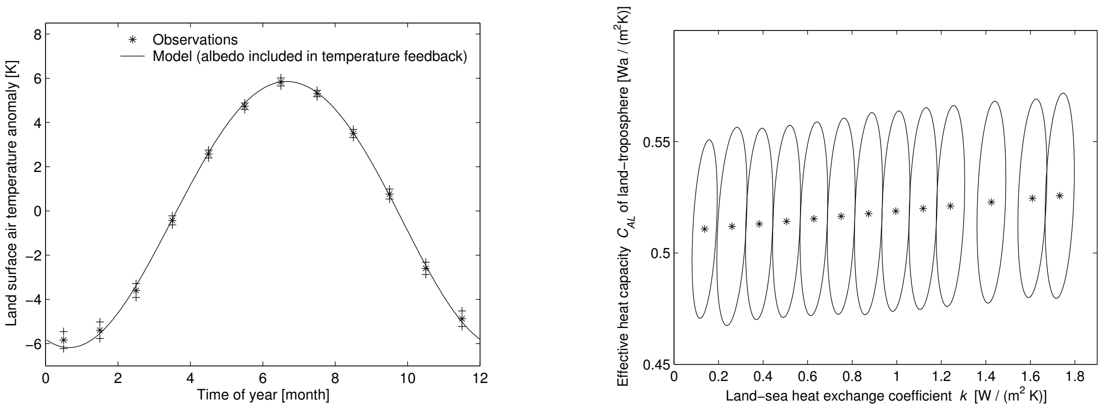
We have used a simple Euler scheme to solve Models (A.16) and (A.17) numerically. The time step of the discretization, i.e. one day, is an order of magnitude smaller than the time scales of the temperature feedback and the heat flux to the ocean. The seasonal anomaly of absorbed solar radiation, for annual average land albedo in Model (A.16) and seasonally varying albedo in Model (A.17), was prescribed (see Fig. A.1.c). Likewise, the observed seasonal anomaly of sea surface temperature (see Fig. A.1.d) was used to calculate the heat flux to the ocean. Models (A.16) and (A.17) generate the seasonal cycle of land surface air temperature anomaly $\Delta T_L(t)$, whose shape and amplitude depends on three parameters: the climate feedback parameter $\lambda_L$, and $\lambda_L'$ respectively, the land-see heat exchange coefficient $k$, and the effective heat capacity $C_{AL}$ of the land-troposphere system. We have estimated the latter two parameters $k$ and $C_{AL}$ from a comparison of the model output with the observed monthly mean land temperature anomalies (see Fig. A.1.d) for fixed values of

$\lambda_L(\lambda'_L) \in [-0.5\,\mathrm{W\,m^{-2}K^{-1}}, 5.1\,\mathrm{W\,m^{-2}K^{-1}}]$.

It can be expected that the seasonal data for the period 1880-2003 scatters around the monthly mean climatology in a highly correlated manner. Therefore, it would be misleading to fit Models (A.16) and (A.17) to the data by directly minimizing the least square sum of the residuals between monthly averaged model output and observations. Instead, we have estimated the covariance matrix $\mathrm{Cov}_{ij}$ between temperature deviations from the mean climatology in months $i$ and $j$. It turned out that $\mathrm{Cov}_{ij}$ was degenerate so that one linear combination of monthly mean deviations from the climatology was fully determined by knowledge of the remaining 11 independent linear combinations. We have calculated this 11D-eigenbasis of the 12D-measurement space by diagonalising the covariance matrix, and projected both observations and model output onto it. Then Models (A.16) and (A.17) were fitted to the data by minimizing the least square sum of the residuals between the projected observations and model outputs weighted by the variance of the data along the directions of the eigenbasis. Error ellipsoids around the minimizing point $(k^*, C^*_{AL})$ were calculated from the curvature of the least square sum around the minimum. Under the assumption of normally distributed residuals on this eigenbasis, we expect the least square sum to constitute a realisation of a $\chi^2$ distribution with 9 degrees of freedom (11 data points, 2 parameters). Hence, a 95% confidence ellipsoid encloses the parameter region, for which $\chi^2(k, C_{AL}) < 16.92$.

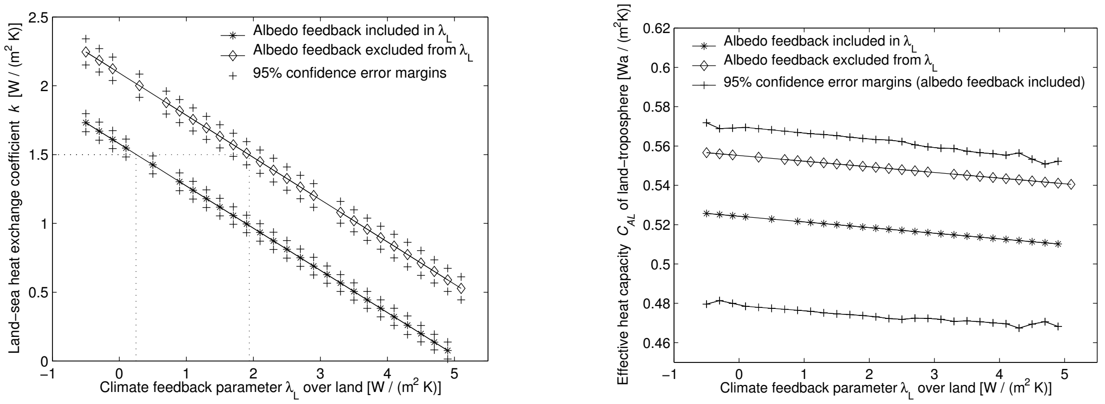
Fig. A.3 summarizes the result of the model data comparison. A typical fit to the data is shown in Fig. A.3.a for Model (A.16) and $\lambda_L = 1.5\,\mathrm{W\,m^{-2}K^{-1}}$. The agreement is generally very good except for NH winter, for which the model shows a slightly larger cooling. The least square minima for Models (A.16) and (A.17) and different values of climate feedback strength $\lambda_L$ exhibited $\chi^2(k, C_{AL})$-values between 4 and 7. 60% to 90% of the realizations of a $\chi^2$ random variable with 9 degrees of freedom would have yielded higher values. This indicates that the agreement between model and data is too good given the error estimates. The errors might be slightly overestimated due to a small trend in some months of the seasonal cycle during the period 1880-2003. Nevertheless, the data strongly constrains the two parameters $k$ and $C_{AL}$. Fig. A.3.b shows the '95% confidence'-ellipsoids for Model (A.16) and 13 different values of $\lambda_L$. The ellipsoids show a small positive correlation between $k$ and $C_{AL}$.

The best estimate of the land-sea heat exchange coefficient $k^*$ scales strictly linear with the climate feedback parameter $\lambda_L$, and $\lambda'_L$ respectively (see Fig. A.3.c). The error margins on the $k^*$ values (depicted by +) constitute the projection of the '95% confidence'-ellipsoids onto the $k$-axis. The two Models (A.16) and (A.17) produce exactly the same slope of the linear dependence between $k^*$, and $\lambda_L$ and $\lambda'_L$, respectively. Their $\lambda$-offset is $\lambda'_L(k^*) - \lambda_L(k^*) = 1.69\,\mathrm{W\,m^{-2}K^{-1}}$ (see dotted lines in Fig. A.3.c). It reflects the fact that $\lambda_L$ includes all temperature feedbacks, while $\lambda'_L$ does not contain the positive ($\lambda$ reducing) albedo feedback. Hence, the offset can be used to estimate the impact of the albedo feedback on the radiation balance to amount to approx. $1.7\,\mathrm{W\,m^{-2}}$ per unit temperature. This value agrees well with estimates that can be deduced from Fig. (A.2). An albedo reduction $\Delta\alpha = 0.06 - 0.07$ for a temperature increase of



a. Modelled seasonal cycle and observations

b. 95%-ellipsoids around the $(k^*, C^*_{AL})$ estimates



c. Estimated land-sea heat exchange coefficient $k$

d. Estimated effective heat capacity $C_{AL}$

Figure A.3: Comparison of Models (A.16) and (A.17) with observed seasonal cycle in the period 1880-2003.

$\Delta T = 12\,\mathrm{K}$ amounts to an increase $\Delta\alpha\overline{DQ_L}S_0/4 \approx 20 - 23\,\mathrm{W\,m^{-2}}$ of absorbed solar radiation over land.

We conclude that Model (A.16) provides a very good representation of the seasonal temperature anomaly over land. Including the albedo feedback in the linear temperature feedback terms does not bias the estimates for $k^*$. A strictly linear relationship between the land-sea heat exchange coefficient $k^*$ and the climate feedback parameter $\lambda_L$ over land can be identified:

$$k = b_k - a_k\,\lambda_L\ , \qquad \text{with}\ \ b_k = 1.59 \pm 0.063\,\mathrm{W\,m^{-2}K^{-1}}\ , \quad a_k = 0.31 \qquad \text{(A.18)}$$

Since the best fit values $k^*$ do not scatter around the linear relationship, we assume that the slope estimate contains only a negligible error. The error margin on the intersect was calculated from the average range of the '95% confidence'-intervals around the best fit values. The relative error on the intersect is approximately 4%, and will be neglected in the following.

The best estimates $C^*_{AL}$ for the effective heat capacity of the land-troposphere system are shown in Fig. A.3.d. They exhibit only a small dependence on the climate feedback parameter $\lambda_L$. As a consequence, they also do not depend strongly on whether the albedo feedback is included or excluded in $\lambda_L$. The estimates generated with Model (A.17) lie within the range of the 95%-confidence error margins of the $C^*_{AL}$ values estimated with Model (A.16). For the purpose of our study, it suffices to choose a value $C_{AL} = 0.52\,\mathrm{Wyr\,m^{-2}K^{-1}}$ from the middle of the range of possible values. Such a heat capacity corresponds to a land column of approx. 8 m. Since the atmospheric heat capacity equals the heat capacity of a land column of approx. 5 m, the remaining 3 m are the effective depth of the land surface that contributes to the overall heat capacity of the land-troposphere system on seasonal time scales. It needs to be noted that the effective land heat capacity increases with the time scale as the heat can penetrate deeper into the ground. However, this effect will play only a minor role on secular and shorter time scales.

It would be desirable to estimate the effective heat capacity $C_{AS}$ of the ocean mixed layer-troposphere system in the same manner. In principle, we could try to fit Equation (2.6) to the observed seasonal cycle of global sea surface temperature (SST) (see Fig. A.1.d) by prescribing the seasonal variation of absorbed solar radiation over the oceans (see Fig. A.1.c), and the seasonal anomalies of land temperature. However, the sea surface temperature anomaly shows a flat maximum during NH summer, when the global mean absorbed solar radiation is at its minimum. The anti-correlation results from the difference in land fraction between Northern and Southern hemisphere. The strong increase of SSTs in the 30N to 60N latitudinal band during NH summer more than outweighs the SST cooling in the 30S to 60S latitudinal band, even when correcting for the larger ocean area in the Southern hemisphere. Such a behaviour can not be explained solely on the basis of heat fluxes between two aggregated land and ocean boxes. Its description would require, as a minimum, the resolution of Northern and Southern hemisphere.

Hence, Equation (2.6) can reproduce the amplitude, but not the phase of the seasonal cycle of global mean SST. If we restrict ourselves to the amplitude, we find an effective heat capacity in the range $C_{AS} = 5.5\,\mathrm{Wyr\,m^{-2}K^{-1}} - 10.5\,\mathrm{Wyr\,m^{-2}K^{-1}}$ depending on different assumptions about the strength of the land-sea heat exchange. This corresponds to an effective mixed layer depth of approximately $40\,\mathrm{m} - 80\,\mathrm{m}$, including the thermal inertia of the troposphere (which is equivalent to the inertia of a water column with $\approx 2.5\,\mathrm{m}$ depth). For the purpose of our analysis, we choose an effective mixed layer depth of 60 m ($7.8\,\mathrm{Wyr\,m^{-2}K^{-1}}$) from the center of this range. This value has been adopted in the energy balance model MAGICC for emulating the behavior of several AOGCMs (Raper et al., 2001). Schlesinger et al. (1997) find an effective value of 55.9 m (53.9 m ocean mixed layer + 2 m tropospheric contribution) from a fit of their hemispherically resolved energy balance model to the seasonal cycle of land and sea surface temperatures in both hemispheres.

Given the difficulty to constrain the effective heat capacity $C_{AS}$ by seasonal data,

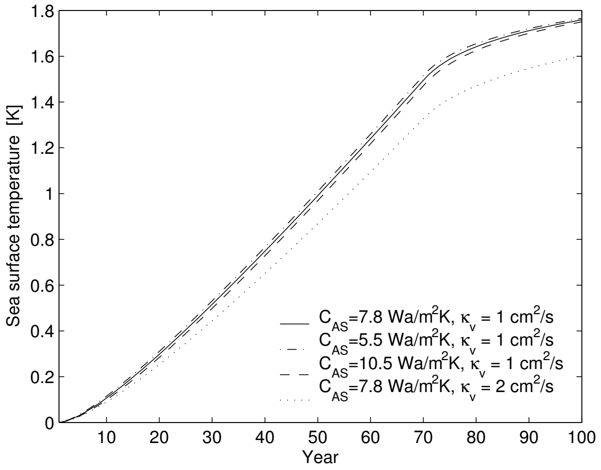


Figure A.4: Dependence of sea surface temperature response on the effective heat capacity $C_{AS}$ of the ocean mixed layer-troposphere system compared to its dependence on the effective vertical heat diffusivity $\kappa_v$.

it is fortunate that it does not influence markedly the model response to radiative perturbations on decadal to secular time scales. The transient temperature behaviour on these time scales is dominated by the heat flux into the interior ocean. Fig. A.4 shows the sea surface warming in $CO_2$ doubling experiments with two different values of vertical diffusivity, and three different values of the troposphere-ocean mixed layer heat capacity which span the plausible range of $C_{AS}$ identified above. The experiments consisted in a linear increase of radiative forcing $Q$ from a presupposed equilibrium state with 280 ppm $CO_2$ ($Q = 0\,\mathrm{W\,m^{-2}}$) to 580 ppm $CO_2$ in the atmosphere ($Q_{2x} = 3.7\,\mathrm{W\,m^{-2}}$) in 70 years, with stabilized radiative perturbation thereafter. It can be seen that the influence of effective heat capacity on the model response can be neglected when compared to the influence of vertical diffusivity. We want to point out, however, that this observation is concerned with decadal to secular time scales. The picture looks different on shorter time scales. The statistical properties of annual variability in SSTs due to stochastic weather forcing will depend strongly on the value of $C_{AS}$.

### A.4.2 Estimation of marine air warming enhancement

Raper and Cubasch (1996) have estimated that retreating sea ice leads to a 20% larger warming of marine surface air than of the sea surface itself ($b_{SI} \approx 1.2$). To assess their estimate, we have compared the increase in sea surface temperature (SST) to the corresponding increase in marine surface air temperature (MAT) in a dataset of $CO_2$ doubling experiments with the model CLIMBER-2 (Schneider von Deimling et al., 2005). CLIMBER-2 is a 2.5-dimensional climate system model with a statistical-dynamical atmosphere module coupled to ocean, sea ice, and terrestrial vegetation modules (Petoukhov et al., 2000). The dataset comprised 62 individual model runs with perturbed parameter values covering a wide range of climate sensitivity. The parameters related to a variety of processes like the radiative scheme, heat transport in the ocean, and atmospheric lapse rate, cloud cover, and water vapour. Despite the heterogeneity among individual experiments, they all agreed very well on the ratio between MAT and SST anomalies. At the time of $CO_2$ doubling (year 70), all experiments showed ratios in the range $b_{SI} = 1.43 \pm 0.05$, which decreased to $b_{SI} = 1.32 \pm 0.05$ after 300 years, and remained constant (with narrowing range) thereafter. During the first hundred years, a small time dependence of $b_{SI}$ could be identified indicating that SST response slightly lags behind MAT response. In our model, we have neglected this time dependence. It will reduce even further, when considering less extreme greenhouse gas scenarios than a $CO_2$ doubling within 70 years. On the basis of the CLIMBER-2 $CO_2$ doubling experiments, we choose a value of $b_{SI} = 1.3$, which is close to the estimate of Raper and Cubasch (1996). We neglect the associated uncertainty, probably less than ±10%, since the marine air warming enhancement $b_{SI}$ exerts only a small influence on the model behaviour compared to $k$, $\lambda_L$, $\lambda_S$ and $\kappa_v$ (see, e.g., Equation A.22).

### A.4.3 Estimation of climate feedback strength

As mentioned in Section 2.1.2, we have investigated the relationship between the climate sensitivity $T_{L,2x}$ of the surface air over land and the climate sensitivity $T_{S,2x}$ of the sea surface with data from the ensemble of CLIMBER-2 $CO_2$ doubling experiments described above (Schneider von Deimling et al., 2005). It entails model sensitivities $T_{2x}$ that range from 1.6 K to 4.2 K, and therefore provides a particularly useful basis for the investigation. Fig. A.5 plots $T_{L,2x}$ versus $T_{S,2x}$ for all 62 model runs in the data set. The linear relationship between climate sensitivity over land and climate sensitivity of the sea surface is evident. A linear regression of the data yields $T_{L,2x} = R_\lambda T_{S,2x}$ with $R_\lambda = 1.43 \pm 0.02$. The error range specifies the 95% confidence interval that the slope of a supposedly linear relationship lies indeed in this interval under the assumption of independent and identically distributed (IID) normal fluctuations in the model behavior due to subglobal scale processes. We do not further investigate whether this assumption is justified, since the confidence interval is merely used to illustrate how strongly the linear regression is constrained by the data. For our modelling purpose, we neglect the minor uncertainty, and only consider the best estimate of $R_\lambda = 1.43$. This value is close to other land-sea ratios of climate sensitivity employed in EBMs for the emulation of climate change simulations in complex AOGCMs (Raper et al., 2001; Cubasch and Meehl, 2001).

Knowledge of $R_\lambda$ and the global climate sensitivity $T_{2x}$ suffices to determine $T_{L,2x}$ and $T_{S,2x}$:

$$T_{L,2x} = \frac{R_\lambda T_{2x}}{R_\lambda f_L + (1 - f_L) b_{SI}}, \qquad T_{S,2x} = \frac{T_{2x}}{R_\lambda f_L + (1 - f_L) b_{SI}}.$$

By use of Equation (2.8) linking the climate feedback parameters $\lambda_L$ and $\lambda_S$ to $T_{L,2x}$ and $T_{S,2x}$, respectively, and Equation (A.18) specifying the land-sea heat exchange

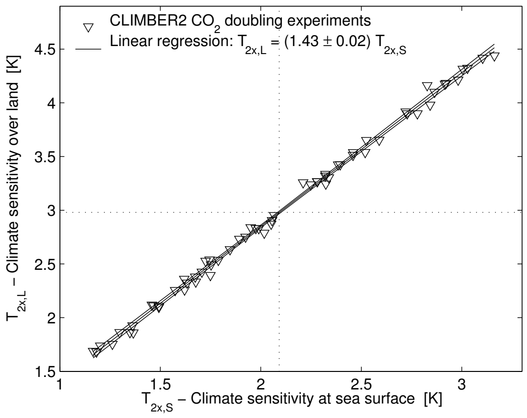


Figure A.5: Linear relationship between climate sensitivity over land and at the sea surface exhibited by an ensemble of 62 CO$_2$ doubling experiments with perturbed versions of the CLIMBER-2 model. Also shown are the linear regression with 95% confidence bounds (under the assumption of normal IID deviations), and the ensemble means for $T_{2x,L}$ and $T_{2x,S}$ (dotted lines).

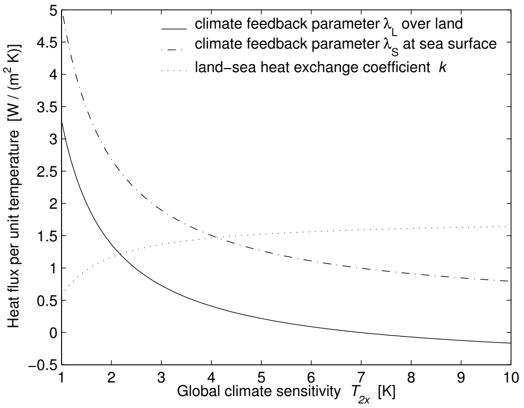


Figure A.6: Dependence of $\lambda_L$, $\lambda_S$ and $k$ on global climate sensitivity $T_{2x}$.

coefficient $k$ as function of $\lambda_L$, we can express $\lambda_L$, $\lambda_S$, and $k$ as functions of global climate sensitivity $T_{2x}$:

$$\lambda_L = f_L \, \frac{R_\lambda \, f_L + (1 - f_L)\, b_{SI}}{R_\lambda \, f_L - a_k \, (R_\lambda - b_{SI})} \, \frac{Q_{2x}}{T_{2x}} - b_k \, \frac{R_\lambda - b_{SI}}{R_\lambda \, f_L - a_k \, (R_\lambda - b_{SI})} \, , \tag{A.19}$$

$$\begin{aligned} \lambda_S = & \left( R_\lambda \, f_L - a_k \, \frac{(R_\lambda - b_{SI})}{1 - f_L} \right) \frac{R_\lambda \, f_L + (1 - f_L)\, b_{SI}}{R_\lambda \, f_L - a_k \, (R_\lambda - b_{SI})} \, \frac{Q_{2x}}{T_{2x}} + \\ & \frac{R_\lambda \, f_L}{1 - f_L} \, b_k \, \frac{R_\lambda - b_{SI}}{R_\lambda \, f_L - a_k \, (R_\lambda - b_{SI})} \, , \end{aligned} \tag{A.20}$$

$$k = b_k \, \frac{R_\lambda \, f_L}{R_\lambda \, f_L - a_k \, (R_\lambda - b_{SI})} - a_k \, f_L \, \frac{R_\lambda \, f_L + (1 - f_L)\, b_{SI}}{R_\lambda \, f_L - a_k \, (R_\lambda - b_{SI})} \, \frac{Q_{2x}}{T_{2x}} \, . \tag{A.21}$$

The dependence of $\lambda_L$, $\lambda_S$ and $k$ on climate sensitivity $T_{2x}$ is shown in Fig. A.6. The climate feedback parameters decrease in proportion to $1/T_{2x}$ for increasing climate sensitivity, with $\lambda_L$ consistently lower than $\lambda_S$. The land-sea heat exchange coefficient increases in proportion to $-1/T_{2x}$ for increasing climate sensitivity.

## A.5 Numerical integration of the model

Equations (2.5)-(2.6) exhibit six different time scales. With

$$\begin{aligned}
\tau_L &:= \frac{C_{AL}}{\lambda_L}\,, & \tau_{LS} &:= f_L\,\frac{C_{AL}}{k}\,, & \tau_{BO} &:= \frac{z_B^2}{\kappa_v} \\
\tau_S &:= \frac{C_{AS}}{\lambda_S}\,, & \tau_{SL} &:= (1-f_L)\,\frac{C_{AS}}{k}\,, & \tau_{FO} &:= \left(\frac{C_{AS}}{c_V}\right)^2\,\frac{\pi}{\kappa_v}\,,
\end{aligned} \tag{A.22}$$

the model can be rewritten as

$$\dot{T}_L = \frac{Q_L}{C_{AL}} - \frac{T_L}{\tau_L} - \frac{T_L - b_{SI}\,T_S}{\tau_{LS}} \tag{A.23}$$

$$\begin{aligned}
\dot{T}_S = {} & \frac{Q_S}{C_{AS}} - \frac{T_S}{\tau_S} - \frac{b_{SI}\,T_S - T_L}{\tau_{SL}} \\
& - \frac{f_{SO}}{\sqrt{\tau_{FO}}} \int_0^t \frac{\dot{T}_S(t')}{\sqrt{t-t'}} \left(1 + 2\sum_{n=1}^{+\infty} (-1)^n\, e^{-\frac{n^2\tau_{BO}}{t-t'}}\right) dt'.
\end{aligned} \tag{A.24}$$

The time scales vary from the order of months ($\tau_{LS}$, and $\tau_L$ for small climate sensitivity $T_{L,2x}$) to decades ($\tau_S$ for large climate sensitivity $T_{S,2x}$, and $\tau_{FO}$ for large diffusivities $\kappa_v$) to even several thousand years ($\tau_{BO}$). A numerical integration with a simple Euler scheme yields only stable results, if the time step $\Delta t$ of the integration is significantly smaller than the time scale of the fastest process in the model. In our case, $\Delta t$ would need to be on the order of weeks. However, we want to choose a time step of one year for the numerical approximation. Therefore, we need to employ an implicit scheme,

$$\dot{T}_i = \frac{T_{i+1} - T_i}{\Delta t} = f\left(t_i + \frac{\Delta t}{2}, \frac{T_i + T_{i+1}}{2}\right)\,, \quad i: \text{ time step index}\,,$$

to integrate Equations (A.23)-(A.24). The resulting difference equations are

$$\begin{aligned}
\frac{T_{L,i+1} - T_{L,i}}{\Delta t} &= \frac{Q_L(t_i + \frac{\Delta t}{2})}{C_{AL}} - \frac{T_{L,i} + T_{L,i+1}}{2\tau_L} - \frac{(T_{L,i} + T_{L,i+1}) - b_{SI}\,(T_{S,i} + T_{S,i+1})}{2\tau_{LS}}\,, \\
\frac{T_{S,i+1} - T_{S,i}}{\Delta t} &= \frac{Q_S(t_i + \frac{\Delta t}{2})}{C_{AS}} - \frac{T_{S,i} + T_{S,i+1}}{2\tau_S} - \frac{b_{SI}\,(T_{S,i} + T_{S,i+1}) - (T_{L,i} + T_{L,i+1})}{2\tau_{SL}} \\
& \quad - \frac{f_{SO}}{\sqrt{\tau_{FO}\Delta t}} \sum_{j=0}^{i} (T_{S,j+1} - T_{S,j}) \int_j^{j+1} \frac{1 + 2\sum_{n=1}^{+\infty} (-1)^n\, e^{-\frac{n^2\tau_{BO}}{\Delta t}\,\frac{1}{i+1-t'}}}{\sqrt{i+1-t'}}\, dt'\,.
\end{aligned} \tag{A.25}$$

By integrating

$$a_{i-j} \;:=\; \int_j^{j+1} \frac{1 + 2\sum_{n=1}^{+\infty} (-1)^n\, e^{-\frac{n^2\tau_{BO}}{\Delta t}\,\frac{1}{i+1-t'}}}{\sqrt{i+1-t'}}\, dt'$$

$$
\begin{aligned}
= \quad & 2\sqrt{i-j+1}\left(1+2\sum_{n=1}^{+\infty}(-1)^n\, e^{-\frac{n^2\tau_{BO}}{\Delta t}\,\frac{1}{i-j+1}}\right) - 2\sqrt{i-j}\left(1+2\sum_{n=1}^{+\infty}(-1)^n\, e^{-\frac{n^2\tau_{BO}}{\Delta t}\,\frac{1}{i-j}}\right) \\
& + 4\sum_{n=1}^{+\infty}(-1)^{n+1}\, n\sqrt{\pi\frac{\tau_{BO}}{\Delta t}}\left(\mathrm{Erf}\left(n\sqrt{\frac{\tau_{BO}}{\Delta t}\,\frac{1}{i-j}}\right) - \mathrm{Erf}\left(n\sqrt{\frac{\tau_{BO}}{\Delta t}\,\frac{1}{i-j+1}}\right)\right) ,
\end{aligned}
$$

and reshuffling the sum, the difference equation for the mixed layer temperature can be transformed to

$$
\begin{aligned}
\frac{T_{S,i+1}-T_{S,i}}{\Delta t} \quad = \quad & \frac{Q_S(t_i+\frac{\Delta t}{2})}{C_{AS}} - \frac{T_{S,i}+T_{S,i+1}}{2\tau_S} - \frac{b_{SI}\,(T_{S,i}+T_{S,i+1})-(T_{L,i}+T_{L,i+1})}{2\tau_{SL}} \\
& - \frac{f_{SO}}{\sqrt{\tau_{FO}\Delta t}}\,a_0\,T_{S,i+1} + \frac{f_{SO}}{\sqrt{\tau_{FO}\Delta t}}\sum_{j=1}^{i}(a_{i-j}-a_{i-j+1})\,T_{S,j}\,. \qquad \text{(A.26)}
\end{aligned}
$$

We have neglected the term $\frac{f_{SO}}{\sqrt{\tau_{FO}\Delta t}}\,a_i\,T_{S,0}$, since $T_{S,0}=0$ for a presumed equilibrium state at $t=0$. Note that $a_0$ is well defined, since $\lim_{x\to 0} e^{-c/x}=0$ and $\lim_{x\to 0}\mathrm{Erf}(c/x)=1$. The infinite sum in the expression for the $a_{i-j}$ converges very fast, if $\tau_{BO}$ is on the order of magnitude of the time period $\Delta t\,(i-j)$. In the extreme case of a very large vertical diffusivity $\kappa_v = 10\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, the time scale of heat penetration to the bottom at $z_B = 4000\,\mathrm{m}$ is still on the order of several centuries ($\tau_{BO} = 507\,\mathrm{yr}$). We have found for this case that an approximation of $a_{i-j}$ up to the fourth order bottom correction term accurately describes the SST response to a doubling of the atmospheric $CO_2$ concentration until 1000 years after the system has reached a new equilibrium.

For integrating Equations (A.23)-(A.24), we need to calculate the successor temperatures $(T_{L,i+1}, T_{S,i+1})$ from their predecessors $(T_{L,i}, T_{S,j\le i})$. Since equations (A.25) and (A.26) are linear, this can be accomplished with standard techniques of linear algebra. Separating the successor and predecessor temperatures on both sides of the equation, it is

$$
\boldsymbol{B}\cdot\begin{pmatrix} T_{L,i+1} \\ T_{S,i+1} \end{pmatrix} \quad = \quad \boldsymbol{Q}+\boldsymbol{A}\cdot\begin{pmatrix} T_{L,i} \\ T_{S,i} \end{pmatrix}, \qquad \text{(A.27)}
$$

$$
\begin{aligned}
\text{with}\quad \boldsymbol{B} \;&:= \begin{pmatrix} 1+\frac{1}{2}\frac{\Delta t}{\tau_L}+\frac{1}{2}\frac{\Delta t}{\tau_{LS}} & -\frac{1}{2}\frac{\Delta t}{\tau_{LS}}\,b_{SI} \\ -\frac{1}{2}\frac{\Delta t}{\tau_{SL}} & 1+\frac{1}{2}\frac{\Delta t}{\tau_S}+\frac{1}{2}\frac{\Delta t}{\tau_{SL}}\,b_{SI}+f_{SO}\sqrt{\frac{\Delta t}{\tau_{FO}}}\,a_0 \end{pmatrix}, \\
\boldsymbol{Q} \;&:= \begin{pmatrix} \frac{\Delta t}{2}\frac{Q_L(t_i+\frac{\Delta t}{2})}{C_{AL}} \\ \frac{\Delta t}{2}\frac{Q_S(t_i+\frac{\Delta t}{2})}{C_{AS}}+f_{SO}\sqrt{\frac{\Delta t}{\tau_{FO}}}\sum_{j=1}^{i-1}(a_{i-j}-a_{i-j+1})\,T_{S,j} \end{pmatrix} \\
\boldsymbol{A} \;&:= \begin{pmatrix} 1-\frac{1}{2}\frac{\Delta t}{\tau_L}-\frac{1}{2}\frac{\Delta t}{\tau_{LS}} & \frac{1}{2}\frac{\Delta t}{\tau_{LS}}\,b_{SI} \\ \frac{1}{2}\frac{\Delta t}{\tau_{SL}} & 1-\frac{1}{2}\frac{\Delta t}{\tau_S}-\frac{1}{2}\frac{\Delta t}{\tau_{SL}}\,b_{SI}+f_{SO}\sqrt{\frac{\Delta t}{\tau_{FO}}}\,(a_0-a_1)\,, \end{pmatrix}.
\end{aligned}
$$

The sum over predecessors $T_{S,j},\ j < i$ is only evaluated, if $i \geq 2$. The temperatures $(T_{L,1}, T_{S,1})$ in the first time step are calculated from the initial equilibrium conditions $T_{L,0} = T_{S,0} := 0$.

The algorithm to integrate Model (A.23)-(A.24) up to time $t_e = n\,\Delta t$ proceeds as follows:

1. Invert $\boldsymbol{B}$, i.e. find $\boldsymbol{B}^{-1}$ with $\boldsymbol{B}^{-1} \cdot \boldsymbol{B} = \mathbf{1}$. Set $i = 0$.
2. For given $(T_{L,i}, T_{S,j \leq i})$, calculate
$$\begin{pmatrix} T_{L,i+1} \\ T_{S,i+1} \end{pmatrix} = \boldsymbol{B}^{-1} \cdot \left( \boldsymbol{Q} + \boldsymbol{A} \cdot \begin{pmatrix} T_{L,i} \\ T_{S,i} \end{pmatrix} \right)$$
3. Set $i = i + 1$. If $i = n$ abort, otherwise go back to step 2.

This concludes our discussion of the construction, calibration and numerical integration of the diffusion ocean energy balance model DOECLIM.

# Appendix B

# The Upwelling-Diffusion Equation

The transport of heat in a medium obeys the conservation of energy: the net heat flux across the boundaries of a volume segment equals the change in heat content $H := c\,m\,T$ of the segment, where $c$ is thermal capacity (at constant volume), $m$ mass, and $T$ temperature. In one-dimensional problems, where only a cross-sectionally uniform heat transport in the direction $z$ is considered, it is

$$\frac{\partial}{\partial t}\int_{z}^{z+\Delta z} c_V\, A(z)\, T(z,t) = A(z)\, F(z,t) - A(z+\Delta z)\, F(z+\Delta z,t)\ , \tag{B.1}$$

where $c_V = c\,\rho$ is the heat capacity per unit volume (in units of $[\,\mathrm{J\,m^{-3}K^{-1}}]$), here assumed to be independent of $z$, $A(z)$ the cross-sectional area of the medium at $z$, $F(z,t)$ the heat flux per unit area entering the volume segment at $z$, and $F(z+\Delta z,t)$ the heat flux per unit area leaving the segment at $z+\Delta z$.

A heat flux $F(z,t)$ can be induced by two mechanisms: *diffusion* of heat from warmer to colder layers, whose amount is proportional to the temperature gradient between neighbouring layers, and *advection* of heat due to physical movement of the medium in $z$-direction. Hence,

$$F(z,t) = -c_V\, \kappa(z)\, \frac{\partial}{\partial z} T(z,t) + c_V\, \tilde{w}(z)\, T(z,t)\ , \tag{B.2}$$

where $c_V\, \kappa(z)$ is the thermal conductivity of the medium (in units of $[\,\mathrm{W\,m^{-1}K^{-1}}]$), and $\tilde{w}(z)$ the velocity of the medium in $z$-direction. In the limit $\Delta z \to 0$, we find the differential form

$$\begin{aligned} A(z)\, T_t(z,t) &= \partial_z\, (\ A(z)\, \kappa(z)\, T_z(z,t) - A(z)\, \tilde{w}(z)\, T(z,t)\ ) \\ &= \partial_z\, (\ A(z)\, \kappa(z)\, T_z(z,t)\ ) - A(z)\, \tilde{w}(z)\, T_z(z,t)\ , \end{aligned} \tag{B.3}$$

where we have used the short hand $T_x := \partial/\partial x\, T$. The conservation of mass requires that $A(z)\,\tilde{w}(z)$ is constant in $z$.

In our particular case, Equation (B.3) constitutes a model for the heat transport in a highly stylized one-dimensional ocean with $z > 0$ the ocean depth. At the boundary to the ocean mixed layer ($z = 0$), the temperature $T(0,t) := T_S(t)$ is prescribed. The heat is diffused to greater depth, while a cross-sectionally uniform *upwelling* of water masses is assumed to transport heat from the ocean floor at $z := z_B$ to the surface. We are interested in the ocean response to a temperature anomaly at the mixed-layer boundary, and possibly also at the ocean floor. The full solution of Equation (B.3) is given by a superposition of

- its equilibrium solution ($T_t(z,t) = 0$) before the temperature anomaly was imposed, i.e.

  $$T_{\mathrm{eql}}(z) = T_{S,\mathrm{eql}}\, e^{-\int_0^z w(z')/\kappa(z')\,dz'}\ , \tag{B.4}$$

  where $w(z) = -\tilde{w}(z)$ is the *upwelling velocity* against the $z$-direction, and

- a solution to the diffusion-convection equation for the anomalous heat fluxes with initial temperature anomaly $T(z,0) = 0$.

The upwelling-diffusion equation for the anomalous heat fluxes is fully specified by

$$\text{for } 0 < z < z_B: \qquad T_t(z,t) = \frac{1}{A(z)}\,\partial_z\,\big(\,A(z)\,\kappa(z)\,T_z(z,t)\,\big) + w(z)\,T_z(z,t)\ , \tag{B.5}$$

$$\text{Bound. cond. (B.C.):} \qquad T(0,t) = T_S(t) \quad F(z_B,t) = -c_V\,w(z_B)\,T_P(t)\ ,$$

$$\text{Initial cond. (I.C.):} \qquad T(z,0) = 0\ .$$

The boundary condition on the heat flux at the ocean floor needs further explanation. To account for the upwelling water masses in the main ocean column, a second polar ocean column is construed, in which downwelling with velocity $w_P$ occurs. The conservation of mass in the volume exchange between main ocean column with cross-sectional area $A$ and polar column with cross-sectional area $A_P$ requires that $A\,w = A_P\,w_P$ at the surface as well as at the bottom. Since $A \gg A_P$, it is also $w_P \gg w$. Hence, the large downwelling velocity will dominate the heat transport in the polar column, which leads to a fast temperature equilibration in the polar ocean with $T_P(z,t) = T_P(0,t)$ throughout the polar column. The polar water enters the main ocean column at its bottom inducing a heat flux $c_V\,A_P(z_B)\,w_P(z_B)\,T_P(z_B,t) = c_V\,A(z_B)w(z_B)\,T_P(0,t)$, which is then transported to the surface by the upwelling water masses. Hence, the boundary condition at the bottom of the main ocean is constituted by the heat flux due to incoming polar water with temperature $T_P(t)$ and area-averaged velocity $w(z_B)$. Since the upwelling-diffusion equation (B.5) is solely concerned with anomalous heat fluxes, the heat flux at the bottom will depend on the mixed layer temperature anomaly in the polar ocean where downwelling occurs. If the downwelling sites simply shift northwards, so that the temperature of the downwelling water masses remains unchanged from its equilibrium value, $T_P(t) = 0$ and thus $F(z_B,t) = 0$. If, however, the water at the downwelling sites warms along with the increasing mixed layer temperature of the main ocean, $T_P(t) > 0$ and a positive heat flux into the main ocean column at its bottom, i.e. against the $z$-direction, is induced. In most upwelling-diffusion ocean models that are used for the assessment of anthropogenic climate change, the temperature anomaly in the polar ocean is taken proportional to the mixed-layer anomaly of the main ocean, i.e. $T_P(t) = \Pi\, T_S(t)$, where $\Pi \geq 0$ is a tunable parameter.

## B.1 Numerical solutions

A numerical approximation to the partial differential equation (B.5) can be obtained by a finite difference scheme. Discretise the ocean column into $n$ layers with thickness $\Delta z_i$, so that $\sum_{i=1}^{n} \Delta z_i = z_B$. Each layer exhibits a uniform cross-sectional area $A_i$, and a temperature $T_i$. At the boundary between layers $i-1$ and $i$, the vertical diffusivity is given by $\kappa_i$ and the upwelling velocity by $w_i$. Conservation of mass requires $w_i = A_1/A_i\, w_1$, so that we only need to consider the upwelling velocity $w := w_1$ at the boundary to the mixed layer. Then, Equation (B.1) and (B.2) can be used to find the following set of ordinary differential equations approximating Problem (B.5) (cf. Murphy, 1995; Schlesinger et al., 1997):

$$1 < i < n: \qquad \dot{T}_i = -\kappa_i \frac{T_i - T_{i-1}}{0.5\,(\Delta z_{i-1} + \Delta z_i)\,\Delta z_i} - \frac{A_1}{A_i}\, w\, \frac{T_i + T_{i-1}}{2\Delta z_i} + \frac{A_{i+1}}{A_i}\, \kappa_{i+1} \frac{T_{i+1} - T_i}{0.5\,(\Delta z_{i+1} + \Delta z_i)\,\Delta z_i} + \frac{A_1}{A_i}\, w\, \frac{T_i + T_{i+1}}{2\Delta z_i}$$

$$\text{top layer:} \qquad \dot{T}_1 = -\kappa_1 \frac{T_1 - T_S}{0.5\,\Delta z_1\,\Delta z_1} - w\, \frac{T_S}{\Delta z_1} + \frac{A_2}{A_1}\, \kappa_2 \frac{T_2 - T_1}{0.5\,(\Delta z_2 + \Delta z_1)\,\Delta z_1} + w\, \frac{T_1 + T_2}{2\Delta z_1}$$

$$\text{bottom layer:} \qquad \dot{T}_n = -\kappa_n \frac{T_n - T_{n-1}}{0.5\,(\Delta z_{n-1} + \Delta z_n)\,\Delta z_n} - \frac{A_1}{A_n}\, w\, \frac{T_n + T_{n-1}}{2\Delta z_n} + \frac{A_1}{A_n}\, w\, \frac{T_P}{\Delta z_n}\,.$$

To simplify the notation, we introduce the diffusion time scales

$$\tau_{\kappa,1} := \frac{0.5\,\Delta z_1\,\Delta z_1}{\kappa_1}\,, \quad \tau_{\kappa,i>1} := \frac{0.5\,(\Delta z_{i-1} + \Delta z_i)\,\Delta z_i}{\kappa_i}\,, \quad \tau'_{\kappa,i} := \frac{A_i}{A_{i+1}}\, \frac{0.5\,(\Delta z_{i+1} + \Delta z_i)\,\Delta z_i}{\kappa_{i+1}}$$

and the upwelling time scales $\tau_{w,i} := \frac{A_i}{A_1} \frac{\Delta z_i}{w}$.

Then, the ordinary differential equation for the temperature vector $\boldsymbol{T} = (T_1, ..., T_n)^T$ can be written in the compact form

$$\dot{\boldsymbol{T}} - \boldsymbol{A} \cdot \boldsymbol{T} = \boldsymbol{F}\,, \qquad \text{with} \tag{B.6}$$

$$
\begin{aligned}
\boldsymbol{F} &:= \left(\frac{1}{\tau_{\kappa,1}}T_S - \frac{1}{\tau_{w,1}}T_S,\, 0,\, \cdots,\, 0,\, \frac{1}{\tau_{w,n}}\,T_P\right)^T \\
\boldsymbol{A} &:= \begin{pmatrix} -\frac{1}{\tau_{\kappa,1}} - \frac{1}{\tau'_{\kappa,1}} + \frac{1}{2\tau_{w,1}} & \frac{1}{\tau'_{\kappa,1}} + \frac{1}{2\tau_{w,1}} & \cdots & \cdots \\ \frac{1}{\tau_{\kappa,2}} - \frac{1}{2\tau_{w,2}} & -\frac{1}{\tau_{\kappa,2}} - \frac{1}{\tau'_{\kappa,2}} & \frac{1}{\tau'_{\kappa,2}} + \frac{1}{2\tau_{w,2}} & \cdots \\ \ddots & \ddots & \ddots & \ddots \end{pmatrix}
\end{aligned}
$$

The numerical solution to the ordinary differential equation (B.6) requires a close inspection of the time scales involved. An Euler scheme is only a reasonable choice, if the smallest time scale is considerably larger than the time step of the numerical approximation. The smallest time scale in Equation (B.6) is exhibited by the diffusion component for large vertical diffusivity $\kappa$ and small layer thickness $\Delta z_i$. Typical finite difference schemes for a 1D-upwelling-diffusion ocean with $z_B \approx 4000\,\mathrm{m}$ employ layer thicknesses of $\Delta z \approx 100\,\mathrm{m}$. Reasonable values for the effective vertical diffusivity in such models can be as large as $9\,\mathrm{cm}^2\,\mathrm{s}^{-1}$ (Cubasch and Meehl, 2001). Hence, the diffusion time scales $\tau_i \approx (\Delta z)^2/\kappa_i$ can be on the order of months. Since we do not account for seasonal variations, we wish to choose time steps of a year or longer. Therefore, we need to employ an implicit scheme for finding a numerical solution to differential equation (B.6):

$$
\begin{aligned}
\boldsymbol{T}_{j+1} &= \boldsymbol{T}_j + 0.5\,\Delta t\;(\,\boldsymbol{A}\cdot(\boldsymbol{T}_j + \boldsymbol{T}_{j+1}) + \boldsymbol{F}_j + \boldsymbol{F}_{j+1}\,) \\
&\qquad \text{with initial cond. } \boldsymbol{T}_0 = \boldsymbol{0}\,,\;\; \text{time } t \in \{t_0, ..., t_e := t_0 + m\,\Delta t\}\,,\;\; 0 \le j \le m\,.
\end{aligned}
$$

Reshuffling the left- and right-hand side leads to

$$
(\boldsymbol{1} - 0.5\,\Delta t\,\boldsymbol{A})\cdot\boldsymbol{T}_{j+1} = (\boldsymbol{1} + 0.5\,\Delta t\,\boldsymbol{A})\cdot\boldsymbol{T}_j + 0.5\Delta t\;(\boldsymbol{F}_j + \boldsymbol{F}_{j+1}\,)\,. \tag{B.7}
$$

Linear equation (B.7) can be solved with standard techniques, e.g. by diagonalizing $\boldsymbol{B} = (\boldsymbol{1} - 0.5\,\Delta t\,\boldsymbol{A})$ with a transformation $\boldsymbol{P}^{-1}\cdot\boldsymbol{B}\cdot\boldsymbol{P}$, and solving the resulting decoupled set of linear equations for the transformed temperature vectors $\boldsymbol{P}^{-1}\boldsymbol{T}$. We have solved Equation (B.7) for a "bucket" ocean with a uniform diffusivity and cross-section of the water column, a depth of $z_B = 4000\,\mathrm{m}$, a choice of 40 layers with $\Delta z = 100\,\mathrm{m}$ (see Fig. B.1), a choice of time step $\Delta t = 1\,\mathrm{yr}$, and idealized boundary conditions: an instantaneous warming of the mixed layer by one temperature unit at $t = 0$, i.e., $T_S(t) = H(t)$ with $H$ the Heavyside-function, and, in the presence of upwelling, an increase of polar bottom water temperature by $T_P(t) = \Pi\,H(t)$ with $\Pi = \{0, 0.2, 0.8\}$.

Fig. B.2 depicts the resulting ocean temperature anomaly profiles after 100, 500, and 2000 years of applying the instantaneous heat forcing at the surface and, in the cases depicted in Fig. B.2.e to .h, at the bottom. A pure diffusion ocean with vanishing upwelling velocity $w = 0\,\mathrm{m\,yr}^{-1}$ will lead to a uniform warming of the water column in equilibrium. For an effective vertical diffusivity of $\kappa = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, which has been sug-

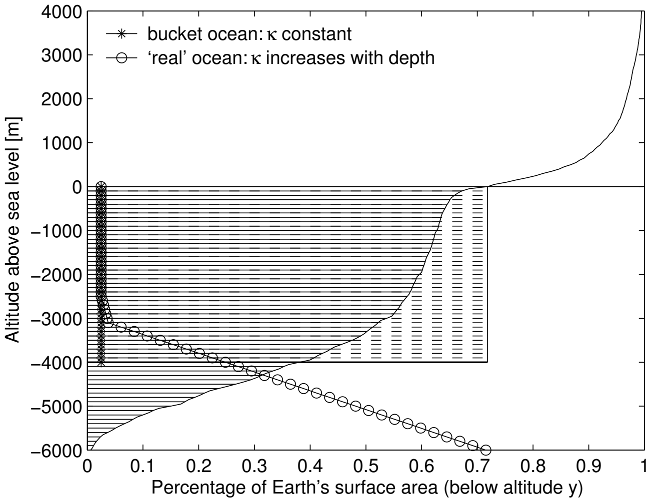


Figure B.1: Hypsometric diagram of the earth surface. Also shown are the "bucket" ocean approximation (depth 4000 m, 40 layers, uniform cross-section and diffusivity), and the "real" ocean approximation (depth 6000 m, 60 layers, depth-dependent cross-section and diffusivity). The diffusivity profiles are given in units of $40 \times \kappa(0)$.

gested as a vertical and global average in several studies (Munk, 1966; Simmons et al., 2004), the heat anomaly reaches the bottom after several hundred years (Fig. B.2.a). It will take $\tau_D = z_B^2/\kappa \approx 5000$ years, however, until the bottom water has warmed as much as the surface, and the ocean equilibrates. After 2000 years, for instance, the ocean bottom will have received only approx. 60% of the surface warming. These results, however, are very sensitive to the value of $\kappa$. When the diffusivity is doubled to a value of $\kappa = 2\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, the heat uptake proceeds much faster (Fig. B.2.b). In this case, the ocean is close to equilibrium after 2000 years.

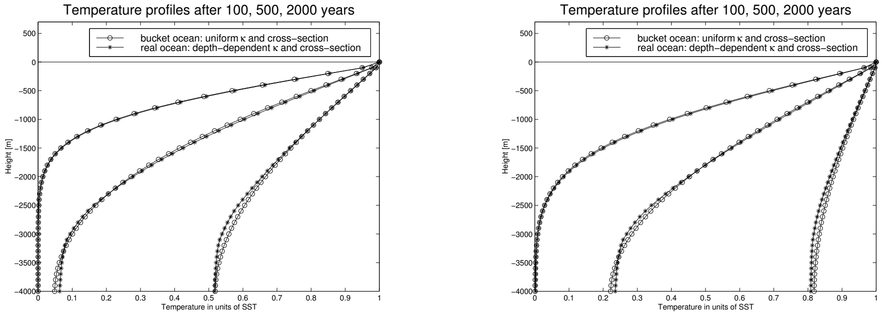
We have also investigated the sensitivity of the temperature profile, when we drop the simplifying assumption of a uniform diffusivity and uniform cross-sectional area of the ocean column with depth. We have solved Equation (B.7) for a "real" ocean approximation with an increasing diffusivity below 2500 m due to tidally driven mixing close to the ocean bottom (Simmons et al., 2004), and a depth-dependent cross-section of the water column down to $z_B = 6000\,\mathrm{m}$ (see Fig. B.1). Accounting for the shape of the ocean basin delays the warming at depths below 3000 m, since heat can now diffuse into the deep ocean that was not included in the idealized "bucket" ocean. This effect is only partially cancelled by heat accumulation due to the narrowing ocean basin profile. However, the increasing vertical diffusivity $\kappa$ at depths below 2500 m counteracts the delayed ocean warming by transporting heat much faster to the bottom. Therefore, the ocean temperature anomaly profiles for the "bucket" ocean and the "real" ocean approximation are in good agreement for the case of pure diffusion (Fig. B.2.a-b) as well as for upwelling-diffusion with no or little warming at the bottom from downwelling polar water (Fig. B.2.c-d and to a lesser degree also Fig. B.2.e-g). Only for a substantial heat source at the bottom, the two profiles deviate notedly from each other below 3000 m depth (Fig. B.2.f-h). This is due to the large diffusivities and upwelling velocities in the lower part of the "real" ocean approximation which transfer heat from the bottom to the interior much faster than in the "bucket" ocean.

When upwelling of water masses takes places, the downward diffusion of heat is partially offset. If no warming at the bottom from downwelling polar water is assumed (Fig. B.2.c: $w = 2\,\mathrm{m\,yr^{-1}}$, Fig. B.2.d: $w = 4\,\mathrm{m\,yr^{-1}}$), the heat anomaly is mainly constrained to the upper 2000 m of the ocean column (the larger $w$, the more constrained). Moreover, equilbrium is reached much faster than for pure diffusion, namely on the time scale $\tau_{UD} = 4\,\kappa/w^2$ for upwelling to balance diffusion , i.e. approx. 3160 years for $w = 2\,\mathrm{m\,yr^{-1}}$, $\kappa = 1\,\mathrm{cm^2\,s^{-1}}$ and approx. 790 years for $w = 4\,\mathrm{m\,yr^{-1}}$, $\kappa = 1\,\mathrm{cm^2\,s^{-1}}$ (Dickinson and Schaudt, 1998). As can be already deduced from the equilibrium solution (B.4) for the temperature anomaly profile, little or no heat is transferred to the bottom water at $z_B = 4000\,\mathrm{m}$ (the larger $w$, the less). Such a behaviour is clearly not to be seen in $CO_2$ doubling experiments with general circulation models (GCMs) (e.g. Raper et al., 2001). In these experiments, the warming anomaly penetrates to the bottom after a few hundred years of $CO_2$ stabilization, with continued warming thereafter. Hence, the assumption of bottom heating from downwelling polar water seems to be an important prerequisite to account for the simulated bottom warming in GCMs, but also for the overall heat uptake of the ocean.

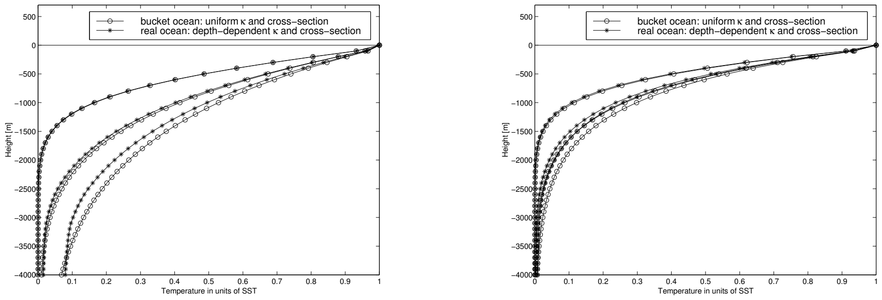
Fig. B.2.e and .g show the ocean temperature anomaly profile for incoming polar bottom water with $T_P = 0.2\,T_S$ and two different upwelling velocities, and Fig. B.2.f and .h the corresponding profiles for $T_P = 0.8\,T_S$. The "U-shaped" profiles reflect the fact that heat sources are now located at the surface as well as at the bottom. The shape will be less pronounced, if a gradual instead of an instantaneous temperature rise is assumed, but, nevertheless, will remain a characteristic of this type of ocean model formulation (Raper et al., 2001, Fig. 5 and 11a). Equilibrium is reached as fast as in the cases without warming from polar bottom water (the larger $w$, the faster), but exhibits now a bottom temperature anomaly $T(z_B) = T_P + e^{-w/\kappa\, z_B}$. Hence, the bottom warming can be tuned to reproduce GCM simulations. The typical "U-Shape" of the profiles, however, is not confirmed by these simulations (Raper et al., 2001). In this respect, the simple case of pure heat diffusion can better reproduce the ocean temperature anomaly that is seen in $CO_2$ doubling experiments with GCMs.

## B.2 Analytical solutions

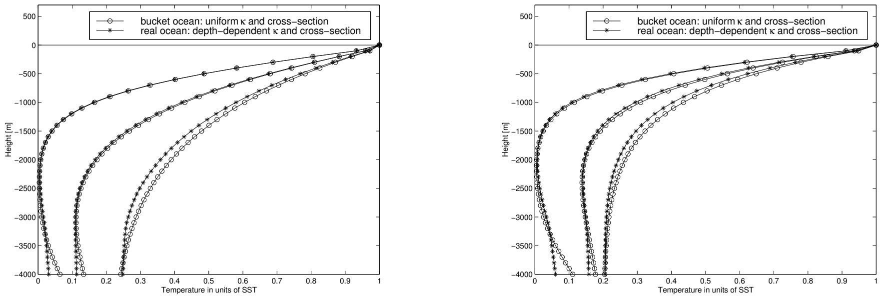
The preceding discussion has shown that an idealized "bucket" ocean model, as used in many models for the assessment of anthropogenic climate change, captures the anomalous heat fluxes well enough for being considered an interesting application. In the case of uniform diffusivity and cross-section of the ocean column, the upwelling-diffusion problem (B.5) for the anomalous heat fluxes simplifies to (with parameteriza-



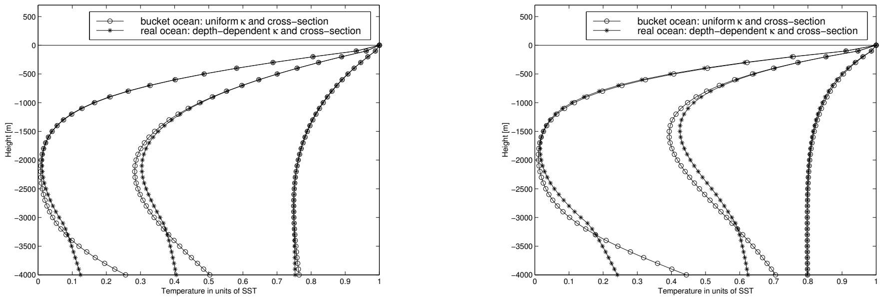
a. $\kappa = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=$0\,\mathrm{m}\,\mathrm{yr}^{-1}$ b. $\kappa = 2\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=$0\,\mathrm{m}\,\mathrm{yr}^{-1}$



c. $\kappa = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=$2\,\mathrm{m}\,\mathrm{yr}^{-1}$, $\Pi = 0$ d. $\kappa = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=$4\,\mathrm{m}\,\mathrm{yr}^{-1}$, $\Pi = 0$



e. $\kappa = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=$2\,\mathrm{m}\,\mathrm{yr}^{-1}$, $\Pi = 0.2$ f. $\kappa = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=$4\,\mathrm{m}\,\mathrm{yr}^{-1}$, $\Pi = 0.2$



g. $\kappa = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=$2\,\mathrm{m}\,\mathrm{yr}^{-1}$, $\Pi = 0.8$ h. $\kappa = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=$4\,\mathrm{m}\,\mathrm{yr}^{-1}$, $\Pi = 0.8$

Figure B.2: Numerical solutions for the ocean temperature anomaly profiles under different assumptions of vertical diffusivity $\kappa$, upwelling velocity $w$ and polar bottom water warming relative to surface warming $\Pi$. Shown are the solutions for an idealized “bucket” ocean and a “real” ocean approximation with depth-dependent diffusity and ocean cross-section profile.

tion $T_P = \Pi\, T_S$)

$$\begin{aligned} \text{for } 0 < z < z_B: \quad & T_t(z,t) = \kappa\, T_{zz}(z,t) + w\, T_z(z,t)\ , \\ \text{B.C.:} \quad & T(0,t) = T_S(t)\ , \quad F(z_B,t) = -c_V\, w\, \Pi\, T_S(t)\ , \\ \text{I.C.:} \quad & T(z,0) = 0\ , \end{aligned} \tag{B.8}$$

Problem (B.8) is accessible, at least for the case $z_B \to \infty$, to analytical solution methods. Analytical solutions are of particular interest for our application, since they will allow us to directly calculate the heat flux from the mixed layer into the interior ocean at $z = 0$, i.e.

$$F_O(t) = -c_V\, w\, T_S(t)\,(1-\Pi) - c_V\, \kappa\ T_z(z,t)\big|_{z=0}\ , \tag{B.9}$$

in terms of the mixed layer temperature $T_S(t)$. Knowledge of $F_O$ in terms of $T_S$ closes the energy balance model (A.8)-(A.9), and hence will save us to model the ocean column explicitely which would require the addition of a large number of ordinary differential equations to the model (see Equation B.6). The usage of analytical solutions of the heat diffusion equation in energy balance models of anthropogenic climate change was proposed by Wigley and Schlesinger (1985). Analytical solutions of the full upwelling-diffusion problem, albeit without an ocean mixed layer, were studied by Dickinson and Schaudt (1998).

Problem (B.8) can be solved, at least partly, by converting it to a form accessible to the Green's function approach. It is well known that the inhomogenuous heat diffusion equation

$$\begin{aligned} \text{for } -\infty < z < \infty: \quad & u_t - \kappa\, u_{zz} = \tilde{f}(z,t)\ , \\ \text{I.C.} \quad & u(z,0) = \tilde{\Phi}(z) \end{aligned} \tag{B.10}$$

on the real line can be solved by means of the *retarded Green's function* (e.g., Honerkamp and Römer, 1989, Chap. 10),

$$G(z,t) = H(t)\, D(z,t)\ , \quad \text{with heat diffusion kernel } \ D(z,t) = \frac{e^{-z^2/(4\kappa t)}}{2\sqrt{\pi\kappa t}}\ , \tag{B.11}$$

and Heavyside function $H(t)$. The retarded Green's function constitutes the temperature response to a $\delta$-heat pulse at $t = 0$ and $z = 0$, i.e.

$$G_t - \kappa\, G_{zz} = \delta(t)\, \delta(z)\ .$$

The solution to problem (B.10) has the form:

$$u(z,t) = \int_{-\infty}^{\infty} D(z-z',t)\, \tilde{\Phi}(z')\, dz' + \int_0^t \int_{-\infty}^{\infty} D(z-z',t-t')\, \tilde{f}(z',t')\, dz'\, dt' \tag{B.12}$$

The method of images allows to extend the Green's function formalism to the heat difffusion problem on the half line, i.e.,

$$\begin{aligned} \text{for } 0 < z < \infty: \quad & u_t - \kappa\, u_{zz} = f(z,t)\,, \\ \text{B.C.} \quad & u(0,t) = 0\,, \\ \text{I.C.} \quad & u(z,0) = \Phi(z)\,. \end{aligned} \tag{B.13}$$

Its basic idea is to extend the initial and boundary condition onto the domain $-\infty < x < 0$, so that they constitute odd functions on the real line. Define $\tilde{\Phi}(z) = \Phi(z)$ on $0 < z < +\infty$, $\tilde{\Phi}(0) = 0$ and $\tilde{\Phi}(z) = -\Phi(-z)$ on $-\infty < z < 0$, and similarly $\tilde{f}(z,t) = f(z,t)$ on $0 < z < +\infty$, $\tilde{f}(0,t) = 0$ and $\tilde{f}(z) = -f(-z,t)$ on $-\infty < z < 0$. Then, the solution (B.12) of the real line problem (B.10), including a solution of the partial differential equation on the half line, is also an odd function with $u(-z,t) = -u(z,t)$. As a result, it is $u(0,t) = 0$, and $u(z,t)$ fulfills the boundary condition at $z = 0$. Moreover, $u(z,0) = \tilde{\Phi}(z) = \Phi(z)$ on $0 < z < +\infty$, and the initial condition is fulfilled. Hence, $u(z,t)$ constitutes the solution to Problem (B.13). Since $\tilde{\Phi}(z)$ and $\tilde{f}(z,t)$ are odd functions, they can be rewritten on the half line $0 < z < +\infty$ as

$$\begin{aligned} u(z,t) = \int_0^\infty & \big(D(z-z',t) - D(z+z',t)\big)\ \Phi(z')\, dz' + \\ & \int_0^t \int_0^\infty \big(D(z-z',t-t') - D(z+z',t-t')\big)\ f(z',t')\, dz'\, dt'\,. \end{aligned} \tag{B.14}$$

The task is now to transform the initial problem (B.8) for the idealized case of an infinitely deep ocean with $z_B \rightarrow +\infty$ into a problem of the type (B.13). To simplify notation, we non-dimensionalize the upwelling-diffusion equation by indexing depth $z^* := z/\lambda$ in units of its characteristic length scale $\lambda = \kappa/w$, and time $t^* := t/\tau$ in units of its characteristic time scale $\tau = \kappa/w^2$. Then, the upwelling-diffusion problem without bottom reads

$$\begin{aligned} \text{for } 0 < z < +\infty: \quad & T_{t^*}(z^*,t^*) = T_{z^*z^*}(z^*,t^*) + T_{z^*}(z^*,t^*)\,, \\ \text{B.C.:} \quad & T(0,t^*) = T_S(t^*)\,, \\ \text{I.C.:} \quad & T(z^*,0) = 0\,, \end{aligned} \tag{B.15}$$

Problem (B.15) can be transformed by the ansatz $T(z^*,t^*) = u(z^*,t^*)\, e^{-z^*/2 - t^*/4} + T_S(t^*)$. The resulting problem for $u(z^*,t^*)$ is indeed of the type (B.13), namely

$$\begin{aligned} \text{for } 0 < z < +\infty: \quad & u_{t^*} - u_{z^*z^*} = -T_{S,t^*}(t^*)\, e^{z^*/2 + t^*/4}\,, \\ \text{B.C.} \quad & u(0,t^*) = 0\,, \\ \text{I.C.} \quad & u(z^*,0) = -T_S(0)\, e^{z^*/2} := 0\,. \end{aligned} \tag{B.16}$$

Inserting Solution (B.14) for $u(z^*,t^*)$ into the ansatz for $T(z^*,t^*)$, Problem (B.15) is solved by

$$\begin{aligned}
T(z^*,t^*) &= T_S(t^*) - \int_0^{t^*}\int_0^{+\infty} T_{S,t'}(t')\, e^{-(z^*-z')/2-(t^*-t')/4} \\
&\qquad \times \left( \frac{e^{-0.25\,(z^*-z')^2/(t^*-t')}}{2\sqrt{\pi(t^*-t'}} - \frac{e^{-0.25\,(z^*+z')^2/(t^*-t')}}{2\sqrt{\pi(t^*-t')}} \right) dz'\, dt' \\
&= T_S(t^*) - \int_0^{t^*}\int_0^{+\infty} T_{S,t'}(t') \\
&\times \left( \frac{e^{-0.25\,(z^*-z'+t^*-t')^2/(t^*-t')}}{2\sqrt{\pi(t^*-t')}} - \frac{e^{-0.25\,(z^*+z'-t^*+t')^2/(t^*-t')}}{2\sqrt{\pi(t^*-t')}}\, e^{-z^*} \right) dz'\, dt'
\end{aligned}$$

The integration over depth $z$ can be performed analytically:

$$\begin{aligned}
\int_0^{+\infty} \frac{e^{-0.25\,(z^*-z'+t^*-t')^2/(t^*-t')}}{2\sqrt{\pi(t^*-t')}}\, dz' &= 0.5 + 0.5\,\mathrm{Erf}\left(\frac{z^*+t^*-t'}{2\sqrt{t^*-t'}}\right), \\
\int_0^{+\infty} \frac{e^{-0.25\,(z^*+z'-t^*+t')^2/(t^*-t')}}{2\sqrt{\pi(t^*-t')}}\, dz' &= 0.5 - 0.5\,\mathrm{Erf}\left(\frac{z^*-t^*+t'}{2\sqrt{t^*-t'}}\right).
\end{aligned}$$

Hence, the solution to Problem (B.15) is given by

$$\begin{aligned}
T(z^*,t^*) = 0.5 &\left( T_S(t^*) - \int_0^{t^*} T_{S,t'}(t')\,\mathrm{Erf}\left(\frac{z^*+t^*-t'}{2\sqrt{t^*-t'}}\right) dt' \right) \\
&+ 0.5\, e^{-z^*} \left( T_S(t^*) - \int_0^{t^*} T_{S,t'}(t')\,\mathrm{Erf}\left(\frac{z^*-t^*+t'}{2\sqrt{t^*-t'}}\right) dt' \right) \qquad \text{(B.17)}
\end{aligned}$$

Expression (B.17) yields the temperature anomaly profile in the interior ocean at any time $t$ for an arbitrary warming trajectory $T_S(t)$ of the mixed layer on top. It can be used to calculate the heat flux (B.9) into the interior ocean for the case of an infinitely deep ocean without polar bottom water ($\Pi = 0$):

$$\begin{aligned}
F_O(t) = c_V \sqrt{\frac{\kappa}{\pi}} \int_0^t T_{S,t'}(t')\, \frac{e^{-\frac{w^2(t-t')}{4\kappa}}}{\sqrt{t-t'}}\, dt' \\
- c_V\, \frac{w}{2} \left( T_S(t) - \int_0^t T_{S,t'}(t')\,\mathrm{Erf}\left(\frac{w\sqrt{t-t'}}{2\sqrt{\kappa}}\right) dt' \right), \qquad \text{(B.18)}
\end{aligned}$$

where we have reintroduced the dimension of time $t$ and depth $z$. The analytical solution (B.18) allows us to close the energy balance model (A.8)-(A.9) with an integro-differential equation for the mixed layer temperature.

After the mixed layer temperature has been stabilized at some value $T_S^*$, Solutions (B.17) and (B.18) approach the new equilibrium asymptotically, i.e.,

$$\lim_{t\to+\infty} T(z,t) = T_S^*\, e^{-\frac{w}{k}\, z}\,, \qquad \lim_{t\to+\infty} F_O(t) = 0\,.$$

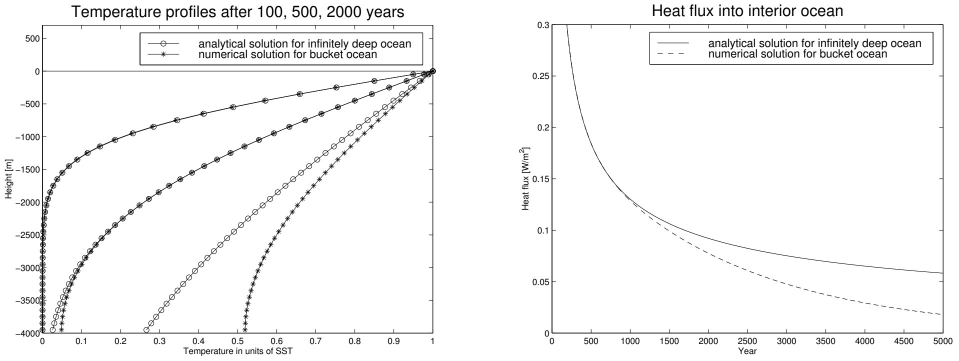
This behaviour is a consequence of assuming an infinitely deep ocean in the underlying problem (B.15). If the ocean has finite depth, the new equilibrium will be reached in finite time. To compare the transient behavior of an infinitely deep ocean with the behavior of a more realistic "bucket" ocean with depth $z_B = 4000\,\mathrm{m}$, we compare the analytical solutions (B.17) and (B.18) with the numerical solution to Equation (B.7) for the idealized boundary condition of an instantaneous mixed layer warming by one temperature unit at $t = 0$. Hence, $T_S(t) = H(t)$ is specified by the Heavyside-function $H$, and the time derivative $T_{S,t}(t) = \delta(t)$ equals the Dirac $\delta$-function (Honerkamp and Römer, 1989, Appendix E). For this boundary condition, Expressions (B.17) and (B.18) greatly simplify to

$$T(z,t) \;=\; 0.5\left(1 - \mathrm{Erf}\left(\frac{z+wt}{2\sqrt{\kappa t}}\right)\right) + 0.5\, e^{-\frac{w}{\kappa}\, z}\left(1 - \mathrm{Erf}\left(\frac{z-wt}{2\sqrt{\kappa t}}\right)\right)\,, \tag{B.19}$$

$$F_O(t) \;=\; c_V\sqrt{\frac{\kappa}{\pi}}\,\frac{e^{-\frac{w^2 t}{4\kappa}}}{\sqrt{t}} - c_V\,\frac{w}{2}\left(1 - \mathrm{Erf}\left(\frac{w\sqrt{t}}{2\sqrt{\kappa}}\right)\right)\,. \tag{B.20}$$

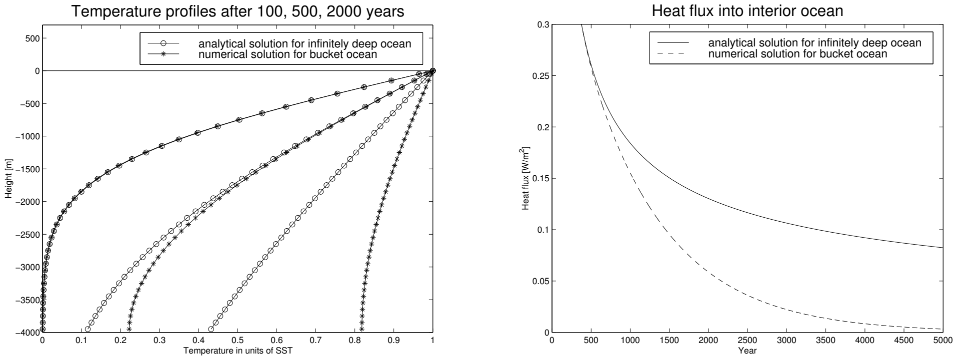
Fig. B.3 shows the ocean temperature anomaly profiles after 100, 500, and 2000 years of the instantaneous warming (Fig. B.3.a, .c, .e), and the heat flux into the interior ocean (Fig. B.3.b, .d, .f) for different combinations of effective heat diffusivity $\kappa$ and upwelling velocity $w$. Analytical and numerical solutions deviate increasingly with increasing depth. This has to be expected, since the "bucket" ocean stores the heat penetrating to the ocean floor in the bottom layers, while an infinitely deep ocean allows it to escape to ever greater depths. Moreover, the deviations grow in time until the "bucket" ocean warming slows down upon approaching a new equilibrium. Most strikingly, the deviations decrease with increasing upwelling velocity. This is due to the fact that the upwelling induces an upward heat flux counteracting the downward heat diffusion, so that the net heat flux at the bottom is significantly reduced. Hence, the error made by the analytical expression by assuming that the heat flux at the bottom can escape to infinite depths is reduced as well. By the same token, the error grows substantially with increasing strength of downward diffusion as captured by the effective vertical diffusivity $\kappa$.

We have discussed above that a pure diffusion ocean without upwelling yields the most realistic picture of the ocean temperature anomaly profile in comparison with GCM experiments. Therefore, we are particularly interested in this special case. For $w = 0\,\mathrm{m\,yr^{-1}}$, the analytical solutions (B.17) for the temperature profile and (B.18) for the heat flux into the interior ocean simplify considerably:



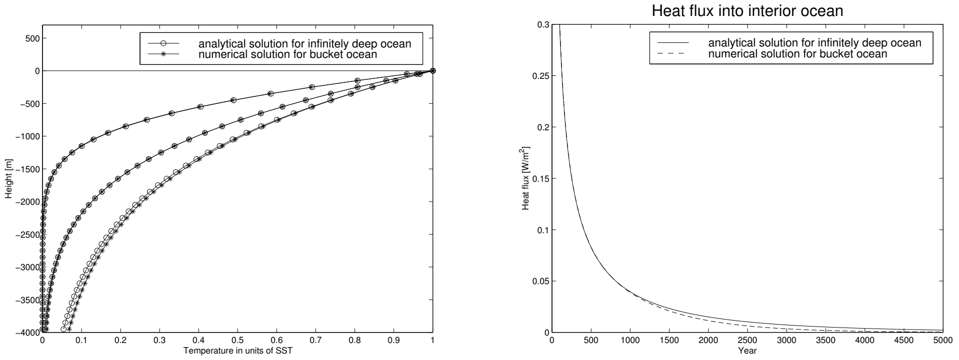
a. $\kappa = 1\,\mathrm{cm^2\,s^{-1}}$, w=$0\,\mathrm{m\,yr^{-1}}$

b. $\kappa = 1\,\mathrm{cm^2\,s^{-1}}$, w=$0\,\mathrm{m\,yr^{-1}}$



c. $\kappa = 2\,\mathrm{cm^2\,s^{-1}}$, w=$0\,\mathrm{m\,yr^{-1}}$

d. $\kappa = 2\,\mathrm{cm^2\,s^{-1}}$, w=$0\,\mathrm{m\,yr^{-1}}$



e. $\kappa = 1\,\mathrm{cm^2\,s^{-1}}$, w=$2\,\mathrm{m\,yr^{-1}}$, $\Pi = 0$

f. $\kappa = 1\,\mathrm{cm^2\,s^{-1}}$, w=$2\,\mathrm{m\,yr^{-1}}$, $\Pi = 0$

Figure B.3: Analytical solutions for the ocean temperature anomaly profiles (left panels) and the heat flux into the interior ocean (right panels) compared to the numerical "bucket" ocean solution.

$$T(z,t) = T_S(t) - \int_0^t T_{S,t'}(t')\,\mathrm{Erf}\left(\frac{z}{2\sqrt{\kappa(t-t')}}\right)dt' \,, \tag{B.21}$$

$$F_O(t) = c_V\sqrt{\frac{\kappa}{\pi}}\int_0^t \frac{T_{S,t'}(t')}{\sqrt{t-t'}}\,dt' \,. \tag{B.22}$$

We could plug Expression (B.22) into the differential equation (A.9) for the mixed layer ocean temperature in order to close our energy balance model (A.8)-(A.9). However, we have just established that the idealised assumption of an infinitely deep ocean leads to particularly large deviations from the behaviour of a finitely deep ocean in the case of pure diffusion. It can be seen from Fig. B.3.b,d that the heat fluxes into the interior ocean begin to deviate from each other after approx. 1300 years for $\kappa = 1\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, and approx. 550 years for $\kappa = 2\,\mathrm{cm}^2\,\mathrm{s}^{-1}$. Since we want to use the energy balance model for assessing global mean temperature changes since the preindustrial period up to the end of the 21st century, i.e. approx. 300 years, such late deviations do not matter here. In the context of the uncertainty analysis, however, we will consider vertical diffusivities up to $k = 10\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, for which analytical (infinite depth) and numerical (finite depth) solutions can deviate already after approx. 100 years.

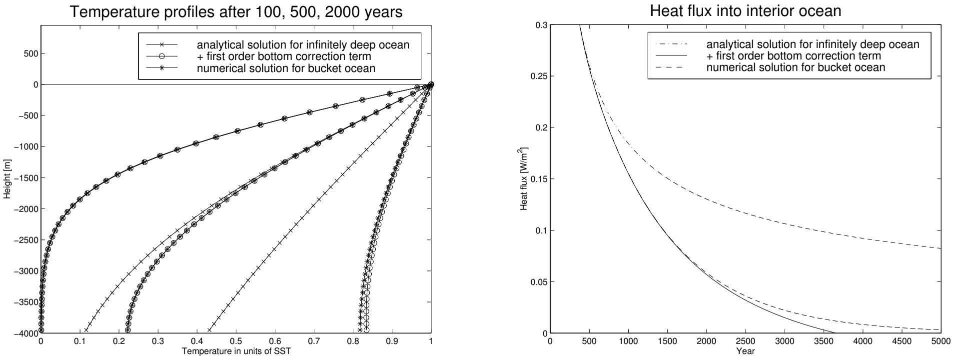
Therefore, it is desirable to correct the analytical solutions (B.21) and (B.22) for the finite depth of the pure diffusion ocean. To put it more precisely, we are looking for the solution to the problem

$$\begin{aligned} \text{for } 0 < z < z_B: \quad & T_t(z,t) = \kappa\, T_{zz}(z,t)\ , \\ \text{B.C.:} \quad & T(0,t) = T_S(t)\ , \quad T_z(z_B,t) = 0\ , \\ \text{I.C.:} \quad & T(z,0) = 0\ , \end{aligned} \tag{B.23}$$

with Dirichlet boundary condition at the top of the water column, and Neumann boundary condition at its bottom. The latter boundary condition simply requires that no heat flows into or out of the ocean floor. An analytical solution of Problem (B.23) can be obtained by superposition of an infinite series of solutions on the half line. Consider Solution (B.21) for an infinitely deep diffusive ocean with a heat source of temperature $T_S(t)$ at the top of the water column ($z = 0$). It entails a heat flux

$$F_O(z_B,t) := -c_V\,\kappa\ T_z(z,t)\big|_{z=z_B} \;=\; c_V\,\sqrt{\frac{\kappa}{\pi}}\int_0^t \frac{T_{S,t'}(t')}{\sqrt{t-t'}}\,e^{-\frac{z_B^2}{4\kappa(t-t')}} \tag{B.24}$$

into the ocean bottom at $z = z_B$. This heat flux can be cancelled by positioning an image of the heat source with temperature $T_S(t)$ at depth $z = 2z_B$. The imaginary heat source below the ocean floor emulates the larger warming of the bottom layer due to the fact that no heat can penetrate the bottom. It induces, however, a larger warming at the mixed-layer boundary so that the upper Dirichlet boundary condition is violated. The additional warming at $z = 0$ is also produced by an additional heat source with temperature $T_S(t)$ at $z = -2z_B$ above the surface. Hence, we can substract the contribution of such a third heat source from the warming induced by the first two heat sources to restore the boundary condition at $z = 0$. Substracting a heat source is equivalent to adding a heat sink, which leads to a cooling of the water column. However, now we are again in violation of the Neumann boundary condition at $z = z_B$, since the cooling induces a heat flux from the ocean floor. But this heat flux is much



a. $\kappa = 2\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=0 m yr$^{-1}$ b. $\kappa = 2\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, w=0 m yr$^{-1}$

Figure B.4: Analytical solution with and without bottom correction term compared to the numerical “bucket” ocean solution.

smaller in magnitude than the heat flux introduced by the original heat source at $z = 0$, which can be seen by replacing $z_B$ in Expression (B.24) with $3z_B$, i.e. the distance from the imaginary heat sink at $z = -2z_B$ to the ocean floor. Therefore, we have gained something from introducing an imaginary heat source below the bottom and an imaginary heat sink above the surface. To cancel the remaining heat flux from the ocean floor, we can repeat the procedure by introducing another imaginary heat sink at $z = 4z_B$. The resulting additional cooling of the mixed-layer boundary can be eliminated by an imaginary heat source at $z = -4z_B$. By introducing heat sources and sinks below the bottom and above the surface infinetely many times to restore the boundary coundations, we arrive at the following exact solution to Problem (B.23):

$$\begin{aligned} T(z,t) &= T_S(t) - \int_0^t T_{S,t'}(t')\,\mathrm{Erf}\left(\frac{z}{2\sqrt{\kappa(t-t')}}\right) dt' \qquad \text{(B.25)} \\ &\quad + \sum_{n=1}^{+\infty} (-1)^n \int_0^t T_{S,t'}(t') \left(\mathrm{Erf}\left(\frac{2nz_B - z}{2\sqrt{\kappa(t-t')}}\right) - \mathrm{Erf}\left(\frac{2nz_B + z}{2\sqrt{\kappa(t-t')}}\right)\right) dt' \end{aligned}$$

$$F_O(t) = c_V \sqrt{\frac{\kappa}{\pi}} \int_0^t \frac{T_{S,t'}(t')}{\sqrt{t-t'}} \left(1 + 2\sum_{n=1}^{+\infty} (-1)^n\, e^{-\frac{n^2 z_B^2}{\kappa(t-t')}}\right) dt' \qquad \text{(B.26)}$$

For an ocean depth of $z_B = 4000\,\mathrm{m}$, Series (B.25) and (B.26) converge very fast. Fig. (B.4) illustrates the enormous improvement that can be achieved by considering only the first order bottom correction term. The resulting estimates for the ocean temperature anomaly profiles are now in good accordance with the numerical “bucket” ocean solution, and the analytically and numerically derived heat fluxes begin to deviate from each other only after 2500 years.

# Appendix C

# A Survey of Imprecise Probability Theory

## C.1 Basic concepts of imprecise probability

A fairly general theory of imprecise probability can be based on *convex sets of probabilities* (Good, 1962; Levi, 1980), or *coherent lower previsions* (Walley, 1991)[1]. For the reader used to Kolmogorovian probability on event spaces, convex sets of probabilities are intuitively easiest to access. Consider a measurable space $(\Omega, \mathcal{A})$ with universal set $\Omega$ and a field (or $\sigma$-field) $\mathcal{A}$ of subsets of $\Omega$[2]. The field $\mathcal{A}$ constitutes the *space of events*, on which a probability measure is defined. We recall the well-known axioms of probability (Kolmogorov 1933; deviating from Kolmogorov, we also include the case of finitely additive probability below).

**Definition C.1** *A (finitely or countably) additive* ***probability measure*** *$P$ on $(\Omega, \mathcal{A})$ is a set function $P : \mathcal{A} \to [0, 1]$ with*

1. *$P(A) \geq 0$ for all $A \in \mathcal{A}$,*
2. *$P(\Omega) = 1$, and*

---

[1] Walley (2000) points out that theories based on sets of desirable gambles (Walley, 1991) and partial preference orderings (Giron and Rios, 1980; Seidenfeld et al., 1995) lead to an even more general uncertainty model than can be derived from convex sets of probabilities and lower previsions. The more general theory allows to condition on events with probability zero, and to distinguish preference between gambles whose outcomes agree almost everywhere, i.e., except on a subset with measure zero. Like classical probability theory, models based on convex sets of probabilities and lower previsions cannot capture such boundary cases adequately. A discussion of theories based on sets of desirable gambles and partial preference orderings is beyond the scope of this thesis.

[2] A collection of subsets of $\Omega$ is called a field, or algebra, if it contains $\Omega$ itself and is closed under the formation of complement and finite union. A field is called a $\sigma$-field, or $\sigma$-algebra if it is closed not only under the formation of finite unions, but also of countably infinite unions (Billingsley, 1995). Finitely additive probabilities are defined on fields, while countably additive probabilities are defined on $\sigma$-fields. Hence, the type of field constituting a measurable space depends on the type of additivity of the measure.

3. *$P(\cup_i A_i) = \sum_i P(A_i)$ for any (finite or countably infinite) collection of disjoint sets $A_i \in \mathcal{A}$.*

These axioms suffice to derive the well-known properties that the probability of an event constitutes a value in the unit interval $[0, 1]$, $P(\emptyset) = 0$, and the probability over every partition of the universal set $\Omega$ adds to unity.

**Definition C.2** *A **convex set of probabilities**, $\mathcal{M}$, is constituted by an arbitrary collection of (finitely or countably) additive probability measures $P : \mathcal{A} \to [0, 1]$ with the property that for any two probabilities $P, Q \in \mathcal{M}$ and an arbitrary $\lambda \in (0, 1)$, also the mixture probability $R = \lambda P + (1 - \lambda)Q$ (pointwise addition on the field $\mathcal{A}$) is contained in $\mathcal{M}$.*

If a collection $\mathcal{C}$ of probability measures is not convex, it is always possible to construct its *convex hull* $\mathcal{M} = \mathrm{co}(\mathcal{C})$ by adding all possible mixture probabilities to it. The requirement of convexity is imposed for mathematical convenience. It does not impose an additional restriction, since a collection $\mathcal{C}$ has the same behavioural implication in a decision making context than its convex hull (as long as the decision maker's values for the outcomes are determinate). Sets of probabilities have also been called *credal sets* in the literature (Levi, 1980).

The concept of *lower prevision* stands in the tradition of the theory of linear previsions by de Finetti (1937). While probabilities are defined on event spaces, previsions exist on the space of real-valued *gambles* $X : \Omega \to \mathbb{R}$. In a behavioural context, gambles are associated with acts under uncertainty linking each uncertain state of the world $\omega \in \Omega$ with the utility $X(\omega)$ from performing action $X$ in state $\omega$. In the frequentist context, gambles constitute random variables. A gamble is called $\mathcal{A}$-measurable if the events $\{\omega : X(\omega) < x\}$ and $\{\omega : X(\omega) > x\}$ are contained in the field $\mathcal{A}$ for all $x \in \mathbb{R}$. Let $\mathcal{K}(\mathcal{A})$ be the (function) space of all $\mathcal{A}$-measurable gambles that are bounded, i.e., for which there exist real numbers $x_1$ and $x_2$ with $x_1 < X < x_2$. This space contains the subset of *indicator gambles* $I_A$ for all measurable events $A \in \mathcal{A}$, with $I_A(\omega) = 1$ if $\omega \in A$, and $I_A(\omega) = 0$ otherwise. In the context of game theory, the indicator gambles are called *unanimity games*.

**Definition C.3** *A **linear prevision** $Pr$ on $\mathcal{K}(\mathcal{A})$ is a functional $Pr : \mathcal{K}(\mathcal{A}) \to \mathbb{R}$, which fulfils*

1. *$Pr(X + Y) = Pr(X) + Pr(Y)$ for any two gambles $X, Y \in \mathcal{K}(\mathcal{A})$, and*
2. *$Pr(X) \geq \inf_{\omega\in\Omega} X(\omega)$ for all gambles $X \in \mathcal{K}(\mathcal{A})$.*

The two conditions imply $Pr(\lambda X) = \lambda Pr(X)$ for all $X \in \mathcal{A}$ and $\lambda \in \mathbb{R}$ (de Finetti, 1974, Section 3.1.5). In particular, $Pr$ is self-conjugate, i.e., $Pr(-X) = -Pr(X)$. It can be shown that the value of a linear prevision for a gamble $X$ is nothing else than the *expectation value* $E(X) := \int_{-\infty}^{\infty} x dF_{P,X}(x)$ for some probability $P : \mathcal{A} \to [0, 1]$ on the event space $\mathcal{A}$, where $F_{P,X}$ is the distribution function of $X$ under $P$ defined by

$F_{P,X}(x) := P(\{\omega : X(\omega) \leq x\})$ (de Finetti, 1974). In particular, it is $Pr(I_A) = P(A)$ for the indicator gambles $I_A$. Hence, linear previsions on the space of $\mathcal{A}$-measurable gambles, and additive probability measures on the space of events are in one-to-one correspondence. Knowledge of $Pr$ uniquely determines $P$ (via the indicator gambles), and knowledge of $P$ uniquely determines $Pr$ (via the expectation operation).

Since linear previsions and additive probabilities contain the same information, it is natural to try to base a theory of imprecise probabilities on a generalisation of linear previsions. The resulting theory is presented in Walley (1991).

**Definition C.4** *A* ***coherent lower prevision*** $\underline{Pr}$ *on a linear space of bounded gambles*[3]*, $\mathcal{K}$, is a functional $\underline{Pr} : \mathcal{K} \to \mathbb{R}$, which fulfils*

1. $\underline{Pr}(X + Y) \geq \underline{Pr}(X) + \underline{Pr}(Y)$ *for any two gambles $X, Y \in \mathcal{K}$ (superadditivity),*
2. $\underline{Pr}(\lambda X) = \lambda \underline{Pr}(X)$ *for $\lambda > 0$ and $X \in \mathcal{K}$, and*
3. $\underline{Pr}(X) \geq \inf_{\omega \in \Omega} X(\omega)$ *for all gambles $X \in \mathcal{K}$.*

It is evident that lower previsions generalize de Finetti's concept of prevision by dropping the linearity condition. The conjugate upper prevision is defined by $\overline{Pr}(X) := -\underline{Pr}(-X)$. Due to superadditivity of $\underline{Pr}$, it is $\underline{Pr}(-X) + \underline{Pr}(X) \leq 0$, and therefore $\overline{Pr}(X) \geq \underline{Pr}(X)$. Moreover, since $\underline{Pr}$ is superadditive, $\overline{Pr}$ is subadditive. It can be shown that a lower (upper) prevision is self-conjugate, i.e., $\underline{Pr}(X) = -\underline{Pr}(-X)$ for all $X \in \mathcal{K}$, if and only if it is a linear prevision (Walley, 1991, Section 2.8). If the lower prevision is strictly smaller than the conjugate upper prevision for some gamble $X$, the captured information contains imprecision. Obviously, a theory of imprecise probability can be equally build on coherent lower or upper previsions.

It was the achievement of Peter Walley to provide a simple behavioural foundation for coherent lower previsions in terms of *axioms for desirability* of gambles (Walley, 1991, Section 2.2.3). The behavioural foundation transfers to convex sets of probabilities, because they are in one-to-one correspondence to coherent lower previsions - just like additive probabilities and linear previsions. To see this let us note that every lower prevision defines a closed convex set $\mathcal{PR}(\underline{Pr})$ of linear previsions by

$$\mathcal{PR}(\underline{Pr}) := \{\, Pr : \ \forall\, X \in \mathcal{K} \quad \underline{Pr}(X) \leq Pr(X)\,\}$$

It can be shown that $\underline{Pr}$ is the *lower envelope* of $\mathcal{PR}(\underline{Pr})$, i.e.,

$$\forall\, X \in \mathcal{K} \quad \underline{Pr}(X) = \inf_{Pr \in \mathcal{PR}(\underline{Pr})} Pr(X)$$

if and only if it is coherent, i.e., if and only if it fulfils Definition C.4 (Walley, 1991, Theorem 3.3.3). Moreover, every closed convex set $\mathcal{PR}$ of linear previsions equals

---

[3] Walley (1991, Section 2.5) provides a general definition of coherent lower prevision for arbitrary spaces of gambles $X : \Omega \to \mathbb{R}$. For the purpose of this exposition, it suffices to consider the simplest case of linear subspaces ($X \in \mathcal{K}$, $Y \in \mathcal{K}$, $\lambda \in \mathbb{R}$, then $\lambda X \in \mathcal{K}$ and $X + Y \in \mathcal{K}$). The space $\mathcal{K}(\mathcal{A})$ of $\mathcal{A}$-measurable gambles is linear, if $\mathcal{A}$ is a $\sigma$-field.

a set $\mathcal{PR}(\underline{Pr})$ for some coherent lower prevision $\underline{Pr}$, and therefore can be uniquely represented by $\underline{Pr}$ (Walley, 1991, Theorem 3.6.1). Since linear previsions determine additive probabilities and vice versa, every closed convex set $\mathcal{PR}$ of linear previsions corresponds to a closed convex set $\mathcal{M}$ of probabilities, and therefore the representation in terms of $\underline{Pr}$ extends to $\mathcal{M}$.

To illustrate the introduced concepts and the further concepts of more specialized versions of imprecise probability theory, we present the nice football example of Walley (1991, 1996b). A football game has three possible outcomes for the home team, W(in), (D)raw, and L(oss), which constitute the universal set $\Omega_f = \{W, D, L\}$ (the subscript $f$ stands for football). The set of all possible probability distributions for the outcome of the game is a two-dimensional simplex in the unit cube $[0,1]^3$, defined by $p(W) + p(D) + p(L) = 1$.

We now ask a gambler to put his stakes on a finite set of gambles about the outcome of the game. Let $I_A$ be the indicator gamble of the event $A \subseteq \Omega_f$. The gambler indicates that she (marginally) desires the following three gambles:

1. $X_1 = I_{D,L} - I_W$, indicating that she (marginally) believes the home team not to win rather than to win. Hence, $\underline{Pr}_f(I_{D,L} - I_W) = 0$.

2. $X_2 = I_W - I_D$, indicating that she (marginally) believes the home team to win rather than to play a draw. Hence, $\underline{Pr}_f(I_W - I_D) = 0$.

3. $X_3 = I_D - I_L$, indicating that she (marginally) believes the home team to draw a match rather than to lose. Hence, $\underline{Pr}_f(I_D - I_L) = 0$.

The choice of (marginally) desirable gambles specifies a set of three constraints for the gambler's coherent lower prevision $\underline{Pr}_f$ on the space $\mathcal{K}(\mathcal{P}(\Omega_f))$ of all possible gambles. In the absence of any further statements, this is all we know about the gambler's assessment of the outcome of the football game. The corresponding closed convex set $\mathcal{PR}(\underline{Pr}_f)$ contains all linear previsions which fulfil $Pr_f(I_{D,L}) \geq Pr_f(I_W)$, $Pr_f(I_W) \geq Pr_f(I_D)$, and $Pr_f(I_D) \geq Pr_f(I_L)$ simultaneously. This corresponds to a closed convex set of probabilities $\mathcal{M}(\underline{Pr}_f)$ including all probability measures with $P(\{D, L\}) \geq P(\{W\})$, $P(\{W\}) \geq P(\{D\})$, and $P(\{D\}) \geq P(\{L\})$.

Since the coherent lower prevision $\underline{Pr}_f$ is the lower envelope of the convex set $\mathcal{PR}(\underline{Pr}_f)$ of linear previsions, we can use $\mathcal{PR}(\underline{Pr}_f)$ to find the value of $\underline{Pr}_f$ for gambles $X \in \mathcal{K}(\mathcal{P}(\Omega_f))$ that were not considered by the gambler. Consider, e.g., the gamble $I_{W,D}$. It is $\underline{Pr}_f(I_{W,D}) = \inf_{p \in \mathcal{M}(\underline{Pr}_f)} \left( p(W) + p(D) \right) = \frac{2}{3}$. Hence, the gambler's assessment implied that she believes the home team not to lose rather than to lose. In this way, an assessment for a finite set of gambles, which will be everything what is available in practice, implies a unique extension onto the entire space $\mathcal{K}(\mathcal{P}(\Omega_f))$ of possible gambles.

**Definition C.5** *The **natural extension** of a lower prevision $\underline{Pr}$ specified on some (possibly finite) set $\mathcal{K}$ of gambles is the minimal coherent lower prevision $\underline{E}$ on the space* *of all* $\mathcal{A}$*-measurable gambles* $\mathcal{K}(A)$ *that dominates* $\underline{Pr}$ *on* $\mathcal{K}$*, i.e.,* $\forall\ X \in \mathcal{K} \quad \underline{E}(X) \geq \underline{Pr}(X)$*. "Minimal" means in this context that whenever some other coherent lower prevision* $\underline{Pr}^{\dagger}$ *dominates* $\underline{Pr}$ *on* $\mathcal{K}$*, it also dominates* $\underline{E}$ *on the entire space* $\mathcal{K}(A)$*.*

The natural extension will always exist if the lower prevision $\underline{Pr}$ is dominated by at least one linear prevision on $\mathcal{K}$. If this is not the case, $\mathcal{PR}(\underline{Pr}) = \emptyset$, and $\mathcal{M}(\underline{Pr}) = \emptyset$. In this situation, $\underline{Pr}$ does not *avoid sure loss*, i.e., the marginally desirable gambles $X - \underline{Pr}(X)$ in $\mathcal{K}$ can be combined in such a way that the gambler always incurs a loss (Walley, 1991, Section 2.4). The revealed belief underlying the choice of lower previsions is unreasonable.

The concept of natural extension is a cornerstone of Walley's theory of coherent lower previsions. Most problems of statistical reasoning, including conditioning, updating and combining information can be framed in terms of natural extension. Consider, for example, the set of all possible values of climate sensitivity, e.g., $\Omega_{T_{2x}} = [0.5\,\mathrm{K}, 10\,\mathrm{K}]$. Assume that there are assessments available for the (lower or precise) probability $p_i$ of the event that $T_{2x} \in A_i$ for a finite number of subsets $A_i \subset \Omega_{T_{2x}}$. For the sake of this argument, let us assume that the absolute amount of global mean temperature increase in the 21st century depends solely on climate sensitivity and the range $\Omega_Q$ of future scenarios for the anthropogenic interference with the climate system. We are now asking the question: Given those assessments about climate sensitivity, what is the lower probability that the warming in the year 2100 lies in the range $[1.4\,\mathrm{K}, 5.8\,\mathrm{K}]$, which was specified by the IPCC in its Third Assessment Report? This is a problem of natural extension. To show this, let us convert the assessments about climate sensitivity into marginally desirable gambles $I_{A_i} - p_i$ with $\underline{Pr}(I_{A_i}) = p_i$. On the basis of some reputed climate model, we can try to find the set $B \subset \Omega_{T_{2x}} \times \Omega_Q$, for which the warming in the year 2100, indeed, lies in the range $[1.8\,\mathrm{K}, 5.6\,\mathrm{K}]$. Knowledge of the set $B$ enables us to construct the IPCC-gamble $X_{IPCC} = I_B$ which pays a good reputation (1 "reputation point") if the IPCC projection included the realised warming, and a bad reputation otherwise (0 "reputation points"). Hence, we need to find the natural extension from the marginal desirable gambles $I_{A_i} - p_i$ onto the lower prevision $\underline{Pr}(X_{IPCC})$, which specifies the lower probability that the IPCC was right given the available assessments of climate sensitivity.

This particular problem of natural extension will be very hard to implement in practice, of course, not the least due to the intricate task of identifying the gamble $X_{IPCC}$. We wanted to point out, however, that lower previsions and their natural extension are fairly general concepts that can be employed to questions of statistical inference seemingly remote from the concept of gambles and previsions. Walley (1991, Section 3.1.1) provides an operational definition of natural extension in terms of a linear program for previsions. Utkin and Kozine (2001) have shown that Walley's definition is the dual form of an intuitively more accessible but often more difficult program to solve, which consists in finding the probability distribution that minimises the prevision $Pr(X)$, while observing the constraints $Pr(X_i) \geq \underline{Pr}(X_i)$.

No matter what form is used, the calculation of the natural extension becomes increasingly difficult with growing complexity of the universal set, and an increasing number of constraining previsions. In this situation, it is useful to identify a subset of the closed convex set $\mathcal{M}(\underline{Pr})$ of probabilities, and $\mathcal{PR}(\underline{Pr})$ of linear previsions, respectively, on which the statistical reasoning can be restricted without loss of generality.

**Definition C.6** *The set of* ***extreme points*** $\operatorname{ext}\mathcal{M}(\underline{Pr})$ ( $\operatorname{ext}\mathcal{PR}(\underline{Pr})$ ) *contains the elements of a closed convex set* $\mathcal{M}(\underline{Pr})$ ( $\mathcal{PR}(\underline{Pr})$ ) *of probabilities (linear previsions) which cannot be written as a convex combination of other elements. Hence,* $\mathcal{M}(\underline{Pr})$ ( $\mathcal{PR}(\underline{Pr})$ ) *constitutes the convex hull* $\operatorname{co}(\operatorname{ext}\mathcal{M}(\underline{Pr}))$ ( $\operatorname{co}(\operatorname{ext}\mathcal{PR}(\underline{Pr}))$ ) *of the set of extreme points. If the number of extreme points is finite, the set* $\operatorname{ext}\mathcal{M}(\underline{Pr})$ ( $\operatorname{ext}\mathcal{PR}(\underline{Pr})$ ) *is called* ***finitely generated****.*

It can be shown that for every gamble $X \in \mathcal{K}$ there exists an extreme point $P_{ext} \in \operatorname{ext}\mathcal{M}(\underline{Pr})$, or equivalently $Pr_{ext} \in \operatorname{ext}\mathcal{PR}(\underline{Pr})$, which fulfils $Pr_{ext}(X) = \underline{Pr}(X)$ (Walley, 1991, Theorem 3.6.2). Due to this property, it suffices to perform operations like natural extension, Bayesian updating of information, and decision making, on the set of extreme points $\operatorname{ext}\mathcal{M}(\underline{Pr})$, or equivalently $\operatorname{ext}\mathcal{PR}(\underline{Pr})$ (see, e.g., Walley, 1991, Sections 3.6.2, 8.4.8). If the number of extreme points is small, the statistical reasoning with imprecise probabilities is greatly simplified. In our football example, the gambler has specified a closed convex set $\mathcal{M}(\underline{Pr}_f)$ of probabilities with just three extreme points: $p_1 = \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$, $p_2 = \{\frac{1}{2}, \frac{1}{2}, 0\}$, and $p_3 = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\}$. The extreme points of the corresponding set $\mathcal{PR}(\underline{Pr}_f)$ of linear previsions are determined by the three (marginally) desirable gambles that were specified by the gambler. In general, the extreme points of convex sets of linear previsions are generated by (marginally) desirable gambles that are linear combinations of indicator gambles. However, not all convex sets need to be finitely generated.

The reader might ask why we have emphasized coherent lower previsions $\underline{Pr}$ on the space of measurable gambles $\mathcal{K}(\mathcal{A})$ as the representation of closed convex sets $\mathcal{M}(\underline{Pr})$ of probabilities, but have not yet paid attention to the lower envelope of $\mathcal{M}(\underline{Pr})$ on the event space $\mathcal{A}$ itself. From a historical perspective, this is indeed an omission since the lower envelope on the event space has received much more attention in the literature.

**Definition C.7** *A* ***coherent lower probability*** *on* $(\Omega, \mathcal{A})$ *is a set function* $\underline{P} : \mathcal{A} \to [0, 1]$ *constituting the lower envelope of a convex set* $\mathcal{M}$ *of additive probabilities on* $(\Omega, \mathcal{A})$*, i.e.,*

$$\forall\, A \in \mathcal{A} \quad \underline{P}(A) = \inf_{P \in \mathcal{M}} P(A)\,.$$

In contrast to coherent lower previsions, it does not seem possible to characterize coherent lower probabilities by a set of “Kolmogorov style” axioms. A necessary, but not sufficient condition for the coherence of lower probabilities is superadditivity, i.e., for any two disjoint sets $A, B \in \mathcal{A}$, it is $\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$ (Walley, 1991, Section 2.7.4). It can be seen from the definition of $\underline{P}$ that the upper envelope $\overline{P}$ of the convex set of probabilities is the conjugate set function to $\underline{P}$, i.e., $\overline{P}(A) = 1 - \underline{P}(A^c)$ for all $A \in \mathcal{A}$. Since $\underline{P}$ is superadditive, $\overline{P}$ needs to be subadditive. Due to the conjugacy relationship, coherent lower and upper probabilities contain the same information (if specified on the entire event space $\mathcal{A}$), so that it suffices to consider only one of the two envelopes.

A coherent lower prevision $\underline{Pr}$ uniquely determines a coherent lower probability $\underline{A}$ by use of the indicator gambles, i.e., $\underline{P}(A) := \underline{Pr}(I_A)$ for all $A \in \mathcal{A}$. However, the inverse is not true! There exists no unique way to calculate the lower expected value $\underline{Pr}(X)$ of a gamble $X$ from knowledge of just the coherent lower probability. By the same token, closed convex sets $\mathcal{M}$ of probabilities, which were shown to uniquely determine $\underline{Pr}$, cannot be fully represented by coherent lower probabilities. There exist subclasses of closed convex sets $\mathcal{M}$ which have the same lower envelope on the event space. Only the largest set in this subclass can be reconstructed from knowledge of a coherent lower probability.

**Definition C.8** *A* ***structure*** $\Gamma$ *is a closed convex set of additive probabilities on* $(\Omega, \mathcal{A})$ *that is generated by a coherent lower probability* $\underline{P}$ *on* $(\Omega, \mathcal{A})$ *by*

$$\Gamma(\underline{P}) := \{\, P : \ \forall\, A \in \mathcal{A} \quad \underline{P}(A) \leq P(A) \,\}$$

We have borrowed the term "structure" from the theory of interval probability of Weichselberger (2000, 2001). In the context of Weichselberger's theory, coherent lower probabilities are called *F-probabilities*, and convex subsets $\mathcal{M} \subset \Gamma$ with the lower envelope of the structure $\Gamma$ are called *prestructures*. Weichselberger (2000) showed that every convex set $\mathcal{M}$ of probabilities has a unique extension to a structure $\Gamma$ by constructing its lower envelope, and using Definition C.8 to generate $\Gamma$ from it. In general, this operation involves a loss of information, because $\Gamma$ might include more additive probabilities than were included in the original set $\mathcal{M}$.

As an example, take the convex set $\mathcal{M}(\underline{Pr}_f)$ of probabilities that specifies the gambler's belief about the outcome of the football game for the home team. It is not a structure. This can be seen by constructing its lower envelope on the power set $\mathcal{P}(\{W, D, L\})$. Since the power set only contains the sets $\{W\}, \{D\}, \{L\}, \{W, D\}, \{W, L\}, \{D, L\}$ beside the empty and the universal set, it suffices to specify the lower envelope in terms of lower and upper probabilities on the elementary events $\{W\}, \{D\}, \{L\}$. The upper probability on the elementary events determines the lower probability on the complementary events $\{W, D\}, \{W, L\}, \{D, L\}$ by the conjugacy relationship. Considering the set of extreme points $p_1 = \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$, $p_2 = \{\frac{1}{2}, \frac{1}{2}, 0\}$, and $p_3 = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\}$, we find immediately

$$\begin{aligned} &\underline{P}_f(\{W\}) = \frac{1}{3}, \qquad &&\underline{P}_f(\{D\}) = \frac{1}{4}, \qquad &&\underline{P}_f(\{L\}) = 0 \\ &\overline{P}_f(\{W\}) = \frac{1}{2}, \qquad &&\overline{P}_f(\{D\}) = \frac{1}{2}, \qquad &&\overline{P}_f(\{L\}) = \frac{1}{3}. \end{aligned} \tag{C.1}$$

Equalities (C.1) fully determine the coherent lower probability $\underline{P}_f$ which is the lower envelope of the convex set $\mathcal{M}(\underline{Pr}_f)$ of probabilities. The associated structure $\Gamma(\underline{P}_f)$ contains all probability distributions $p = \{p(W), p(D), p(L)\}$ with $\underline{P}_f(\{W\}) \leq p(W) \leq \overline{P}_f(\{W\})$, $\underline{P}_f(\{D\}) \leq p(D) \leq \overline{P}_f(\{D\})$, and $\underline{P}_f(\{L\}) \leq p(L) \leq \overline{P}_f(\{L\})$. $\Gamma(\underline{P}_f)$ has five extreme points: $p_1$, $p_2$, $p_3$ of $\mathcal{M}(\underline{Pr}_f)$, and in addition $p_4 = \{\frac{1}{3}, \frac{1}{2}, \frac{1}{6}\}$, and $p_5 = \{\frac{5}{12}, \frac{1}{4}, \frac{1}{3}\}$. Clearly, the structure $\Gamma(\underline{P}_f)$ is a true superset of the convex set $\mathcal{M}(\underline{Pr}_f)$ of probabilities.

Hence, the one-to-one correspondence between linear previsions on the space of gambles and additive probabilities on the space of events no longer persists when moving to coherent lower previsions and coherent lower probabilities. Coherent lower previsions are more informative then coherent lower probabilities, and constitute the only candidate to represent all closed convex sets of additive probabilities adequately. As a rule, uncertainty assessments involving *comparative probability statements* like, e.g., the gambler's belief that the home team will win the football game rather than drawing the match, can be represented only by coherent lower previsions, but not by coherent lower probabilities. Therefore, it is often argued that the concept of previsions for gambles is more fundamental for describing epistemic uncertainty than the concept of probabilities on events. The fixation on the concept of probability in the tradition of Laplace (1812) and Kolmogorov (1933) might have been a major obstacle to the construction of satisfactory theories of epistemic uncertainty which have matured only in the last decade.

## C.2 Special classes of imprecise probability

The general concepts of closed convex sets of probabilities and coherent lower previsions include a variety of uncertainty models as special cases (Walley, 1996b, 2000). To see this, it is useful to consider the following class of set functions.

**Definition C.9** *A* ***Choquet capacity*** $\mu$ *on* $(\Omega, \mathcal{A})$ *is a set function* $\mu : \mathcal{A} \rightarrow \mathbb{R}$ *which is monotone, i.e.,* $\mu(A) \leq \mu(B)$ *when* $A \subseteq B$*. A normalised Choquet capacity* $\mu : \mathcal{A} \rightarrow [0, 1]$ *assigns the value* $\mu(\emptyset) = 0$ *to the smallest element, and the value* $\mu(\Omega) = 1$ *to the largest element in the field* $\mathcal{A}$*.*

Choquet capacities were characterised by Choquet (1953) in his theory of capacities. Since coherent lower probabilities are monotone due to their superadditivity, they constitute a special case of normalised Choquet capacities. The same is true for coherent upper probabilities, whose monotonicity can be shown by their conjugacy relationship with lower probabilities. However, not every normalised Choquet capacity is necessarily the lower or upper envelope of a convex set of probabilities, or even dominated by a single additive probability. Nevertheless, the lower/upper probability interpretation constitutes the most important source of interest about Choquet Capacities. They have been employed in classical statistics (Huber, 1973; Huber and Strassen, 1973; Augustin, 1998), robust Bayesian analysis (Berger, 1993), decision theory (Gilboa, 1987; Schmeidler, 1989) and game theory, where they are called *monotone games* (Shapley, 1971).

Choquet capacities have also been used under the name *fuzzy measure* in fuzzy measure theory (Wang and Klir, 1992). A summary of the mathematical theory on capacities can be found in Denneberg (1994).

The value of Choquet capacities is threefold. They allow for a characterisation of non-additive set functions that provides a hierarchy of lower probability models. They allow the definition of an integral with respect to non-additive measures. And they exhibit an additive representation. A Choquet capacity can be characterised by the relationship between its value for a union of sets and the sum of its values for the individual sets and their intersections.

**Definition C.10** *A Choquet capacity $\mu : \mathcal{A} \to [0,1]$ is called* ***$n$-monotone*** *if and only if*

$$\mu(\cup_{i=1}^{n} A_i) \quad \geq \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \mu(\cap_{i \in I} A_i) \qquad \text{for all } \; A_i \in \mathcal{A}\,, \;\; 1 \leq i \leq n\,,$$

*and* ***n-alternating*** *if and only if*

$$\mu(\cup_{i=1}^{n} A_i) \quad \leq \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \mu(\cap_{i \in I} A_i) \qquad \text{for all } \; A_i \in \mathcal{A}\,, \;\; 1 \leq i \leq n\,.$$

*If the property of $n$-monotonicity can be extended to arbitrary $n \in \mathbb{N}$, a Choquet capacity $\mu$ is called* ***$\infty$-monotone*** *or* ***totally monotone****. Similarly, if the property of $n$-alternation can be extended to arbitrary $n \in \mathbb{N}$, $\mu$ is called* ***$\infty$-alternating*** *or* ***totally alternating****.*

The definition is intuitively easiest to access when considering the simplest case of 2-monotonicity, which requires $\mu(A \cup B) \geq \mu(A) + \mu(B) - \mu(A \cap B)$. 2-monotone capacities have also been called *supermodular*, or *convex capacities* in the literature. If $\mu$ is a probability measure, left and right hand side of the $n$-monotonicity condition, and consequently also of the $n$-alternation condition, are equal to each other. Additive probabilities are the only type of Choquet capacities which are totally monotone and totally alternating.

It has been shown that $n$-monotone capacities constitute special cases of coherent lower probabilities. The condition of 2-monotonicity is sufficient, but not necessary, for a set function to be the lower envelope of some convex set of probabilities (Walley, 1981). In his theory of interval probability, Weichselberger (2000, 2001) introduces 2-monotone lower probabilities under the name *C-probabilities*. Since $n+1$-monotonicity implies $n$-monotonicity, the monotonicity property induces a hierarchy of imprecise probability models, with the class of $n+1$-monotone lower envelopes being included in the more general class of $n$-monotone lower envelopes. The conjugate upper probability of an $n$-monotone lower probability is $n$-alternating.

2-monotone capacities are the most general class of coherent lower probabilities, whose coherence can be checked without showing that they constitute the lower envelope of a convex set of probabilities. They are also the most general class of coherent lower probabilities, which can be extended uniquely to a coherent lower prevision. The extension is performed on the basis of a generalised integral proposed by Choquet (1953).

**Definition C.11** *The **Choquet integral** of an $\mathcal{A}$-measurable gamble $X : \Omega \to \mathbb{R}$ with respect to a normalised Choquet capacity $\mu : \mathcal{A} \to [0,1]$ is given by*

$$E_\mu(X) = \int_0^\infty \mu(\{\omega | X(\omega) > x\})\, dx + \int_{-\infty}^0 (\, \mu(\{\omega | X(\omega) > x\}) \, - 1\,)\, dx\,.$$

*In terms of the conjugate capacity defined by $\mu^c(A) = 1 - \mu(A^c)$ for all events $A \in \mathcal{A}$, the Choquet integral reads*

$$E_\mu(X) = \int_0^\infty (\, 1 - \mu^c(\{\omega | X(\omega) \leq x\})\,)\, dx - \int_{-\infty}^0 \mu^c(\{\omega | X(\omega) \leq x\})\, dx\,.$$

The Choquet integral generalises the expectation formation about the outcome of a gamble from additive probabilities to Choquet capacities. If $\mu$ is an additive probability, the Choquet integral collapses to the usual expectation operation. In the context of fuzzy measure theory, the Choquet integral is employed under the name *Sugeno* or *fuzzy integral* (Sugeno, 1974).

As mentioned in the preceding section, there exists no one-to-one correspondence between coherent lower probabilities and coherent lower previsions in general. However, it has been shown (Huber, 1981; Walley, 1981) that the extension of 2-monotone lower probabilities $\underline{P} : \mathcal{A} \to [0,1]$ onto the space of $\mathcal{A}$-measurable gambles $K(\mathcal{A})$ by use of the Choquet Integral equals the natural extension (Definition C.5; note that a coherent lower probability determines the lower prevision for the indicator gambles). For coherent lower probabilities that are not 2-monotone the application of the Choquet integral is too conservative. It would yield an extension onto the space $K(\mathcal{A})$ that is strictly dominated by the natural extension, i.e., $E_\mu(X) < \underline{E}(X)$ for some gamble $X$.

Consider once more the football example. For universal sets with only three elements, as contained in $\Omega_f = \{W, D, L\}$, every coherent lower probability is 2-monotone. This can be derived from superadditivity of the lower probability, and the fact that either $A \cap B = \emptyset$, $A \subseteq B$ or $B \subset A$ for any pair $A, B$ in the power set $\mathcal{P}(\Omega_f)$. Hence, the coherent lower probability $\underline{P}_f$ described by the Equalities C.1 is 2-monotone. If we want to calculate its natural extension for some gamble on the outcome of the football game, we can use the Choquet integral. Consider the gamble $X = 3\, I_W + I_D$, which gives the home team three points if they win, one point if they draw the match and nothing if they lose. This is the typical gamble faced by European football teams in their national leagues. Taking the coherent lower probability as a representation of the epistemic uncertainty about the outcome of the game, the lower expected value of the points received by the home team is given by

$$\underline{E}_{\underline{P}_f}(X) = 1 \cdot \underline{P}_f(\{W, D\}) + (3-1) \cdot \underline{P}_f(\{W\}) = \frac{4}{3}\,.$$

The equality between Choquet integration and natural extension for the special class of 2-monotone lower probabilities is not a trivial result. It can greatly simplify the use of imprecise probabilities in the statistical inference process as well as in decision analysis. Therefore, 2-monotone capacities have constituted the dominant imprecise probability model in economic applications (see Mukerji and Tallon 2004 for an overview). An early application to the economics of climate change can be found in Lange (2000).

Although the hierarchic classification of Choquet capacities proves useful in identifying special classes of lower probabilities with favourable properties, there is still the need for simplifying the representation on the event space. Simplified representations have been studied in particular on finite universal sets $\Omega_n = \{\omega_1, ..., \omega_n\}$, for which the space of all possible events $A \subseteq \Omega_n$ is constituted by the power set $\mathcal{P}(\Omega_n)$ containing $2^n$ elements. Therefore, we will restrict ourselves to finite universal sets in the following. Since climate change assessments usually involve uncertain quantities on uncountable universal sets $\Omega$, this will require transforming the uncountable $\Omega$ to a finite *partition* $\Omega_n = \{A_1, ..., A_n\}$ with $n$ atoms[4]. The choice of the partition depends on the application, and we have explained our particular choice for this analysis in Section 3.4.1. Finite partitions of continuous spaces have to be used in any numerical calculation. For instance, non-parametric probability distributions for climate sensitivity in the literature are constituted by histogrammed data on a finite partition.

It was shown by Shafer (1976, Chapter 2) that any set function $\mu : \mathcal{P}_n \to \mathbb{R}$ on a finite power set $\mathcal{P}_n := \mathcal{P}(\Omega_n)$ has an additive representation in terms of another set function $\nu : \mathcal{P}_n \to \mathbb{R}$.

**Definition C.12** *The* ***Möbius inverse*** $\nu : \mathcal{P}_n \to \mathbb{R}$ *of a set function* $\mu : \mathcal{P}_n \to \mathbb{R}$ *is uniquely defined by*

$$\forall\, A \in \mathcal{P}_n \quad \nu(A) := \sum_{B \subseteq A} (-1)^{|A-B|}\, \mu(B)\,.$$

In turn, a set function $\mu : \mathcal{P}_n \to \mathbb{R}$ is represented by its Möbius inverse $\nu : \mathcal{P}_n \to \mathbb{R}$ as follows (Shafer, 1976):

$$\forall\, A \in \mathcal{P}_n \quad \mu(A) := \sum_{B \subseteq A} \nu(B)\,. \tag{C.2}$$

The properties of the Möbius inverse have been studied in Chateauneuf and Jaffray (1989) and Gilboa and Schmeidler (1994). For a generalisation of the Möbius representation to uncountable spaces see Gilboa and Schmeidler (1995). If a set function $\mu$ is normalised, it is $\mu(\Omega_n) := \sum_{B \in \mathcal{A}} \nu(B) = 1$, i.e., the Möbius inverse adds to unity over the event space $\mathcal{A}$. A set function $\mu$ is a Choquet capacity if and only if $\nu(\{\omega_i\}) \geq 0$ for all $\omega_i \in \Omega_n$ (Chateauneuf and Jaffray, 1989, Proposition 2). The Möbius inverse provides an efficient representation of a capacity if it is *sparse*, i.e., if it contains non-zero values $\nu(A) \neq 0$ only for a limited number $k$ of events $A$ in the power set $\mathcal{P}_n$ ($k \ll 2^n - 1$; $\nu(\emptyset) = 0$ iff $\mu(\emptyset) = 0$). The Möbius inverse of a probability measure, for

---

[4] A collection of sets $\{A_1, ..., A_n\}$ is a finite partition of $\Omega$, if $\cup_{i=1}^{n} A_i = \Omega$, and $A_i \cap A_j = \emptyset$ for $i \neq j$.

instance, has only non-zero values $\nu(\{\omega_i\}) > 0$ on the elementary events $\omega_i \in \Omega_n$. In this case, the Möbius inverse coincides with the probability mass function.

Consider once more the football example. Since the universal set $\Omega_f$ contains only three elements, the Möbius inverse of the 2-monotone lower probability $\underline{P}_f$ defined by equalities (C.1) can be easily calculated by use of Definition C.12. We find

$$\nu_f(\{W\}) = \frac{1}{3}\,, \qquad \nu_f(\{D\}) = \frac{1}{4}\,, \qquad \nu_f(\{L\}) = 0\,, \qquad \nu_f(\{W,D\}) = \frac{1}{12}$$
$$\nu_f(\{W,L\}) = \frac{1}{6}\,, \qquad \nu_f(\{D,L\}) = \frac{1}{4}\,, \qquad \nu_f(\{W,D,L\}) = -\frac{1}{12}\,. \tag{C.3}$$

The power set $\mathcal{P}(\Omega_f)$ contains 6 out of 8 elements with non-zero Möbius inverse. Therefore, the Möbius representation does not provide a large improvement over the coherent lower probability representation in this particular case. However, it can simplify the evaluation of the Choquet integral considerably . The Choquet integral of a $\mathcal{P}_n$-measurable gamble $X : \Omega_n \to \mathbb{R}$ with respect to a normalised Choquet capacity $\mu : \mathcal{P}_n \to [0,1]$ reads (e.g., Chateauneuf and Jaffray, 1989, Corollary 4)

$$E_\mu(X) = \sum_{B \in \mathcal{P}_n} \nu(B) \inf_{\omega \in B} X(\omega) \tag{C.4}$$

in terms of the Möbius inverse $\nu : \mathcal{P}_n \to [0,1]$ of $\mu$. In the case of the football example, we find once more $E_\mu(X) = 3\,\nu_f(\{W\}) + \nu_f(\{D\}) + \nu_f(\{W,D\}) = \frac{4}{3}$ for the gamble $X = 3\,I_W + I_D$.

The concept of the Möbius inverse becomes particularly valuable if the Choquet capacity is totally monotone (see Definition C.10).

**Definition C.13** *A totally monotone Choquet capacity* $bel : \mathcal{A} \to [0,1]$ *is called a* ***belief function****. Its conjugate totally alternating capacity* $pl : \mathcal{A} \to [0,1]$ *is called a* ***plausibility function****.*

Belief and plausibility functions are the least general class of coherent lower and upper probabilities that encompass additive probability as special case. It has been shown by Shafer (1976) that a Choquet capacity $\mu : \mathcal{P}_n \to [0,1]$ on a finite event space $\mathcal{P}_n$ is a belief function if and only if its Möbius inverse $\nu : \mathcal{P}_n \to \mathbb{R}$ contains only non-negative values, i.e., $\nu(A) \geq 0$ for all $A \in \mathcal{P}_n$. Since also $bel(\Omega) := \sum_{A \in \mathcal{P}_n} \nu(A) = 1$, the Möbius inverse of a belief function has been called a *probability mass assignment* in the literature. In contrast to additive probability, it is not defined on elementary events $\omega_i \in \Omega_n$, but on events $A \subseteq \Omega_n$.

The concept of belief and plausibility functions has emerged in the context of the *Dempster-Shafer theory of evidence*, which encompasses a variety of uncertainty models with different semantics. The original idea was proposed by Dempster (1967) who considered a probability mass assignment on an underlying space $\Psi = \{\psi_1, ..., \psi_n\}$ that is transferred onto a field $\mathcal{A}$ by means of a multivalued mapping $K : \Psi \to \mathcal{A}$. Later on, belief functions were given a non-probabilistic interpretation by Shafer (1976).

The last two decades have seen more variants of semantic models for belief functions, most notably the *transferable belief model* by Smets and Kennes (1994). Hence, the interpretation of belief functions needs to be considered carefully when studying their applications. An early application to climate change on the basis of Dempster's model was provided by Luo and Caselton (1997). In this analysis, we interpret belief functions exclusively as a special class of coherent lower probabilities.

**Definition C.14** *Let* $bel : \mathcal{P}_n \to [0,1]$ *a belief function on a finite event space, and* $\nu : \mathcal{P}_n \to [0,1]$ *its Möbius inverse. The events* $E \in \mathcal{P}_n$ *with non-zero Möbius inverse* $\nu(E) > 0$ *are called* ***focal elements****. The collection of focal elements along with their Möbius assignment is called a (finite support)* ***focal set*** *or* ***random set*** $(\mathcal{E}, \nu) := \{\, (E_1, \nu_1 := \nu(E_1)), ..., (E_k, \nu_k := \nu(E_k)) \,\}$.

The random set contains the full information of the Möbius inverse $\nu$ of *bel*. By means of Equation (C.2), it completely determines belief and plausibility function.

$$bel(A) := \sum_{B \subseteq A} \nu(B) \quad = \quad \sum_{i \,|\, E_i \subseteq A} \nu_i \,, \tag{C.5}$$

$$pl(A) := \sum_{B \cap A \neq \emptyset} \nu(A) \quad = \quad \sum_{i \,|\, E_i \cap A \neq \emptyset} \nu_i \,. \tag{C.6}$$

Equations (C.5) and (C.6) will be used frequently in this analysis to calculate lower and upper probabilities from a random set representation of the uncertainty.

Random sets are the preferable choice for representing the special class of imprecise probability models that can be described by a totally monotone lower probability. This is particularly true if the number $k$ of focal elements is much smaller than the number of events in the power set $\mathcal{P}_n$ ($k \ll 2^n$). We will show in this analysis that the use of random sets simplifies the statistical reasoning with belief functions by allowing for rather simple algorithms to combine, update and project their information content. In addition, they can be interpreted as a probability mass assignment on the focal elements. We are only able to assign probability masses with a level of precision described by the focal elements, but not on a smaller, more precise scale. Within a focal element, the probability mass can be arbitrarily distributed on the elementary events contained in it. This interpretation has been backed formally by the results of Chateauneuf and Jaffray (1989, Proposition 5 and Corollary 3). They showed that a probability distribution dominates a belief function if and only if it can be determined by a set of unit weighting functions on the focal elements and the probability mass assignment. Given this interpretation, how much mass will be contained in an arbitrary set $A \subset \mathcal{P}_n$? In the best case, every focal element which intersects $A$ will have its probability mass located entirely in the intersection (upper probability, Equation C.6). In the worst case, every focal element that is not fully enclosed by $A$ will have its mass situated outside the intersection (lower probability; Equation C.5).

As mentioned above, belief functions are the simplest representation of imprecise probabilities that incorporate classical probabilities as a special case. Assume that all focal elements of the random set are disjoint. Then the random set reduces to a probability mass distribution on the elementary events $\omega_i \in \Omega_n$, and belief and plausibility functions collapse into one additive probability measure. There exists another interesting special case, which is rather complementary to classical probability theory. Assume that the focal elements are nested, i.e., there exists a chain $E_1 \subseteq ... \subseteq E_k$. In this case, the imprecision in the uncertainty representation is very large. If an event $A$ does not intersect all focal elements, it includes none of them. Hence,

$$pl(A) < 1 \;\Rightarrow\; bel(A) = 0 \qquad \Leftrightarrow \qquad bel(A) > 0 \;\Rightarrow\; pl(A) = 1\,.$$

If the entire random set is constituted by a chain of nested focal elements, the plausibility function becomes a possibility measure $\Pi : \mathcal{P}_n \to [0,1]$ (Dubois and Prade, 1990).

**Definition C.15** *A* ***possibility measure*** $\Pi : \mathcal{P} \to [0,1]$ *on the power set of a (possibly uncountable) universal set* $\Omega$ *is defined by*

$$\Pi(\cup_{i\in I} A_i) = \sup_{i\in I} \Pi(A_i)$$

*for any index set* $I$ *and any family of subsets* $A_i \subseteq \mathcal{P}$*. Its conjugate set function is called a* ***necessity measure*** $N : \mathcal{P} \to [0,1]$*.*

Possibility and necessity measures are the primitives of *possibility theory* (Zadeh, 1978; Dubois and Prade, 1988; de Cooman, 1997). It is mathematically convenient in the sense that the set functions $\Pi$ and $N$ can be represented by a single *possibility distribution* $\pi : \Omega_n \to [0,1]$,

$$\forall\, A \in \mathcal{P}_n \quad \Pi(A) = \sup_{\omega\in A} \pi(\omega)\,. \tag{C.7}$$

Here, the convention $\sup_{\omega\in\emptyset}(.) := 0$ has been silently assumed. The possibility distribution $\pi$ is normalized by the condition that there exists a $\omega \in \Omega$ with $\pi(\omega) = 1$. Thus, formally it represents a *normal fuzzy set*.

Possibility theory has been proposed by Zadeh (1978) in order to provide an uncertainty model for fuzzy sets. Since then, possibility theory has received considerable attention (Dubois and Prade, 1998), but was rarely given an imprecise probability interpretation. Possibility measures as a special case of totally monotone upper probabilities have been analysed in Dubois and Prade (1992) and de Cooman and Aeyels (1999). They can play only a little role as imprecise probability model since they represent the upper envelope of a rather artificial type of structure $\Gamma_\Pi := \{P \,|\, \forall A \in \mathcal{A} \;\; P(A) \leq \Pi(A)\}$. Due to relationship (C.7) in combination with the normalisation of possibility distributions, $\Gamma_\Pi$ has to contain at least one Dirac measure representing complete information about the true state of the world. We have studied such structures in Kriegler and Held (2003), but a discussion of this work is beyond the scope of this thesis. Since possibility measures constitute only a very limited model of imprecise probability, we will not consider them any further here.

We summarise our survey of imprecise probability theory by recapitulating the hierarchy of imprecise probability models. The most general models that were discussed are closed convex set of probabilities and coherent lower previsions. Both models are equivalent to each other. They are the only models that can capture comparative probability statements. A special class of these models are coherent lower probabilities, which are also a special type of Choquet capacity. Mathematical tractability is considerably improved when restricting coherent lower probabilities further to 2-monotone lower probabilities. Among 2-monotone probabilities, belief functions have a particularly accessible representation in terms of their Möbius inverse. Belief functions are the least general class that generalises additive probability. They also encompass possibility theory when the latter is interpreted as an imprecise probability model (which it usually is not).

# Appendix D

# Proof of Propositions

**Proof of Lemma 3.1:**

(I) Since at least one index $i \in \{1, ..., n\}$, $j \in \{1, ..., m\}$ is raised by one in each iteration, the algorithm enters step (2) at most $n + m - 1$ times before it stops.

(II) Since the indices $i \in \{1, ..., n\}$, $j \in \{1, ..., m\}$ are either raised by one or remain unchanged in each iteration, $i(k) \leq i(l)$ and $j(k) \leq j(l)$ for iterations $k < l$. Thus, $x^*_{j(k)} \leq x^*_{j(l)}$ and $x_{*i(k)} \leq x_{*i(l)}$ for iterations $k < l$.

(III) Consider an arbitrary $x \in \mathbb{R}$.
**Belief function:** If $x < x_{*1}$, then there exists no focal element with $E_k \subseteq (-\infty, x]$. Therefore, we have $bel(-\infty, x] = \underline{SF}(x) = 0$. If $x \geq x_{*n}$, then all focal elements have the property $E_k \subseteq (-\infty, x]$. Therefore, we have $bel(-\infty, x] = \underline{SF}(x) = 1$. Assume $x_{*1} \leq x < x_{*n}$. Let $E_l = (x^*_l, x_{*l}]$ be the focal element with $x_{*l} \leq x < x_{*l+1}$. Then, we have $bel_{\mathcal{E}}(-\infty, x] := \sum_{k|E_k \subseteq (-\infty,x]} \nu_k = \sum_{k \leq l} \nu_k$ due to (II). By construction of the algorithm, $\sum_{k \leq l} \nu_k = p_l := \underline{SF}(x_{*l})$. Since $x_{*l} \leq x < x_{*l+1}$, we have $\underline{SF}(x_{*l}) = \underline{SF}(x)$.

**Plausibility function:** If $x \leq x^*_1$, then all focal elements have the property $E_k \cap (-\infty, x] = \emptyset$. Therefore, we have $pl(-\infty, x] = \overline{SF}(x) = 0$. If $x > x^*_m$, then all focal elements have the property $E_k \cap (-\infty, x] \neq \emptyset$. Therefore, we have $pl(-\infty, x] = \overline{SF}(x) = 1$. Assume $x^*_1 < x \leq x^*_m$. Let $E_l = (x^*_l, x_{*l}]$ be the focal element whith $x^*_l < x \leq x^*_{l+1}$. Then, we have $pl_{\mathcal{E}}(-\infty, x] := \sum_{k|E_k \cap (-\infty,x] \neq \emptyset} \nu_k = \sum_{k \leq l} \nu_k$ due to (II). By construction of the algorithm, $\sum_{k \leq l} \nu_k = p_l := \overline{SF}(x^*_{l+1})$. Since $x^*_l < x \leq x^*_{l+1}$, we have $\overline{SF}(x^*_{l+1}) = \overline{SF}(x)$. $\square$

**Proof of Theorem 3.1:**

**Step 1**: The random set $\mathcal{E}$ has Property (III), Lemma 3.1. Hence, we have $\underline{P}_X(A) =$

$bel_{\mathcal{E}}(A)$ for every event $A = (-\infty, x]$, $x \in \mathbb{R}$. Likewise, we have $\underline{P}_X(\emptyset) = bel_{\mathcal{E}}(\emptyset) = 0$, and $\underline{P}_X(\mathbb{R}) = bel_{\mathcal{E}}(\mathbb{R}) = 1$.

**Step 2**: Consider an arbitrary half closed interval $(a, b] \subset \mathbb{R},\ a < b$. we have to show that $\underline{P}_X(a, b] = bel_{\mathcal{E}}(a, b]$.

$$\underline{P}_X(a,b] = \max[\,0, \underline{SF}(b) - \overline{SF}(a)] = \max[\,0, \sum_{i \,|\, E_i \subseteq (-\infty, b]} \nu_i - \sum_{j \,|\, E_j \cap (-\infty, a] \neq \emptyset} \nu_j\,]\,.$$

If $\underline{SF}(b) < \overline{SF}(a)$, there exists a focal element $\hat{E} = (\hat{x}^*, \hat{x}_*] \in \mathcal{E}$ with $\hat{E} \cap (-\infty, a] \neq \emptyset$ and $\hat{E} \not\subseteq (-\infty, b]$. Assume now that an $\acute{E} \in \mathcal{E}$ with $\acute{E} = (\acute{x}^*, \acute{x}_*] \subseteq (a, b]$ would exist. Then, $a \leq \acute{x}^* < \acute{x}_* \leq b$, and $\hat{x}^* < \acute{x}^* < \acute{x}_* < \hat{x}_*$. The latter, however, contradicts Property (II) in Lemma 3.1, and we conclude that such an $\acute{E} \in \mathcal{E}$ does not exist. Hence, $bel_{\mathcal{E}}(a, b] = 0$.

Assume, vice versa, that there exists a focal element $\hat{E} \in \mathcal{E}$ with $\hat{E} \cap (-\infty, a] \neq \emptyset$ and $\hat{E} \not\subseteq (-\infty, b]$. Let $E$ be an arbitrary focal element with $E \subseteq (-\infty, b]$. Then either $E \subseteq (a, b]$ or $E \cap (-\infty, a] \neq \emptyset$. It was shown in the last paragraph that the existence of $\hat{E}$ excludes $E \subseteq (a, b]$. Hence, all $E_i \subseteq (-\infty, b] \in \mathcal{E}$ intersect $(-\infty, a]$, and $\underline{SF}(b) < \overline{SF}(a)$. Therefore, if $\underline{SF}(b) \geq \overline{SF}(a)$, there is no such focal element $\hat{E} \in \mathcal{E}$, i.e., for all focal elements $E_i \not\subseteq (-\infty, b] \Rightarrow E_i \cap (-\infty, a] = \emptyset$. Then,

$$\underline{P}_X(a,b] = \sum_{s(i) \,|\, E_{s(i)} \subseteq (a,b]} \nu_{s(i)} + \sum_{t(i) \,|\, E_{t(i)} \cap (-\infty, a] \neq \emptyset} \nu_{t(i)} - \sum_{j \,|\, E_j \cap (-\infty, a] \neq \emptyset} \nu_j = bel_{\mathcal{E}}(a, b]\,.$$

**Step 3:** Consider an arbitrary Borel set $B \in \mathcal{R}$. Let $E_1, ..., E_n$ be the focal elements that are fully contained in $B$, and $E_{n+1}, ...E_k$ the remaining focal elements of $\mathcal{E}$. Due to Property (I), Lemma 3.1, $E = \cup_{i=1}^{n} E_i$ is a union of $m \leq n$ disjoint half-closed intervals $E = (a_1, b_1] \cup ... \cup (a_m, b_m],\ a_1 < b_1 < ... < a_m < b_m$, where no pair of half-closed intervals exhibits common accumulation points. Choose a CDF $F' : \mathbb{R} \to [0, 1]$ with $F'(a_1) = \min[\overline{SF}(a_1), \underline{SF}(b_1)]$, $F'(b_1) = \underline{SF}(b_1)$, ..., $F'(a_m) = \min[\overline{SF}(a_m), \underline{SF}(b_m)]$, $F'(b_m) = \underline{SF}(b_m)$. Since $F'(a_1) \leq F'(b_1) \leq ... \leq F'(a_m) \leq F'(b_m)$, such a CDF does exist, and is contained in $\Gamma_X(\underline{SF}, \overline{SF})$. Given this probability specification, we have

$$\begin{aligned} P'(E) &= F'(b_m) - F'(a_m) + ... + F'(b_1) - F'(a_1) \\ &= \max[0, \underline{SF}(b_m) - \overline{SF}(a_m)] + ... + \max[0, \underline{SF}(b_1) - \overline{SF}(a_1)] \\ &= \underline{P}_X(a_m, b_m] + ... + \underline{P}_X(a_1, b_1]\,. \end{aligned}$$

Since the lower envelope $\underline{P}_X$ is super-additive on a union of disjoint sets (Walley, 1991, Ch. 2.7.4), we have $\underline{P}_X(E) = P'(E)$. Since $\underline{P}_X(a_i, b_i] = bel(a_i, b_i]$ as shown in Step 2, and each focal element contained in $E$ is contained in exactly one interval $(a_i, b_i]$,

$$\underline{P}_X(E) \;=\; \sum_{i=1}^{m} \sum_{j | E_j \subseteq (a_i, b_i]} \nu_j \;=\; \sum_{j | E_j \subseteq \bigcup\limits_{i=1}^{m} (a_i, b_i]} \nu_j \;=\; bel_{\mathcal{E}}(E)\,.$$

Since a lower envelope $\underline{P}_X$ is a monotone set function, we have

$$\underline{P}_X(B) \;\geq\; \underline{P}_X(E) \;=\; bel_{\mathcal{E}}(E) \;=\; bel_{\mathcal{E}}(B)\ .$$

**Step 4:** Consider an arbitrary Borel set $B \in \mathcal{R}$. Let $E_1, ..., E_n$ be the focal elements that are fully contained in $B$, and $E_{n+1}, ..., E_k$ the remaining focal elements of $\mathcal{E}$. Choose a right-continuous step function $SF^* : \mathbb{R} \to [0,1]$ as follows. For each focal element $E_i \not\subseteq B$, $n < i \leq k$, introduce a step of height $\nu_i$ at a point $x_i \in E_i$, $x_i \notin B$. For each focal element $E_i \subseteq B$, $1 \leq i \leq n$ introduce a step of height $\nu_i$ at an arbitrary point $x_i \in E_i$. Since $\sum_{i=1}^{k} \nu_i = 1$, $SF^*$ so defined is a CDF of some Dirac $\delta$-measure $P^*$. Moreover, we have

$$\begin{aligned}
\forall\, x \in \mathbb{R} \quad SF^*(x) \;&\geq \sum_{j\,|\,E_j \subseteq (-\infty,x]} \nu_j = bel(-\infty,x] = \underline{SF}(x)\ ,\\
SF^*(x) \;&\leq \sum_{j\,|\,E_j \cap (-\infty,x] \neq \emptyset} \nu_j = pl(-\infty,x] = \overline{SF}(x)\ ,
\end{aligned}$$

so that $P^* \in \Gamma_X(\underline{SF}, \overline{SF})$. Clearly, $P^*(B) = \sum_{i=1}^{n} \nu_i = \sum_{i\,|\,E_i \subseteq B} \nu_i = bel_{\mathcal{E}}(B)$. Since we have established $\underline{P}_X(B) \geq bel_{\mathcal{E}}(B)$ for arbitrary Borel sets $B$ in Step 3, it is $\underline{P}_X(B) = bel_{\mathcal{E}}(B)$. $\square$

**Proof of Corollary 3.1:**

**Show $\Gamma_X(bel_{\mathcal{E}}) \subseteq \Gamma_X(\underline{SF}, \overline{SF})$ :**
Choose an arbitrary probability $P_X \in \Gamma_X(bel_{\mathcal{E}})$.
$\Rightarrow\ \forall\, x \in \mathbb{R}\ \ P_X(-\infty,x] \geq bel_{\mathcal{E}}(-\infty,x]$ and $P_X(-\infty,x] \leq pl_{\mathcal{E}}(-\infty,x]$.
$\Rightarrow\ \ P_X \in \Gamma_X(\underline{SF}, \overline{SF})$ by definition.

**Show $\Gamma_X(bel_{\mathcal{E}}) = \Gamma_X(\underline{SF}, \overline{SF})$, if $(\mathcal{E}, \nu)$ has Properties (I) and (II):**
Use Algorithm 3.1 to construct a random set $(\mathcal{E}', m')$ from $\underline{SF}$ and $\overline{SF}$. $(\mathcal{E}', m')$ and $(\mathcal{E}, \nu)$ both have Properties (I) to (III), Lemma 3.1. This implies, inter alia, that for any half-closed interval $(a, b]$ on the real line we have $bel_{\mathcal{E}'} = bel_{\mathcal{E}} := \max[0, \underline{SF}(b) - \overline{SF}(a)]$ (see Step 1 in Proof of Theorem 3.1). Since both random sets contain only half-closed intervals, it follows $(\mathcal{E}', m') = (\mathcal{E}, \nu)$. Hence, $bel_{\mathcal{E}} = bel_{\mathcal{E}'}$, and according to Theorem 3.1, $bel_{\mathcal{E}}$ is the lower envelope of $\Gamma_X(\underline{SF}, \overline{SF})$.

**Show $\Gamma_X(bel_{\mathcal{E}}) = \Gamma_X(\underline{SF}, \overline{SF})$, only if $(\mathcal{E}, \nu)$ has Properties (I), (II):**
Assume that $(\mathcal{E}, \nu)$ does not fulfil property (I) or (II) in Lemma 3.1. Let $(\mathcal{E}', m')$ be the random set constructed from $\underline{SF}$, and $\overline{SF}$ by Algorithm 3.1, and $bel_{\mathcal{E}'}$ the associated belief function. According to Lemma 3.1, $(\mathcal{E}', m')$ has properties (I) to (III). Then, $(\mathcal{E}, \nu) \neq (\mathcal{E}', m')$, and $bel'_{\mathcal{E}} \neq bel_{\mathcal{E}}$. According to Theorem 3.1, $bel'_{\mathcal{E}}$ is the lower envelope of $\Gamma_X(\underline{SF}, \overline{SF})$. Hence, $bel_{\mathcal{E}}$ cannot be the lower envelope, and $\Gamma_X(bel_{\mathcal{E}}) \neq \Gamma_X(\underline{SF}, \overline{SF})$. $\square$

**Proof of Lemma 3.2:**

**Step 1:** Let $\mathcal{B} = \{(-\infty, x_1], ..., (-\infty, x_m], (y_1, \infty), ..., , (y_{m'}, \infty), A_1, ..., A_k\}$ be the collection of sets, on which the defining lower probability constraints of the p-box $\Gamma(\underline{SF}, \overline{SF})$ and the $\varepsilon$-contamination model $\Gamma(\underline{p})$ are specified. $\{A_1, ..., A_k\}$ constitutes a partition of $S$, and $x_1 < ... < x_m$, $y_1 < ... < y_{m'} \in \mathbb{R}$, Let $\underline{P}|_{\mathcal{B}}$ be a lower probability on $\mathcal{B}$ that is defined by the joint constraints of $\Gamma(\underline{SF}, \overline{SF})$ and $\Gamma(\underline{p})$, i.e.,

$$\underline{P}(-\infty, x_i] := \underline{SF}(x_i) \ , \qquad \underline{P}(y_{i'}, \infty) := 1 - \overline{SF}(y_{i'}) \ , \qquad \underline{P}(A_j) = \underline{p}_j \ .$$

$\underline{P}|_{\mathcal{B}}$ generates a convex set of probabilities $\mathcal{M}(\underline{P}|_{\mathcal{B}}) := \{P \,|\, \forall\ B \in \mathcal{B} \ \ P(B) \geq \underline{P}(B)\}$. Clearly, it is $\mathcal{M}(\underline{P}|_{\mathcal{B}}) \supseteq \Gamma(\underline{SF}, \overline{SF}, \underline{p})$, because every $P \in \Gamma(\underline{SF}, \overline{SF}, \underline{p})$ dominates $\underline{P}$ on $\mathcal{B}$, and therefore needs to be included in $\mathcal{M}(\underline{P}|_{\mathcal{B}})$. In turn, every $P \in \mathcal{M}(\underline{P}|_{\mathcal{B}})$ dominates $bel_{\mathcal{E}}$ on the events $(-\infty, x_i]$, $1 \leq i \leq m$, and $(y_{i'}, +\infty)$, $1 \leq i' \leq m'$. It also dominates $bel_{\underline{p}}$ on the events $A_j$, $1 \leq j \leq k$. Therefore, it will be included in $\Gamma(\underline{SF}, \overline{SF})$ and $\Gamma(\underline{p})$ by definition of these sets of probabilities. Hence, $\mathcal{M}(\underline{P}|_{\mathcal{B}}) = \Gamma(\underline{SF}, \overline{SF}, \underline{p})$.

Since $\Gamma(\underline{SF}, \overline{SF})$ and $\Gamma(\underline{p})$ are compatible, it is $\Gamma(\underline{SF}, \overline{SF}, \underline{p}) = \Gamma(\underline{SF}, \overline{SF}) \cap \Gamma(\underline{p}) \neq \emptyset$. It follows from the Natural Extension Theorem of Walley (1991, Theorem 3.4.1) that the natural extension $\underline{P}_E : \mathcal{R} \to [0, 1]$ of the lower probability $\underline{P}|_{\mathcal{B}} : \mathcal{B} \to [0, 1]$ is the lower envelope of $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$ on the entire Borel field $\mathcal{R}$.

By assumption, $bel_{\mathcal{E}}$ is the lower envelope of $\Gamma(\underline{SF}, \overline{SF})$ (see also Theorem 3.1), and $bel_{\underline{P}}$ the lower envelope of $\Gamma(\underline{p})$. Hence,

$$\underline{P}_E(A) \geq \max[bel_{\mathcal{E}}(A),\, bel_{\underline{p}}(A)] \qquad \text{for all} \quad A \in \mathcal{R} \ .$$

Moreover, since the natural extension is a coherent lower probability (Walley, 1991, Theorem 3.1.2), it is in particular superadditive on the union $A \cup B$ of two disjoints sets $A, B \in \mathcal{R}$ (Walley, 1991, Section 2.7.4). Hence,

$$\underline{P}_E(A \cup B) \geq \max[bel_{\mathcal{E}}(A),\, bel_{\underline{p}}(A)] + \max[bel_{\mathcal{E}}(B),\, bel_{\underline{p}}(B)] \ \text{ for all disjoint } A, B \in \mathcal{R} \ .$$

Consider an arbitrary $A \in \mathcal{R}$ which does not contain a focal element $E_i \in \mathcal{E}$. Then,

$$\underline{P}_E(A) \geq \max[bel_{\mathcal{E}}(A),\, bel_{\underline{p}}(A)] = bel_{\underline{p}}(A) = \underline{P}^*(A) \ .$$

Consider now an event $A \in \mathcal{R}$ which contains at least one focal element $E_i \in \mathcal{E}$. For an arbitrary union $E \in \mathcal{U}$ of focal elements with $E \subseteq A$, it follows from super-additivity of $P_E$ that

$$\begin{aligned} \underline{P}_E(A) &\geq \max[bel_{\mathcal{E}}(E),\, bel_{\underline{p}}(E)] + \max[bel_{\mathcal{E}}(A \cap E^c),\, bel_{\underline{p}}(A \cap E^c)] \\ &\geq \max[bel_{\mathcal{E}}(E) + bel_{\underline{p}}(A \cap E^c),\, bel_{\underline{p}}(A)] + \max[bel_{\mathcal{E}}(A \cap E^c) - bel_{\underline{p}}(A \cap E^c),\, 0] \\ &\geq \max[bel_{\mathcal{E}}(E) + bel_{\underline{p}}(A \cap E^c),\, bel_{\underline{p}}(A)] \,. \end{aligned}$$

Since this is true for arbitrary $E \subseteq A$, we find

$$\underline{P}_E(A) \quad \geq \quad \max_{E\in\mathcal{U},\, E\subseteq A} \Big( \max[bel_{\mathcal{E}}(E) + bel_{\underline{p}}(A \cap E^c),\, bel_{\underline{p}}(A)] \Big) = \underline{P}^*(A)\,. \qquad \square$$

**Proof of Lemma 3.3:**

**Step 1:** Consider an arbitrary union $E^s_{i(s)} \in \mathcal{U}$ of focal elements $E_i \in \mathcal{E}$ with an arbitrary hierarchy level $s$. By construction of Algorithm 3.2, a level $s$-set $E^s_{i(s)}$ will contain only such focal elements $G \in \mathcal{G}$ as true subsets that are either atoms or level-$s'$-sets $E^{s'}_{i(s')} \in \mathcal{U}$ with $s' < s$. Assuming the contrary that $s' \geq s$ would lead to a contradiction. Then $E^{s'}_{i(s')}$ needed to contain more or an equal number of unions $E \in \mathcal{U}$ than $E^s_{i(s)}$ (see Step 4). But this is impossible because $E^{s'}_{i(s')} \subset E^s_{i(s)}$. Hence,

$$\sum_{l|G_l\subseteq E^s_{i(s)}} \nu^*_l = \sum_{\substack{l|\, G_l\in\mathcal{G}^{s-1} \\ G_l\subseteq E^s_{i(s)}}} \nu^*_l + \nu^*_{i(s)} =: \underline{P}^*(E^s_{i(s)})\,,$$

by construction of the Möbius assignment $\nu^*_{i(s)}$ in Step 6. Note that this is also true for the focal elements $E_i \in \mathcal{E}$ themselves, since they constitute level-1-sets.

**Step 2:** Consider an arbitrary $A \in \mathcal{R}$. Let $E \in \mathcal{U}$ be the union of all focal elements $E_i \in \mathcal{E}$ that are contained in $A$ (possibly empty). Let $A_1, ..., A_l$ be the collection of atoms that are contained in $A \cap E^c$ (possibly none). Then,

$$\sum_{l|G_l\subseteq A} \nu^*_l = \sum_{l|G_l\subseteq E} \nu^*_l + \sum_{j=1}^{l} \underline{p}_j = \underline{P}^*(E) + bel_{\underline{p}}(A \cap E^c)\,,$$

where we have made use of Assumption 3.1 to establish the first equality, and utilised the result from Step 1 of this proof to establish the second equality. Assuming first that $A$ contains no focal element $E_i \in \mathcal{E}$. Then $E = \emptyset$, and $\sum_{l|G_l\subseteq A} \nu^*_l = bel_{\underline{p}}(A) = \underline{P}^*(A)$. Assuming next that $A$ contains all focal elements $E_i \in \mathcal{E}$. Then $E = \cup^n_{i=1} E_i$, and $\underline{P}^*(E) = 1$. Moreover, compatibility of $\Gamma(\underline{SF}, \overline{SF})$ and $\Gamma(\underline{p})$ requires $bel_{\underline{p}}(A_j) \leq pl_{\mathcal{E}}(A_j)$ for all atoms $A_j \in \{A_1, ..., A_k\}$, which implies that every atom with $\underline{p}_j > 0$ needs to be contained in at least one focal element $E_i \in \mathcal{E}$. Hence, $bel_{\underline{p}}(A \cap E^c) = 0$, and $\sum_{l|G_l\subseteq A} \nu^*_l = \underline{P}^*(E) = 1$ and $\underline{P}^*(A) = \underline{P}^*(E) = 1$.

Consider now the remaining non-trivial case that $A$ contains some, but not all focal elements $E_i \in \mathcal{E}$. Then, $A$ cannot contain the support $S = A_1 \cup ... \cup A_k$ of the partition, because Assumption 3.1 implies $S \supseteq E_1 \cup ... \cup E_n$. If $A \not\supseteq S$, then it is

$$bel_{\underline{p}}(A) = bel_{\underline{p}}(E) + bel_{\underline{p}}(A \cap E^c)\ ,$$

since each atom $A_j \subseteq A$ is either contained in $E$ or in $A \cap E^c$ because of Assumption 3.1. The same holds true for any other two disjoint subsets $B_1, B_2$ of $A$, when at least one of these subsets is contained in the power set $\mathcal{P}(A_1, ..., A_k)$. Hence,

$$\begin{aligned}
\sum_{l|G_l\subseteq A} \nu_l^* &= \underline{P}^*(E) + bel_{\underline{p}}(A \cap E^c) \\
&= \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E \cap \tilde{E}^c),\, bel_{\underline{p}}(E)] \Big) + bel_{\underline{p}}(A \cap E^c) \\
&= \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E \cap \tilde{E}^c) + bel_{\underline{p}}(A \cap E^c),\, bel_{\underline{p}}(A)] \Big) \\
&= \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(A \cap \tilde{E}^c),\, bel_{\underline{p}}(A)] \Big)\ ,
\end{aligned}$$

where the last equality follows from additivity of $bel_{\underline{p}}$ on $E \cap \tilde{E}^c$, $A \cap E^c$ (see above). Since $E$ is the union of all focal elements in $A$, every union $\tilde{E} \subseteq A$ also needs to be contained in $E$. Therefore, we find

$$\sum_{l|G_l\subseteq A} \nu_l^* = \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq A} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(A \cap \tilde{E}^c),\, bel_{\underline{p}}(A)] \Big) = \underline{P}^*(A)\,. \qquad \square$$

**Proof of Lemma 3.4:**

**Step 1:** We begin by showing that $\underline{P}^* : \mathcal{R} \to [0,1]$ is additive for two arbitrary disjoint sets $E \in \mathcal{U}$ and $F \in \mathcal{U}$.

$$\begin{aligned}
\underline{P}^*(E) + \underline{P}^*(F) &= \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E \cap \tilde{E}^c),\, bel_{\underline{p}}(E)] \Big)\ + \\
&\quad\ \max_{\tilde{F}\in\mathcal{U},\, \tilde{F}\subseteq F} \Big( \max[bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(F \cap \tilde{F}^c),\, bel_{\underline{p}}(F)] \Big) \\
&= \max_{\substack{\tilde{E},\, \tilde{F}\in\mathcal{U}, \\ \tilde{E}\subseteq E\,,\, \tilde{F}\subseteq F}} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E \cap \tilde{E}^c),\, bel_{\underline{p}}(E)] + \\
&\qquad\qquad \max[bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(F \cap \tilde{F}^c),\, bel_{\underline{p}}(F)] \Big) \\
&= \max_{\substack{\tilde{E},\, \tilde{F}\in\mathcal{U}, \\ \tilde{E}\subseteq E\,,\, \tilde{F}\subseteq F}} \Big( \max[\, bel_{\underline{p}}(E) + bel_{\underline{p}}(F)\ \ , \\
&\qquad\qquad bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E \cap \tilde{E}^c) + bel_{\underline{p}}(F)\ , \\
&\qquad\qquad bel_{\underline{p}}(E) + bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(F \cap \tilde{F}^c)\ , \\
&\qquad\qquad bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E \cap \tilde{E}^c) + bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(F \cap \tilde{F}^c)\,] \Big)\ ,
\end{aligned}$$

where we have converted the sum of the individual max-operations between $bel_{\mathcal{D}}(\tilde{D}) + bel_{\underline{p}}(D \cap \tilde{D}^c)$ and $bel_{\underline{p}}(D)$, $D \in \{E, F\}$, to a max-operation over all four possible cases. Since $E$ and $F$ are disjoint, we also have $\tilde{E} \cap \tilde{F} = \emptyset$. It was discussed in the proof of Lemma 3.3 that $bel_{\underline{p}}$ is additive on two disjoint sets, if their union does not include the support $S$ of the partition, and if at least one of them is contained in the power set of the partition $\mathcal{P}(A_1, ..., A_k)$. This is true for all $E \in \mathcal{U}$ due to Assumption (3.1). Hence,

$$\begin{aligned}
bel_{\underline{p}}(E) + bel_{\underline{p}}(F) &= bel_{\underline{p}}(E \cup F) \ , \\
bel_{\underline{p}}(E \cap \tilde{E}^c) + bel_{\underline{p}}(F) &= bel_{\underline{p}}(\,(E \cap \tilde{E}^c) \cup F\,) \ = \ bel_{\underline{p}}(\,(E \cup F) \cap \tilde{E}^c\,) \ , \\
bel_{\underline{p}}(E) + bel_{\underline{p}}(F \cap \tilde{F}^c) &= bel_{\underline{p}}(\,E \cup (F \cap \tilde{F}^c)\,) \ = \ bel_{\underline{p}}(\,(E \cup F) \cap \tilde{F}^c\,) \ , \\
bel_{\underline{p}}(E \cap \tilde{E}^c) + bel_{\underline{p}}(F \cap \tilde{F}^c) &= bel_{\underline{p}}(\,(E \cap \tilde{E}^c) \cup (F \cap \tilde{F}^c)\,) \\
&= bel_{\underline{p}}(\,(E \cup F) \cap (\tilde{E} \cup \tilde{F})^c\,) \ ,
\end{aligned}$$

where the second equalities follow from the fact that

$$\tilde{E}^c \supseteq (E)^c \supset F \supseteq F \cap \tilde{F}^c \ , \qquad \tilde{F}^c \supseteq (F)^c \supset E \supseteq E \cap \tilde{E}^c \ ,$$

due to $E \cap F = \emptyset$, and $\tilde{E} \subseteq E$ and $\tilde{F} \subseteq F$. Moreover, since all focal elements $E_i \in \mathcal{E}$ are intervals of the real-line (see Lemma 3.1), and $\tilde{E}$, $\tilde{F}$ are disjoint, each focal element $E_i \subseteq \tilde{E} \cup \tilde{F}$ is either contained in $\tilde{E}$ or in $\tilde{F}$. Hence, $bel_{\mathcal{E}}$ is additive on the two sets, i.e.,

$$bel_{\mathcal{E}}(\tilde{E}) + bel_{\mathcal{E}}(\tilde{F}) = bel_{\mathcal{E}}(\tilde{E} \cup \tilde{F}) \ .$$

By including all these identities into the expression for $\underline{P}^*(E) + \underline{P}^*(F)$, we find

$$\begin{aligned}
\underline{P}^*(E) + \underline{P}^*(F) \quad = \quad \max_{\substack{\tilde{E},\,\tilde{F} \in \mathcal{U}, \\ \tilde{E} \subseteq E\,,\ \tilde{F} \subseteq F}} \Big( \max[\, & bel_{\underline{p}}(E \cup F) \ \ , \\
& bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(\,(E \cup F) \cap \tilde{E}^c\,) \ , \\
& bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(\,(E \cup F) \cap \tilde{F}^c\,) \ , \\
& bel_{\mathcal{E}}(\tilde{E} \cup \tilde{F}) + bel_{\underline{p}}(\,(E \cup F) \cap (\tilde{E} \cup \tilde{F})^c\,)\,] \Big) \ ,
\end{aligned}$$

By definition, the set $E \cup F$ contains all $\tilde{E}, \tilde{F} \in \mathcal{U}$ with $\tilde{E} \subseteq E$ and $\tilde{F} \subseteq F$, and all their unions $\tilde{E} \cup \tilde{F}$. Moreover, it contains no other $E' \in \mathcal{U}$. This follows from the fact that $E \cup F$ is non-convex and all focal elements $E_i \in \mathcal{E}$ are half-closed intervals of the real line. Therefore, every non-convex $E' \in \mathcal{U}$ must be a union of convex constituents, and these can only be contained in either $E$ or $F$. Hence, we find

$$\begin{aligned}
\underline{P}^*(E) + \underline{P}^*(F) &= \max_{\substack{\tilde{D} \in \mathcal{U}, \\ \tilde{D} \subseteq E \cup F}} \Big( \max[\, bel_{\underline{p}}(E \cup F)\,, bel_{\mathcal{E}}(\tilde{D}) + bel_{\underline{p}}(\,(E \cup F) \cap \tilde{D}^c\,)\,] \Big) \\
&= \underline{P}^*(E \cup F) \ .
\end{aligned}$$

**Step 2:** We proof the lemma by induction over the level of hierarchy $s$ of the unions of focal elements $E \in \mathcal{U}$. The level-1-sets cannot fulfil the condition of the Lemma because they contain no other set $E \in \mathcal{U}$ by definition. Hence, we start the induction with hierarchy level $s = 2$. Consider an arbitrary level-2-set $E^2_{i(2)}$ which is the union of two level-1-sets $E_*$, $E_{**}$ with $E_* \cap E_{**} = \emptyset$. Then,

$$\begin{aligned}
\nu^*_{i(2)} &= \underline{P}^*(E^2_{i(2)}) - \sum_{l \,|\, \substack{G_l \in \mathcal{G}^1 \\ G_l \subseteq E^2_{i(2)}}} \nu^*_l \\
&= \underline{P}^*(E^2_{i(2)}) - \sum_{l | G_l \subseteq E_*} \nu^*_l - \sum_{l' | G_{l'} \subseteq E_{**}} \nu^*_{l'} \;=\; \underline{P}^*(E^2_{i(2)}) - \underline{P}^*(E_*) - \underline{P}^*(E_{**}) \;,
\end{aligned}$$

where the last equality follows from Lemma 3.3. Since the level-1-sets $E_*$ and $E_{**}$ are disjoint, and it is $E^2_{i(2)} = E_* \cup E_{**}$, it follows from Step 1 of the proof that $\nu^*_{i(2)} = 0$.

**Step 3:** Let $s = r$. Let the proposition be true for all level-$s$-sets with a smaller hierarchy level $s < r$. Consider an arbitrary $r$-union $E^r_{i(r)} = E^{s'}_* \cup E^{s''}_{**}$ with $E^{s'}_* \cap E^{s''}_{**} = \emptyset$. Then,

$$\begin{aligned}
\nu^*_{i(r)} &= \underline{P}^*(E^r_{i(r)}) - \sum_{l \,|\, \substack{G_l \in \mathcal{G}^{r-1} \\ G_l \subseteq E^r_{i(r)}}} \nu^*_l \\
&= \underline{P}^*(E^r_{i(r)}) - \left( \sum_{\tilde{s}=2}^{r-1} \; \sum_{i(\tilde{s}) \,|\, \substack{E^{\tilde{s}}_{i(\tilde{s})} \subset E^r_{i(r)} \\ E^{\tilde{s}}_{i(\tilde{s})} \cap E^{s'}_* \neq \emptyset \\ E^{\tilde{s}}_{i(\tilde{s})} \cap E^{s''}_{**} \neq \emptyset}} \nu^*_{i(\tilde{s})} \right) - \sum_{l | G_l \subseteq E^{s'}_*} \nu^*_l - \sum_{l' | G_{l'} \subseteq E^{s''}_{**}} \nu^*_{l'} \;,
\end{aligned}$$

where the sum of the Möbius masses of the focal elements $G_l \subset E^r_{i(r)}$ has been divided into three subsums. We note that the first of these three subsums runs over the level-$\tilde{s}$ sets $E^{\tilde{s}}_{i(\tilde{s})}$ contained in $\mathcal{G}^{r-1}$ that intersect both $E^{s'}_*$ and $E^{s''}_{**}$. Since the latter two sets are disjoint, the intersect of their union with $E^{\tilde{s}}_{i(\tilde{s})}$ will also be disjoint. Since their union constitutes $E^r_{i(r)}$ which contains $E^{\tilde{s}}_{i(\tilde{s})}$, it is also

$$E^{\tilde{s}}_{i(\tilde{s})} = \left( E^{\tilde{s}}_{i(\tilde{s})} \cap E^{s'}_* \right) \cup \left( E^{\tilde{s}}_{i(\tilde{s})} \cap E^{s''}_{**} \right) \;.$$

Hence, $E^{\tilde{s}}_{i(\tilde{s})}$ is non-convex. Since the focal elements $E_i \in \mathcal{E}$ constitute half-closed intervals of the real line (see Lemma 3.1), $E^{\tilde{s}}_{i(\tilde{s})}$ cannot be a level-1-set $E^1 \in \mathcal{U}$. By the same token it needs to be a union of two disjoint sets $E^{\tilde{s}'}_*$ and $E^{\tilde{s}''}_{**} \in \mathcal{U}$. Since $\tilde{s} < r$, it is $\nu^*_{i(\tilde{s})} = 0$ by assumption. Hence, the first of the three subsums is identical to zero, and the expression simplifies to

$$\nu^*_{i(r)} \;=\; \underline{P}^*(E^r_{i(r)}) - \underline{P}^*(E^{s'}_{*}) - \underline{P}^*(E^{s''}_{**}) \,,$$

where we have used the result from Lemma 3.3. Since the sets $E^{s'}_{*}$ and $E^{s''}_{**}$ are disjoint, and it is $E^r_{i(r)} = E^{s'}_{*} \cup E^{s''}_{**}$, it follows from Step 1 of the proof that $\nu^*_{i(r)} = 0$. $\square$

**Proof of Lemma 3.5:**

**Step 1:** We begin the proof by considering the case of an arbitrary level-1-set $E^1_{i(1)}$. Since it contains no other set $E \in \mathcal{U}$, we have

$$\nu^*_{i(1)} = P^*(E^1_{i(1)}) - \sum_{l \,\middle|\, \substack{G_l \in \mathcal{G}^0 \\ G_l \subseteq E^1_{i(1)}}} \nu^*_l \;=\; \max[bel_{\mathcal{E}}(E^1_{i(1)}),\, bel_{\underline{p}}(E^1_{i(1)})] - \sum_{j | A_j \subseteq E^1_{i(1)}} \underline{p}_j \geq 0 \,.$$

It remains to consider the case of level-$s$-sets $E^s_{i(s)}$ with hierarchy level $s \geq 2$. Lemma 3.4 has established $\nu^*_{i(s)} = 0$ for arbitrary level-$s$-sets $E^s_{i(s)}$, $s \geq 2$, that can be separated into $E^s_{i(s)} = E^{s'}_{*} \cup E^{s''}_{**}$ with $E^{s'}_{*} \cap E^{s''}_{**} = \emptyset$.

**Step 2:** Consider the remaining case of an arbitrary level-$s$-set $E^s_{i(s)}$, $s \geq 2$, that cannot be separated into $E^s_{i(s)} = E^{s'}_{*} \cup E^{s''}_{**}$ with $E^{s'}_{*} \cap E^{s''}_{**} = \emptyset$. We know from Lemma 3.1 that each focal element $E_i \subseteq \mathcal{E}$ is a half-closed interval on the real line, i.e., $E_i = (x_{*,i}, x^*_i]$ for some $x_{*,i}, x^*_i \in \mathbb{R}$. Hence, $E^s_{i(s)} = \cup_{i=1}^m (x_{*,i}, x^*_i]$ will be a half-closed interval as well. Moreover, Property (II) in Lemma 3.1 implies that the lower and upper bounds of the focal intervals have the same weak order on the real line, i.e., for all $E_i$, $E_j$, we have $x_{*,i} \leq x_{*,j}$ if and only if $x^*_i \leq x^*_j$, and at least one inequality is strict. Hence, we can construct a strict order of the focal elements $E_1 < ... < E_m$ that are contained in $E^s_{i(s)}$. Use, e.g., the lower bound as primary ordering criterion. If $x_{*,i} = x_{*,j}$, then the upper bound will serve as a secondary criterion that assures a strict ordering of $E_i$ and $E_j$. $E_1$ is the element with the lowest bound $x_{*,1} = \inf_{x \in E^s_{i(s)}} x$ and $E_m$ the element with the largest bound $x^*_m = \max_{x \in E^s_{i(s)}} x$, respectively. Hence, $E^s_{i(s)} = (x_{*,1}, x^*_m]$. Define the sets $E_L := \cup_{i=1}^{m-1} E_i = (x_{*,1}, x^*_{m-1}]$, and $E_R := \cup_{i=2}^{m} E_i = (x_{*,2}, x^*_m]$, which are true subsets of $E^s_{i(s)}$. Then,

$$\nu^*_{i(s)} \;=\; \underline{P}^*(E^s_{i(s)}) - \sum_{l \,\middle|\, \substack{G_l \in \mathcal{G}^{s-1} \\ G_l \subseteq E^s_{i(s)}}} \nu^*_l \;=$$

$$\underline{P}^*(E^s_{i(s)}) \;-\; \sum_{\tilde{s}=2}^{s-1} \; \sum_{i(\tilde{s}) \,\middle|\, \substack{E^{\tilde{s}}_{i(\tilde{s})} \subset E^s_{i(s)} \\ E^{\tilde{s}}_{i(\tilde{s})} \nsubseteq E_L \\ E^{\tilde{s}}_{i(\tilde{s})} \nsubseteq E_R}} \nu^*_{i(\tilde{s})} \;-\; \sum_{l | G_l \subseteq E_L} \nu^*_l \;-\; \sum_{l' | G_{l'} \subseteq E_R} \nu^*_{l'} \;+\; \sum_{l^* | G_{l^*} \subseteq E_L \cap E_R} \nu^*_{l^*} \,,$$

where the sum over the Möbius masses of the focal elements $G_l \subset E^s_{i(s)}$ has been separated into three subsums. The first subsum runs over the Möbius masses of the level-$\tilde{s}$-sets $E^{\tilde{s}}_{i(\tilde{s})}$ with $\tilde{s} < s$ that are contained in $E^s_{i(s)}$, but neither in $E_L$ nor in $E_R$. Hence, these $E^{\tilde{s}}_{i(\tilde{s})}$ need to contain the sets $E_1$ and $E_s$. As a consequence of the ordering of the focal elements, $E^{\tilde{s}}_{i(\tilde{s})}$ has lower bound $x_{*,1}$ and upper bound $x^*_s$. Assuming that $E^{\tilde{s}}_{i(\tilde{s})}$ was convex. Then, $E^{\tilde{s}}_{i(\tilde{s})} = (x_{*,1}, x^*_s] = E^s_{i(s)}$. This is impossible, because $E^s_{i(s)}$ has been declared a level-$s$-set with $s > \tilde{s}$ (see Step 2 in Algorithm 3.2). Hence, the sets $E^{\tilde{s}}_{i(\tilde{s})}$ cannot be convex. Since $E^{\tilde{s}}_{i(\tilde{s})} \in \mathcal{G}$ is also a union of focal elements $E_i \in \mathcal{E}$, it needs to constitute a union of two disjoint subsets $E^{s'}_{*} \in \mathcal{U}$ and $E^{s''}_{**} \in \mathcal{U}$. It follows from Lemma 3.4 that $\nu^*_{i(\tilde{s})} = 0$, and the first of the three subsums is identical to zero. Hence,

$$\nu^*_{i(s)} \quad = \quad \underline{P}^*(E^s_{i(s)}) - \underline{P}^*(E_L) - \underline{P}^*(E_R) + \underline{P}^*(E_L \cap E_R) \ ,$$

where we used the result of Lemma 3.3.

**Step 3:** To simplify the notation, we set $E := E^s_{i(s)}$ in the following.

$$\begin{aligned}
\underline{P}^*(E_L) + \underline{P}^*(E_R) \quad &= \quad \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E_L} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E_L \cap \tilde{E}^c),\, bel_{\underline{p}}(E_L)] \Big) \ + \\
&\qquad \max_{\tilde{F}\in\mathcal{U},\, \tilde{F}\subseteq E_R} \Big( \max[bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(E_R \cap \tilde{F}^c),\, bel_{\underline{p}}(E_R)] \Big) \\
&= \quad \max_{\substack{\tilde{E},\, \tilde{F}\in\mathcal{U},\\ \tilde{E}\subseteq E_L\,,\, \tilde{F}\subseteq E_R}} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E_L \cap \tilde{E}^c),\, bel_{\underline{p}}(E_L)] \ + \\
&\qquad \max[bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(E_R \cap \tilde{F}^c),\, bel_{\underline{p}}(E_R)] \Big) \\
&= \quad \max_{\substack{\tilde{E},\, \tilde{F}\in\mathcal{U},\\ \tilde{E}\subseteq E_L\,,\, \tilde{F}\subseteq E_R}} \Big( \max[\, bel_{\underline{p}}(E_L) + bel_{\underline{p}}(E_R) \ , \\
&\qquad bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E_L \cap \tilde{E}^c) + bel_{\underline{p}}(E_R) \ , \\
&\qquad bel_{\underline{p}}(E_L) + bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(E_R \cap \tilde{F}^c) \ , \\
&\qquad bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E_L \cap \tilde{E}^c) + bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(E_R \cap \tilde{F}^c) \,] \Big) \ ,
\end{aligned}$$

where we have converted the sum of the individual max-operations between $bel_{\mathcal{D}}(\tilde{D}) + bel_{\underline{p}}(D \cap \tilde{D}^c)$ and $bel_{\underline{p}}(D)$, $D \in \{E_L, E_R\}$, to a max-operation over all four possible cases.

We now investigate each case separately. Assuming first that $\underline{P}^*(E_L) + \underline{P}^*(E_R) = bel_{\underline{p}}(E_L) + bel_{\underline{p}}(E_R)$. Then,

$$\begin{aligned} \underline{P}^*(E_L) + \underline{P}^*(E_R) &= bel_{\underline{p}}(E_L) + bel_{\underline{p}}(E_R) \\ &= bel_{\underline{p}}(E) + bel_{\underline{p}}(E_L \cap E_R) \\ &\leq \underline{P}^*(E) + \underline{P}^*(E_L \cap E_R)\ , \end{aligned}$$

where the second equality follows from the fact that $bel_{\underline{p}}$ is additive on disjoint subsets of $E$, if at least one of these subsets constitutes an element of the power set $\mathcal{P}(A_1, ..., A_k)$ (see proof of Lemma 3.3 for details). $E$, $E_L$, $E_R$, and $E_L \cap E_R$ are elements of $\mathcal{P}(A_1, ..., A_k)$ due to Assumption 3.1.

Assume now the second of the four cases in the max-operation.

$$\begin{aligned} \underline{P}^*(E_L) + \underline{P}^*(E_R) &= \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E_L} \Big( bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E_L \cap \tilde{E}^c) + bel_{\underline{p}}(E_R) \Big) \\ &= \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E_L} \Big( bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E_L \cap \tilde{E}^c) + bel_{\underline{p}}(E_R \cap E_L) + bel_{\underline{p}}(E_R - E_L) \Big) \\ &= \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E_L} \Big( bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(\,(E_L \cap \tilde{E}^c) \cup (E_R - E_L)\,) + bel_{\underline{p}}(E_R \cap E_L) \Big) \\ &= \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E_L} \Big( bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(\,(E_L \cup E_R) \cap \tilde{E}^c\,) + bel_{\underline{p}}(E_R \cap E_L) \Big) \end{aligned}$$

where we have used again the additivity of $bel_{\underline{p}}$ for the disjoint sets considered here. The last equality follows from the fact that $\tilde{E} \subseteq E_L$, and therefore $\tilde{E}^c \supset E_R - E_L$. Inserting $E = E_L \cup E_R$, we find

$$\begin{aligned} \underline{P}^*(E_L) + \underline{P}^*(E_R) &\leq \max_{\tilde{E}\in\mathcal{U},\, \tilde{E}\subseteq E} \Big( bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E \cap \tilde{E}^c) + bel_{\underline{p}}(E_R \cap E_L) \Big) \\ &\leq \underline{P}^*(E) + \underline{P}^*(E_L \cap E_R)\ . \end{aligned}$$

By a completely analogous argument, it is found for the third case in the max-operation

$$\begin{aligned} \underline{P}^*(E_L) + \underline{P}^*(E_R) &= \max_{\tilde{F}\in\mathcal{U},\, \tilde{F}\subseteq E_R} \Big( bel_{\underline{p}}(E_L) + bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(E_R \cap \tilde{F}^c) \Big) \\ &\leq \underline{P}^*(E) + \underline{P}^*(E_L \cap E_R)\ . \end{aligned}$$

Consider finally the remaining case in the max-operation.

$$\begin{aligned} \underline{P}^*(E_L) + \underline{P}^*(E_R) &= \max_{\substack{\tilde{E},\, \tilde{F}\in\mathcal{U}, \\ \tilde{E}\subseteq E_L\,,\ \tilde{F}\subseteq E_R}} \Big( bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(E_L \cap \tilde{E}^c) + bel_{\mathcal{E}}(\tilde{F}) + bel_{\underline{p}}(E_R \cap \tilde{F}^c) \Big) \\ &= \max_{\substack{\tilde{E},\, \tilde{F}\in\mathcal{U}, \\ \tilde{E}\subseteq E_L\,,\ \tilde{F}\subseteq E_R}} \Big( bel_{\mathcal{E}}(\tilde{E}) + bel_{\mathcal{E}}(\tilde{F}) \\ &\qquad + bel_{\underline{p}}(\,(E_L \cap E_R) \cap \tilde{E}^c\,) + bel_{\underline{p}}(\,(E_L \cap E_R^c) \cap \tilde{E}^c\,) \\ &\qquad + bel_{\underline{p}}(\,E_R \cap (\tilde{F}^c \cap \tilde{E})\,) + bel_{\underline{p}}(\,E_R \cap (\tilde{F}^c \cap \tilde{E}^c)\,) \Big) \end{aligned}$$

$$
\begin{aligned}
= \max_{\substack{\tilde{E},\,\tilde{F}\in\mathcal{U},\\ \tilde{E}\subseteq E_L\,,\ \tilde{F}\subseteq E_R}} \Big( & bel_{\mathcal{E}}(\tilde{E}) + bel_{\mathcal{E}}(\tilde{F}) \\
& + bel_{\underline{p}}(\,((E_L \cap E_R) \cap \tilde{E}^c) \cup (E_R \cap (\tilde{F}^c \cap \tilde{E}))\,) \\
& + bel_{\underline{p}}(\,(E_R \cap (\tilde{F}^c \cap \tilde{E}^c)) \cup ((E_L \cap E_R^c) \cap \tilde{E}^c)\,)\Big) \ ,
\end{aligned}
$$

where we have used again the additivity of $bel_{\underline{p}}$ for the disjoint sets considered here. By noting that $E_R^c \subseteq \tilde{F}^c$, we find

$$
\begin{aligned}
(E_R \cap (\tilde{F}^c \cap \tilde{E}^c)) \cup ((E_L \cap E_R^c) \cap \tilde{E}^c) &= (E_R \cap (\tilde{F}^c \cap \tilde{E}^c)) \cup ((E_L \cap E_R^c) \cap (\tilde{E}^c \cap \tilde{F}^c)) \\
&= (E_R \cup (E_L \cap E_R^c)) \cap (\tilde{E}^c \cap \tilde{F}^c) \\
&= (E_R \cup E_L) \cap (\tilde{E} \cup \tilde{F})^c\,.
\end{aligned}
$$

Moreover, by noting that $(\tilde{F}^c \cap \tilde{E}) \subseteq E_L$, we also find

$$
\begin{aligned}
((E_L \cap E_R) \cap \tilde{E}^c) \cup (E_R \cap (\tilde{F}^c \cap \tilde{E})) &= ((E_L \cap E_R) \cap \tilde{E}^c) \cup ((E_L \cap E_R) \cap (\tilde{F}^c \cap \tilde{E})) \\
&= (E_L \cap E_R) \cap (\tilde{E}^c \cup (\tilde{F}^c \cap \tilde{E})) \\
&= (E_L \cap E_R) \cap (\tilde{E}^c \cup \tilde{F}^c) \\
&= (E_L \cap E_R) \cap (\tilde{E} \cap \tilde{F})^c\,.
\end{aligned}
$$

By inserting these two identities into the expression for $\underline{P}^*(E_L) + \underline{P}^*(E_R)$, we find for the remaining case of the max-operation

$$
\begin{aligned}
\underline{P}^*(E_L) + \underline{P}^*(E_R) = \max_{\substack{\tilde{E},\,\tilde{F}\in\mathcal{U},\\ \tilde{E}\subseteq E_L\,,\ \tilde{F}\subseteq E_R}} \Big( & bel_{\mathcal{E}}(\tilde{E}) + bel_{\mathcal{E}}(\tilde{F}) \\
& + \ bel_{\underline{p}}(\,(E_R \cup E_L) \cap (\tilde{E} \cup \tilde{F})^c\,) + bel_{\underline{p}}(\,(E_L \cap E_R) \cap (\tilde{E} \cap \tilde{F})^c\,)\Big) \ .
\end{aligned}
$$

Since $bel_{\mathcal{E}}$ is a belief function, it is in particular two-monotone (see Appendix C). Hence,

$$
\begin{aligned}
\underline{P}^*(E_L) + \underline{P}^*(E_R) &\leq \max_{\substack{\tilde{E},\,\tilde{F}\in\mathcal{U},\\ \tilde{E}\subseteq E_L\,,\ \tilde{F}\subseteq E_R}} \Big( bel_{\mathcal{E}}(\tilde{E} \cup \tilde{F}) + bel_{\mathcal{E}}(\tilde{E} \cap \tilde{F}) \\
&\qquad + \ bel_{\underline{p}}(\,(E_R \cup E_L) \cap (\tilde{E} \cup \tilde{F})^c\,) + bel_{\underline{p}}(\,(E_L \cap E_R) \cap (\tilde{E} \cap \tilde{F})^c\,)\Big) \\
&\leq \max_{\substack{\tilde{C},\,\tilde{D}\in\mathcal{U},\\ \tilde{C}\subseteq E\,,\ \tilde{D}\subseteq E_L\cap E_R}} \Big( bel_{\mathcal{E}}(\tilde{C}) + bel_{\mathcal{E}}(\tilde{D}) \\
&\qquad + \ bel_{\underline{p}}(\,(E \cap \tilde{C}^c\,) + bel_{\underline{p}}(\,(E_L \cap E_R) \cap \tilde{D}^c\,)\Big) \\
&\leq \ \underline{P}^*(E) + \underline{P}^*(E_L \cap E_R)\ ,
\end{aligned}
$$

where the second inequality follows from the fact that $bel_{\mathcal{E}}(\tilde{E} \cap \tilde{F}) = bel_{\mathcal{E}}(\tilde{D})$, where $\tilde{D} \in E_L \cap E_R$ is the union of all focal elements contained in $\tilde{E} \cap \tilde{F}$. Hence, we also have $\tilde{D}^c \supseteq (\tilde{E} \cap \tilde{F})^c$, and therefore $bel_{\underline{p}}(\,(E_L \cap E_R) \cap \tilde{D}^c\,) \geq bel_{\underline{p}}(\,(E_L \cap E_R) \cap (\tilde{E} \cap \tilde{F})^c\,)$.

We conclude that $\underline{P}^*(E) + \underline{P}^*(E_L \cap E_R) \geq \underline{P}^*(E_L) + \underline{P}^*(E_R)$ in any case. Together with the result from Step 2 of the proof, it follows that $\nu^*_{i(s)} \geq 0$. $\square$

**Proof of Theorem 3.2:**

**Step 1:** We show first that $\underline{P}^*$ is a belief function. Shafer (1976) has proved for finite spaces $\Omega_n$ that a set function $\underline{P}^* : \mathcal{P}(\Omega_n) \to [0,1]$ is an $\infty$-monotone belief function, if and only if its Möbius inverse $\nu^* : \mathcal{P}(\Omega_n) \to \mathbb{R}$ fulfils $\nu^*(A) \geq 0$ for all $A \in \mathcal{P}(\Omega_n)$. The property of $\infty$-monotonicity extends to an infinite field $\mathcal{R}$, if the belief function can be fully described on a finite subfield $\mathcal{P}(\Omega_n)$, i.e.,

$$\forall\, B \in \mathcal{R} \quad \exists\, A \in \mathcal{P}(\Omega_n), \;\; A \subseteq B \quad \text{with} \quad \mathit{bel}(B) = \mathit{bel}(A)\ .$$

This can be seen easily for the case of 2-monotonicity. Consider arbitrary sets $B_1$, $B_2 \in \mathcal{R}$ and let $A_1$, $A_2$ be their subsets in the finite field $\mathcal{P}(\Omega_n)$ with $\mathit{bel}(B_1) = \mathit{bel}(A_1)$ and $\mathit{bel}(B_2) = \mathit{bel}(A_2)$. Then,

$$\begin{aligned} \mathit{bel}(B_1) + \mathit{bel}(B_2) = \mathit{bel}(A_1) + \mathit{bel}(A_2) \;&\leq\; \mathit{bel}(A_1 \cup A_2) + \mathit{bel}(A_1 \cap A_2) \\ &\leq\; \mathit{bel}(B_1 \cup B_2) + \mathit{bel}(B_1 \cap B_2)\ , \end{aligned}$$

where the first inequality follows from 2-monotonicity on the finite field, and the second inequality from $B_1 \cup B_2 \supseteq A_1 \cup A_2$ and $B_1 \cap B_2 \supseteq A_1 \cap A_2$. The same result for the case of $\infty$-monotonicity can be obtained by induction over the number $n$ of unions $\cup_{i=1} B_i$.

In our case, the set function $\underline{P}^* : \mathcal{R} \to [0,1]$ is fully described on the power set $P(A_1, ..., A_k)$ of the partition $\{A_1, ..., A_k\}$, since it contains all focal elements $G$ of the Möbius inverse $\nu^* : \mathcal{R} \to [0,1]$ with Möbius assignments $\nu^*(G) \neq 0$ (see Lemma 3.3). Hence, it suffices to show that all focal elements $G$ collected in the random set $(\mathcal{G}, \nu^*)$ have Möbius masses $\nu^*(G) > 0$.

By definition of Algorithm 3.2, all focal elements $G \in \mathcal{G}$ are either atoms $A_j$ of the partition $\{A_1, ..., A_k\}$ or unions $E \in \mathcal{U}$ of focal elements $E_i \in \mathcal{E}$. The atoms $A_j$ contained in $(\mathcal{G}, \nu^*)$ carry Möbius masses $\nu^*(A_j) = \underline{p}_j > 0$ (see Step 1 of the Algorithm). Moreover, we have established in Lemma 3.5 that it is also $\nu^*(E) \geq 0$ for all $E \in \mathcal{U}$. Therefore, all focal elements $G \in \mathcal{G}$ must carry Möbius masses $\nu^*(G) > 0$. We conclude that $\underline{P}^* : \mathcal{R} \to [0,1]$ is a belief function.

**Step 2:** We show next that $\underline{P}^* : \mathcal{R} \to [0,1]$ is the lower envelope of $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$. Since $\underline{P}^*$ is a belief function, it is in particular a coherent lower probability (see Appendix C). It follows from the Lower Envelope Theorem of Walley (1991, Theorem 3.3.3) that $\underline{P}^*$ is the lower envelope of a convex set of probabilities $\mathcal{P}^*$, i.e.,

$$\forall\, A \in \mathcal{R} \qquad \underline{P}^*(A) = \inf_{P \in \mathcal{P}^*} P(A)\ .$$

We now from Lemma 3.2 that $\underline{P}^*$ is also a lower bound of $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$. Therefore, $\Gamma(\underline{SF}, \overline{SF}, \underline{p}) \subseteq \mathcal{P}^*$. On the other hand, we find for each $P \in \mathcal{P}^*$,

$$\begin{aligned} P(A) \;&\geq\; \max_{\tilde{E} \in \mathcal{U},\, \tilde{E} \subseteq A} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(A \cap \tilde{E}^c),\, bel_{\underline{p}}(A)] \Big) \;\geq\; bel_{\underline{p}}(A) \\ P(A) \;&\geq\; \max_{\tilde{E} \in \mathcal{U},\, \tilde{E} \subseteq A} \Big( \max[bel_{\mathcal{E}}(\tilde{E}) + bel_{\underline{p}}(A \cap \tilde{E}^c),\, bel_{\underline{p}}(A)] \Big) \;\geq\; bel_{\mathcal{E}}(E) \;=\; bel_{\mathcal{E}}(A)\ , \end{aligned}$$

where $E \subseteq A$ is the union of all focal elements $E_i \in \mathcal{E}$ that are subsets of $A$. As a consequence, $P \in \mathcal{P}$ is contained in the p-box $\Gamma(\underline{SF}, \overline{SF})$ as well as in the $\varepsilon$-contamination model $\Gamma(\underline{p})$, and hence also in $\Gamma(\underline{SF}, \overline{SF}, \underline{p}) = \Gamma(\underline{SF}, \overline{SF}) \cap \Gamma(\underline{p})$. Therefore, we also find $\mathcal{P}^* \subseteq \Gamma(\underline{SF}, \overline{SF}, \underline{p})$. It follows that $\mathcal{P}^* = \Gamma(\underline{SF}, \overline{SF}, \underline{p})$, and $\underline{P}^* : \mathcal{R} \to [0,1]$ is the lower envelope of $\Gamma(\underline{SF}, \overline{SF}, \underline{p})$. □

### Proof of Lemma 4.1:

We note first that $bel(\cdot||_*\hat{y}) : \mathcal{A}(\Omega) \to [0,1]$ is indeed a belief function, since $\sum_{i=1}^n \nu_{*,i} = 1$, and $\nu_{*,i} \geq 0$ for all $1 \leq i \leq n$. The latter is true because $\mathcal{L}_{*,i} \geq 0$ (the likelihood cannot be negative), $\nu_i > 0$ ($bel$ is a belief function), and $\sum_{i=1}^n \mathcal{L}_{*,i}\, \nu_i = \underline{C_P}(\mathcal{L}) > 0$ by assumption.

It remains to show that $\underline{P}(A||_*\hat{y}) = bel(A||_*\hat{y})$ for all sets $A \in \mathcal{A}(\Omega)$. Consider an arbitrary such $A$. Then,

$$\underline{P}(A||_*\hat{y}) := \frac{\underline{C_P}(\mathcal{L}_A)}{\underline{C_P}(\mathcal{L})} = \frac{\sum_{E_j \subseteq A} \mathcal{L}_{*,j}\, \nu_j}{\sum_{i=1}^n \mathcal{L}_{*,i}\, \nu_i} = \sum_{E_j \subseteq A} \nu_{*,j} \;=\; bel(A||_*\hat{y})\ . \qquad \square$$

**Proof of Lemma 4.2:**

We need to show that $\overline{P}(A||^*\hat{y}) = pl(A||^*\hat{y})$ for all sets $A \in \mathcal{A}(\Omega)$. Consider an arbitrary such $A$. Let $1 \leq i \leq n$ be the running index over the focal elements $E_i \in \mathcal{E}$. Let $1 \leq j \leq k$ be the running index over the atoms $A_j$ of the partition $\{A_1, ..., A_k\}$, the constant likelihood values $\mathcal{L}^*_j$ on $A_j$, and the associated $j$-cuts $C_j$ of the likelihood function. Define $\mathcal{L}^*_0 = 0$. Then,

$$\begin{aligned}
\overline{P}(A||^*\hat{y}) := \frac{\overline{C}_{\overline{P}}(\mathcal{L}'_A)}{\overline{C}_{\overline{P}}(\mathcal{L}')} \quad &= \quad \frac{\sum\limits_{i\,|\,E_i \cap A \neq \emptyset} \sup\limits_{\theta \in E_i \cap A} \mathcal{L}'(\theta;\hat{y}) \ \nu_i}{\sum_{i'=1}^{n} \sup\limits_{\theta \in E_{i'}} \mathcal{L}'(\theta;\hat{y}) \ \nu_{i'}} \\
&= \quad \frac{\sum\limits_{i\,|\,E_i \cap A \neq \emptyset} \sup\limits_{j\,|\,A_j \cap (E_i \cap A) \neq \emptyset} \mathcal{L}^*_j \ \nu_i}{\sum_{i'=1}^{n} \sup\limits_{j'\,|\,A_{j'} \cap E_{i'} \neq \emptyset} \mathcal{L}^*_{j'} \ \nu_{i'}} \\
&= \quad \frac{\sum\limits_{i\,|\,E_i \cap A \neq \emptyset} \left( \sum_{j\,|\,C_j \cap (E_i \cap A) \neq \emptyset} \mathcal{L}^*_j - \mathcal{L}^*_{j-1} \right) \nu_i}{\sum_{i'=1}^{n} \left( \sum_{j'\,|\,C_{j'} \cap E_{i'} \neq \emptyset} \mathcal{L}^*_{j'} - \mathcal{L}^*_{j'-1} \right) \nu_{i'}} \\
&= \quad \frac{\sum_{ij\,|\,(C_j \cap E_i) \cap A \neq \emptyset} \left( \mathcal{L}^*_j - \mathcal{L}^*_{j-1} \right) \nu_i}{\sum_{i'=1}^{n} \mathcal{L}^*_{l(i')} \ \nu_{i'}} \\
&= \quad \sum_{ij\,|\,E_{ij} \cap A \neq \emptyset} \frac{\left( \mathcal{L}^*_j - \mathcal{L}^*_{j-1} \right) \nu_i}{\sum_{i'=1}^{n} \mathcal{L}^*_{l(i')} \ \nu_{i'}} \quad = \quad pl(A||^*\hat{y}) \ . \qquad \square
\end{aligned}$$

**Proof of Theorem 4.1:**

**1. Show $\underline{P}(\cdot||_*\hat{y}) = bel(\cdot||_*\hat{y})$:** Since $\mathcal{L}'_{i,*} := \inf_{\theta \in E_i} \mathcal{L}'(\theta|\hat{y}) \ \nu_i$ for $1 \leq i \leq n$ (see Definition 4.16), it is $0 < \underline{C}(\mathcal{L}') = \sum_{i=1}^{n} \mathcal{L}_{*,i} \, \nu_i$ by assumption. Hence, the construction of the random set $(\mathcal{E}, \nu)_{||_*\hat{y}}$ associated with the belief function $bel(\cdot||_*\hat{y})$ in Lemma 4.1 is defined (Denominator larger than zero). It follows from Lemma 4.1 that $\underline{P}(\cdot||_*\hat{y}) = bel(\cdot||_*\hat{y})$, where $\underline{P}(\cdot||_*\hat{y})$ is given by Definition 4.15.

**2. Show $\overline{P}(\cdot||^*\hat{y}) = pl(\cdot||^*\hat{y})$:** Since the likelihood function $\mathcal{L}'(\cdot|\hat{y})$ is assumed to be constant on the atoms of the partition $\{A_1, ..., A_k\}$ of $\Omega$, it is $\sup_{\theta \in A} \mathcal{L}'(\theta|\hat{y}) \ \nu_i = \max_{j\,|\,A_j \cap A \neq \emptyset} \mathcal{L}^*_j$ for arbitrary $A \in \mathcal{A}(\Omega)$ (recall Definition 4.18 of $\mathcal{L}^*_j$). In particular, it is $\sup_{\theta \in E_i} \mathcal{L}'(\theta|\hat{y}) \ \nu_i = \mathcal{L}^*_{l(i)}$ for $1 \leq i \leq n$, since the latter constitutes the supremum likelihood value that can be achieved on any atom intersecting $E_i$ (see Algorithm 4.1). Hence, $0 < \underline{C}(\mathcal{L}') \leq \overline{C}(\mathcal{L}') = \sum_{i=1}^{n} \mathcal{L}^*_{l(i)} \, \nu_i$, and the construction of the random set $(\mathcal{E}, \nu)_{||^*\hat{y}}$ associated with the belief function $bel(\cdot||^*\hat{y})$ in Algorithm 4.1 is defined (Denominator larger than zero).

Moreover, $\mathcal{L}'(\cdot|\hat{y})$ fulfils the condition of Lemma 4.2. Hence, it follows from Lemma 4.2 that $\overline{P}(\cdot||^*\hat{y}) = pl(\cdot||^*\hat{y})$, where $\overline{P}(\cdot||^*\hat{y})$ is given by Definition 4.15.

**3. Show Equation (4.19):** The prior belief function is $\infty$-monotone. It follows from the theorem of Wasserman and Kadane (1990) that

$$\underline{P}(A|\hat{y}) \quad = \quad \frac{\underline{C}_{\underline{P}}(\mathcal{L}_A)}{\underline{C}_{\underline{P}}(\mathcal{L}_A) + \overline{C}_{\overline{P}}(\mathcal{L}_{A^c})} \qquad \text{for all} \quad A \in \mathcal{A}(\Omega)$$

According to Definition (4.15), it is

$$\underline{C}_{\underline{P}}(\mathcal{L}_A) = \underline{C}(\mathcal{L}')\,\underline{P}(A||_*\hat{y}), \qquad \overline{C}_{\overline{P}}(\mathcal{L}_{A^c}) = \overline{C}(\mathcal{L}')\,\overline{P}(A||^*\hat{y}) \qquad \text{for all} \quad A \in \mathcal{A}(\Omega)\ .$$

Equation (4.19) follows from these identities, and Part 1 and 2 of the proof. □

**Proof of Lemma 5.1:**

The following properties hold in general for inverse images $f^{-1}(B) = \{x\,|\,f(x) \in B\}$ (Billingsley, 1995, A8): $f^{-1}(\cup_\theta B_\theta) = \cup_\theta f^{-1}(B_\theta)$, $f^{-1}(\cap_\theta B_\theta) = \cap_\theta f^{-1}(B_\theta)$ for arbitrary (possibly uncountable) unions and intersections, and $B_1 \cap B_2 = \emptyset \;\Rightarrow\; f^{-1}(B_1) \cap f^{-1}(B_2) = \emptyset$.

**Part (a):** $\mathcal{F} \subseteq \mathcal{R}^n$ follows from the definition of $\mathcal{F}$, the measurability of $f$ and the fact that $\mathrm{Rg}(f)$ is a Borel set. $\mathcal{F}$ is a $\sigma$-field on $\mathbb{R}^n$, if

1. $\mathbb{R}^n \in \mathcal{F}$: True, since $\mathbb{R}^n = f^{-1}(\mathrm{Rg}(f))$ and $\mathrm{Rg}(f) \in \mathcal{R}^m \cap \mathrm{Rg}(f)$.

2. $A \in \mathcal{F} \;\Rightarrow\; A^c \in \mathcal{F}$: Consider an arbitrary $A \in \mathcal{F}$. By definition of $\mathcal{F}$, it exists $B \in \mathcal{R}^m \cap \mathrm{Rg}(f)$ with $A = f^{-1}(B)$. Define $B' = \mathrm{Rg}(f) - B$, $B' \in \mathcal{R}^m \cap \mathrm{Rg}(f)$. It is $f^{-1}(B \cup B') = f^{-1}(B) \cup f^{-1}(B') = A \cup f^{-1}(B')$. Since $B \cap B' = \emptyset$, $A \cap f^{-1}(B') = \emptyset$. Since $f^{-1}(B \cup B') = \mathbb{R}^n$, $A^c = f^{-1}(B')$. Thus, $A^c \in \mathcal{F}$.

3. For any collection $A_i \in \mathcal{F}$, it is $\cup_i A_i \in \mathcal{F}$: For each $A_i$ in $\mathcal{F}$, it exists $B_i \in \mathcal{R}^m \cap \mathrm{Rg}(f)$ with $A_i = f^{-1}(B_i)$. Thus, $\cup_i A_i = \cup_i f^{-1}(B_i) = f^{-1}(\cup_i B_i)$. Since for any collection $B_i \in \mathcal{R}^m \cap \mathrm{Rg}(f)$ it is also $\cup_i B_i \in \mathcal{R}^m \cap \mathrm{Rg}(f)$, we have $\cup_i A_i \in \mathcal{F}$.

**Part (b):** $\{f^{-1}(y)\,|\,y \in \mathrm{Rg}(f)\}$ is the set of atoms of $\mathcal{F}$, if

1. $f^{-1}(y) \cap f^{-1}(y') = \emptyset$ for $y \neq y'$: True, since $\{y\} \cap \{y'\} = \emptyset$, if $y \neq y'$.

2. For all $A \in \mathcal{F}$, there exists a set $B \in \mathcal{P}(\mathbb{R}^m) \cap \mathrm{Rg}(f)$ with $A = \cup_{y \in B} f^{-1}(y)$: True, since by definition of $\mathcal{F}$, there exists a set $B \in \mathcal{R}^m \cap \mathrm{Rg}(f)$ for all $A \in \mathcal{F}$ with $A = f^{-1}(B) = f^{-1}(\cup_{y \in B}\, y) = \cup_{y \in B} f^{-1}(y)$.

**Part (c):** Note that $P_{X|\mathcal{F}}(A) := P_Y(f(A))$ is always defined on $\mathcal{F}$, since $\mathrm{Rg}(f)$ is a Borel set, and $\forall\, A \in \mathcal{F}\;\; f(A) \in \mathcal{R}^m \cap \mathrm{Rg}(f)$ by definition of $\mathcal{F}$. $P_{X|\mathcal{F}}$ is a countably additive probability measure on $\mathcal{F}$, if

1. For all $A \in \mathcal{F}\;\; P_{X|\mathcal{F}}(A) \geq 0$: True, by definition of $P_{X|\mathcal{F}}$, and the fact that $P_Y$ is a probability measure.
2. $P_{X|\mathcal{F}}(\mathbb{R}^n) = 1$: True, since $f(\mathbb{R}^n) = \mathrm{Rg}(f)$ is a Borel set, and $P_Y(\mathrm{Rg}(f)) = 1$.
3. For any collection $A_i \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$, it is $P_{X|\mathcal{F}}(\cup_i A_i) = \sum_i P_{X|\mathcal{F}}(A_i)$:
   We have $P_{X|\mathcal{F}}(\cup_i A_i) := P_Y(f(\cup_i A_i)) = P_Y(\cup_i f(A_i))$. It follows from Part (b) that $A_i = f^{-1}(f(A_i))$. Hence, we have $A_i = f^{-1}(\cup_{y \in f(A_i)} y) = \cup_{y \in f(A_i)} f^{-1}(y)$. Then, $A_i \cap A_j = \emptyset$ for $i \neq j$ implies $f(A_i) \cap f(A_j) = \emptyset$ for $i \neq j$. Thus, $P_Y(\cup_i f(A_i)) = \sum_i P_Y(f(A_i)) =: \sum_i P_{X|\mathcal{F}}(A_i)$, since $P_Y$ is a countably additive probability measure. □

**Proof of Theorem 5.1:**

**Part (a)**: Consider an arbitrary probability $P_Y \in f(\Gamma_X(bel_{\mathcal{E}}))$. There exists a probability $P_X \in \Gamma_X(bel_{\mathcal{E}})$ with $P_Y(B) = P_X(f^{-1}(B))$ for all $B \in \mathcal{R}^m$. For arbitrary $B \in \mathcal{R}^m$, we have

$$\begin{aligned} P_Y(B) &= P_X(f^{-1}(B)) \;\geq\; bel_{\mathcal{E}}(f^{-1}(B)) \;=\; \sum_{E_i \subseteq f^{-1}(B)} \nu_i \\ &= \sum_{f(E_i) \subseteq B} \nu_i \;=\; bel_{f(\mathcal{E})}(B)\,. \quad \text{Thus, } P_Y \in \Gamma_Y(bel_{f(\mathcal{E})})\,. \end{aligned}$$

**Part (b)**: Consider an arbitrary probability $P_{X|\mathcal{F}} \in f^{-1}(\Gamma_Y(bel_{f(\mathcal{E})}))$. There exists a probability $P_Y \in \Gamma_Y(bel_{f(\mathcal{E})})$ with $P_{X|\mathcal{F}}(A) = P_Y(f(A))$ for all $A \in \mathcal{F}$. For arbitrary $A \in \mathcal{F}$, we have

$$\begin{aligned} P_{X|\mathcal{F}}(A) = P_Y(f(A)) \;\geq\; bel_{f(\mathcal{E})}(f(A)) &= \sum_{f(E_i) \subseteq f(A)} \nu_i \\ &= \sum_{E_i \subseteq f^{-1}(f(A))} \nu_i \;=\; bel_{\mathcal{E}}(A)\,, \end{aligned}$$

since $A = f^{-1}(f(A))$ by definition of $\mathcal{F}$. Thus, $P_{X|\mathcal{F}} \in \Gamma_{X|\mathcal{F}}(bel_{\mathcal{E}})$. □

# Bibliography

Allan, R., N. Nicholls, P. Jones, and I. Butterworth (1991). A further extension of the Tahiti-Darwin SOI, early SOI results and Darwin pressure. *Journal of Climate 4*, 743–749. 31

Allen, M. R. and W. J. Ingram (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature 419*, 224–232. 3

Allen, M. R. and P. A. Stott (2003). Estimating signal amplitudes in optimal fingerprinting. Part I: Theory. *Climate Dynamics 21*, 477–491. 12

Allen, M. R. and S. F. B. Tett (1999). Checking for model consistency in optimal fingerprinting. *Climate Dynamics 15*, 419–434. 4, 12, 31

Ammann, C. M., G. A. Meehl, and W. M. Washington (2003). A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophysical Research Letters 30(12)*, 1657. 27

Anderson, T. L., C. R. J., S. E. Schwartz, R. Knutti, O. Boucher, H. Rodhe, and J. Heintzenberg (2003). Climate forcing by aerolsols - a hazy picture. *Science 300*, 1103–1104. 81

Andronova, N. G., E. V. Rozanov, F. Yang, M. E. Schlesinger, and G. L. Stenchikov (1999). Radiative forcing by volcanic aerosols from 1850 to 1994. *Journal of Geophysical Research 104(D14)*, 16,807–16,826. 27, 28, 58

Andronova, N. G. and M. E. Schlesinger (2000). Causes of global temperature changes during the 19th and 20th centuries. *Geophysical Research Letters 27(14)*, 2137–2140. 10, 28, 58, 59

Andronova, N. G. and M. E. Schlesinger (2001). Objective estimation of the probability density function for climate sensitivity. *Journal of Geophysical Research 106(D19)*, 22605–22611. 11, 46, 51, 56, 58, 59, 68, 70, 74, 81, 87, 88, 90, 91, 92, 102, 121, 124, 151

Annan, J. D. and J. C. Hargreaves (2003). Some practical aspects of quantifying uncertainty. Presentation at the joint EGU/AGU assembly, 06.-11. April 2003, Nice. 95, 96

Anscombe, F. J. and R. J. Aumann (1963). A definition of subjective probability. *The Annals of Mathematics and Statistics 34*, 199–205. 144, 145

Arrow, K. J. and L. Hurwicz (1972). An optimality criterion for decision making under ignorance. In C. F. Carter and J. L. Ford (Eds.), *Uncertainty and Expectations in Economics.* Basil Blackwell, Oxford. 144

Augustin, T. (1998). *Optimale Tests bei Intervallwahrscheinlichkeit.* Vandenhoeck & Ruprecht, Göttingen. XLVI

Basu, S. and A. DasGupta (1990). Bayesian analysis under distribution bands. Technical Report 90–48, Department of Statistics, Purdue University. 61

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transaction of the Royal Society London A53*, 370–418. 4

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis.* Series in Statistics. Springer, New York. 11, 21

Berger, J. O. (1993). An overview of robust Bayesian analysis. Technical Report #93-53C, Purdue University. 70, 97, XLVI

Berliner, L. M., R. A. Levine, and D. J. Shea (2000). Bayesian climate change assessment. *Journal of Climate 13*, 3805–3820. 4, 11, 12

Billingsley, P. (3rd edition, 1995). *Probability and Measure.* Wiley Interscience, New York. 61, 62, XXXIX, LXIX

Brooke, A., D. Kendrick, and A. Meeraus (1992). *GAMS: A User's Guide, Release 2.25.* Scientific Press, San Francisco. 39, 116

Bruckner, T., G. Petschel-Held, F. L. Tóth, H. M. Füssel, C. Helm, M. Leimbach, and H. J. Schellnhuber (1999). Climate change decision support and the tolerable windows approach. *Environmental Modeling and Assessment 4*, 217–234. 146

Budyko, M. I. (1969). The effect of solar radiation variations on the climate of the earth. *Tellus 21*, 611–619. 10

Casadesus-Masanell, R., P. Klibanoff, and E. Ozdenoren (2000). Maxmin expected utility over Savage acts with a set of priors. *Journal of Economic Theory 92*, 35–66. 144

Chateauneuf, A. and J. Y. Jaffray (1989). Some characterization of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences 17*, 263–283. 53, XLIX, L, LI

Chevé, M. and R. Congar (2000). Optimal pollution control under imprecise environmental risk and irreversibility. *Risk, Decision and Policy 5*, 151–164. 145

Chevé, M. and R. Congar (2003). Managing environmental risks under scientific uncertainty and controversy. Paper presented at the EARE 2003 conference, Bilbao, Spain. 6

Choquet, G. (1953). Theory of capacities. In *Ann. Inst. Fourier*, Volume 5, No. 4, pp. 131–295. Institut Fourier, Grenoble. 50, XLVI, XLVIII

Chrisman, L. (1995). Incremental conditioning of lower and upper probabilities. *International Journal of Approximate Reasoning 13*, 1–25. 106, 107, 109, 110, 149

Collins, M., S. F. B. Tett, and C. Cooper (2001). The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics 17*, 61–81. 29, 30, 34, 38

Cook, J. and E. J. Highwood (2004). Climate response to tropospheric absorbing aerosols in an intermediate general-circulation model. *Quarterly Journal of the Royal Meteorological Society 130*, 175–191. V

Couso, I., S. Moral, and P. Walley (2000). A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy 5*, 165–181. 100, 141

Cozman, F. (1999a). A brief introduction to the theory of sets of probability measures. Upgraded version of technical report CMU-RI-TR 97-24, Carnegie Mellon University. URL: www-2.cs.cmu.edu/∼qbayes/Tutorial/. 50

Cozman, F. G. (1999b). Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions. *Journal of Computational and Graphical Statistics 8*, 824–838. 150

Cubasch, U. and G. Meehl (2001). Projections of future climate change. In J. Houghton and Y. Ding (Eds.), *Climate Change 2001: The Scientific Basis*, Contribution of Working Group I to the Third Assessment Report of the IPCC, pp. 525–582. Cambridge University Press, Cambridge. 14, 21, 22, 54, 130, 135, 136, 137, VI, XX, XXVIII

de Cooman, G. (1997). Possibility theory: I The measure- and integral-theoretic groundwork, II Conditional possibility, III Possibilistic independence. *International Journal of General Systems 25*, 291–371. LII

de Cooman, G. (2004). Random set independence is weaker than epistemic irrelevance. Personal communication. 101, 102

de Cooman, G. and D. Aeyels (1999). Supremum preserving upper probabilities. *Information Sciences 118*, 173–212. LII

de Finetti, B. (1937). La prevision: ses lois logiques, ses sources subjectives. In *Ann. Inst. Henri Poincaré*, Volume 7, pp. 1–68. Institut Henri Poincaré, Paris. 50, XL

de Finetti, B. (1974). *Theory of Probability.* Wiley, London. 2, XL, XLI

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics 38*, 325–339. 55, L

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B 30*, 205–247. 104, 105, 109

Denneberg, D. (1994). *Non-additive Measure and Integral.* Kluwer, Boston. XLVII

Dessai, S. and M. Hulme (2003). Does climate policy need probabilities? Tyndall Centre Working Paper 34, Tyndall Centre. 3, 55

Dickinson, R. E. (1981). Convergence rate and stability of ocean-atmosphere coupling schemes with a zero-dimensional climate model. *Journal of Atmospheric Science 38*, 2112–2120. XI

Dickinson, R. E. and K. J. Schaudt (1998). Analysis of timescales of response of a simple climate model. *Journal of Climate 11*, 97–107. VIII, XXX, XXXII

Dong, W. M. and H. C. Shah (1987). Vertex method for computing functions of fuzzy variables. *Fuzzy Sets and Systems 24*, 65–78. 134

Dowlatabadi, H. and M. G. Morgan (1995). Integrated assessment of climate change. *Science 259*, 1813 & 1932. 1

Drud, A. S. (1992). CONOPT: A large-scale GRG code. *ORSA Journal on Computing 6*, 207–216. 39

Dubois, D. and H. Prade (1988). *Possibility Theory: An Approach to Computerized Processessing of Uncertainty.* Plenum Press, New York. LII

Dubois, D. and H. Prade (1990). Consonant approximations of belief functions. *International Journal of Approximate Reasoning 4*, 419–449. LII

Dubois, D. and H. Prade (1991). Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems 42*, 87–101. 128

Dubois, D. and H. Prade (1992). When upper probabilities are possibility measures. *Fuzzy Sets and Systems 49*, 203–244. LII

Dubois, D. and H. Prade (1998). Possibility theory: Qualitative and quantitative aspects. In D. M. Gabbay and P. Smets (Eds.), *Quantified Representation of Uncertainty and Imprecision*, Volume 1 of *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pp. 169–226. Kluwer Academic Publishers, Dordrecht. LII

Dutton, J. A. (1995). An analytical model of atmospheric feedback and global temperature change. *Journal of Climate 8*, 1122–1139. 10, X, XIV

Edenhofer, O., N. Bauer, and E. Kriegler (2005). The impact of technological change on climate protection and welfare - Insights from the model MIND. *Ecological Economics 54*, 277–292. vii, 10, 140

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics 75*, 643–669. 6, 7, 144

EPICA community members (2004). Eight glacial cycles from an Antarctic ice core. *Nature 429*, 623–628. 21

European Commission (2000). Communication of the Commission on the precautionary principle. COM(2000) 1. 6

Ferson, S., L. R. G. V. Kreinovich, D. S. Meyers, and K. Sentz (2002). Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002-4015, Sandia National Laboratories, Albuquerque, NM. 62, 64, 65, 69, 147

Fetz, T. and M. Oberguggenberger (2004). Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. *Reliability Engineering and System Safety 85*, 73–87. 101, 102, 129, 150

Fligge, M. and S. K. Solanki (2000). The solar spectral irradiance since 1700. *Geophysical Research Letters 27(14)*, 2157–2160. 26

Folland, C. K. and T. R. Karl (2001). Observed climate variability and change. In J. Houghton and Y. Ding (Eds.), *Climate Change 2001: The Scientific Basis*, Contribution of Working Group I to the Third Assessment Report of the IPCC, pp. 99–181. Cambridge University Press, Cambridge. 131

Folland, C. K., N. A. N.A.Rayner, S. J. S.J.Brown, T. M. T.M.Smith, S. S. P. Shen, D. E. Parker, I. Macadam, P. D. Jones, R. N. Jones, N. Nicholls, and D. M. H. Sexton (2001). Global temperature change and its uncertainties since 1861. *Geophysical Research Letters 28*, 2621–2624. 29

Forest, C. E., P. H. Stone, A. P. Sokolow, M. R. Allen, and M. D. Webster (2002). Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science 295*, 113–117. 4, 11, 46, 51, 56, 57, 58, 59, 65, 68, 70, 74, 81, 87, 88, 90, 91, 101, 102, 121, 124, 151, 152, 153

Forster, P. M., F. De, M. Blackburn, R. Glover, and K. P. Shine (2000). An examination of climate sensitivity for idealised climate change experiments in an intermediate general circulation model. *Climate Dynamics 16*, 833–849. IV

Foukal, P., G. North, and T. Wigley (2004). A stellar view on solar variations and climate. *Science 306*, 68–69. 26

Frame, D. J., B. B. B. Booth, J. A. Kettleborough, D. A. Stainforth, J. M. Gregory, M. Collins, and M. R. Allen (2005). Constraining climate forecasts. the role of prior assumptions. *Geophysical Research Letters 32*, doi:10.1029/2004GL022241. 5

Fröhlich, C. (2000). Solar irradiance. *Space Science Reviews 94*, 15–24. 26, I

Fröhlich, C. and J. Lean (1998). The sun's total irradiance: cycles, trends, and related climate change uncertainties since 1976. *Geophysical Research Letters 25(23)*, 4377–4380. 26, 131

Fu, Q., C. M. Johanson, S. G. Warren, and D. J. Seidel (2004). Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends. *Nature 429*, 55–58. VII

Genest, C. and J. V. Zidek (1986). Combining probability distributions: A critique and annotated bibliography. *Statistical Science 1*, 114–148. 3

Gilboa, I. (1987). Expected utility theory with purely subjective non-additive probabilities. *Journal of Mathematical Economics 16*, 65–88. 7, 144, XLVI

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*, 141–153. 7, 144

Gilboa, I. and D. Schmeidler (1993). Updating ambiguous beliefs. *Journal of Economic Theory 59*, 33–49. 105, 109, 117, 149

Gilboa, I. and D. Schmeidler (1994). Additive representations on nonadditive measures and the Choquet Integral. *Annals of Operational Research 52*, 43–65. XLIX

Gilboa, I. and D. Schmeidler (1995). Canonical representations of set functions. *Mathematics of Operations Research 20*, 197–212. XLIX

Giron, F. J. and S. Rios (1980). Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, J. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics*, pp. 17–38. University Press, Valencia. XXXIX

Good, I. J. (1962). Subjective probability as the measure of a non-measurable set. In *Logic, Methodology and Philosophy of Science*, pp. 319–329. Stanford University Press, California. 50, XXXIX

Gordon, C., C. Cooper, and C. A. Senior (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics 16*, 147–168. 19

Gregory, J. M., W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, J. A. Lowe, T. C. Johns, and K. D. Williams (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters 31*, doi:10.1029/2003GL018747. 20, VI

Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stot, and N. A. Rayner (2002). An observationally based estimate of climate sensitivity. *Journal of Climate 15*, 3117–3121. 60, 81

Grübler, A. and N. Nakicenovic (2001). Identifying dangers in an uncertain climate. *Nature 412*, 15. 3

Ha Duong, M. (2003). Imprecise probability bridges scenario-forecast gap. Working Paper, CIRED. 147

Hall, J., C. Twyman, and A. Kay (2005). Influence diagrams for representing uncertainty in climate-related propositions. *Climatic Change 69*, 343–365. 147

Hall, J. W. and J. Lawry (2004). Generation, combination and extension of random set approximations to coherent lower and upper probabilities. *Reliability Engineering and System Safety 85*, 89–101. 71

Hallin, M. and A. Saidi (2003). Optimal tests for non-correlation between multivariate time series. Technical Report 0428, IAP Statistics Network. 37, 38, 41

Hallin, M. and A. Saidi (2005). Testing non-correlation and non-causality between multivariate ARMA time series. *Journal of Time Series Analysis 26*, 83–105. 37

Hansen, J., M. Sato, L. Nazarenko, R. Ruedy, A. Lacis, D. Koch, I. Tegen, T. Hall, D. Shindell, B. Santer, P. Stone, T. Novakov, L. Thomason, R. Wang, Y. Wang, D. Jacob, S. Hollandsworth, L. Bishop, J. Logan, A. Thompson, R. Stolarski, J. Lean, R. Willson, S. Levitus, J. Antonov, R. Rayner, D. Parker, and J. Christy (2002). Climate forcings in Goddard Institute for Space Studies SI2000 simulations. *Journal of Geophysical Research 107*, 4347. 25, 27, 28, 91, VII

Hansen, J., M. Sato, and R. Ruedy (1997). Radiative forcing and climate response. *Journal of Geophysical Research 102*, 6831–6864. 28, IV

Harvey, D., J. Gregory, M. Hoffert, A. Jain, M. Lal, R. Leemans, S. Raper, T. Wigley, and J. de Wolde (1997). An introduction to simple climate models used in the IPCC Second Assessment Report. In J. T. Houghton, L. G. Meira Filho, D. J. Griggs, and K. Maskell (Eds.), *IPCC Technical Paper II.* 23

Harvey, D. D. and T. M. L. Wigley (2003). Characterizing and comparing control-run variability of eight coupled AOGCMs and of observations. Part 1: Temperature. *Climate Dynamics 21*, 619–646. 29, 36

Harvey, L. D. D. (2000). *Global Warming. The Hard Science.* Prentice Hall, Harlow. 14, 22, 23, 25, 26, 101, II, IV, VI, VII

Harvey, L. D. D. (2004). Declining temporal effectiveness of carbon sequestration: Implications for compliance with the United National Framework Convention on Climate Change. *Climatic Change 63*, 259–290. 10

Harvey, L. D. D. and Z. Huang (2001). A quasi-one-dimensional coupled climate-change cycle model, 1. Description and behavior of the climate component. *Journal of Geophysical Research 106*, 22,339–22,353. 10, IX

Harvey, L. D. D. and R. K. Kaufmann (2002). Simultaneously constraining climate sensitivity and aerosol radiative forcing. *Journal of Climate 15*, 2837–2861. 28, 46

Harvey, L. D. D. and S. H. Schneider (1985). Transient climate response to external forcing on $10^0$–$10^4$ year time scales. Part 2: Sensitivity experiments with a seasonal, hemispherically averaged, coupled atmosphere, land, and ocean energy balance model. *Journal of Geophysical Research 90*, 2207–2222. 10

Hasselmann, K. (1976). Stochastic climate models. Part I: Theory. *Tellus 28*, 473–484. 29, 30

Hasselmann, K. (1993). Optimal fingerprints for the detection of time-dependent climate change. *Journal of Climate 6*, 1957–1969. 4, 12, 31

Hasselmann, K. (1997). Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics 13*, 601–611. 12

Hasselmann, K. (1998). Conventional and Bayesian approach to climate-change detection and attribution. *Quarterly Journal of the Royal Meteorological Society 124*, 2541–2565. 11

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: a univariate residual cross-correlation approach. *Journal of the American Statistical Association 71*, 378–385. 37

Hegerl, G. C., H. v. Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones (1996). Detecting greenhouse gas induced climate change with an optimal fingerprint method. *Journal of Climate 9*, 2281–2306. 12

Henry, C. and M. Henry (2002). Formalization and applications of the precautionary principle. Discussion Paper No 2002009, Institut de Recherches Economiques et Sociales (IRES), Université catholique de Louvain. 6

Herron, T., T. Seidenfeld, and L. Wasserman (1997). Divisive conditioning: Further results on dilation. *Philosophy of Science 64*, 411–444. 69, 71, 106, 126

Hoffert, M. I., A. J. Callegari, and C. T. Hsieh (1980). The role of deep sea heat storage in the secular response to climate forcing. *Journal of Geophysical Research 85*, 6667–6679. 10, VIII, IX

Honerkamp, J. (1994). *Stochastic Dynamical Systems: Concepts, Numerical Methods, Data Analysis.* VCH, New York. 30

Honerkamp, J. and H. Römer (1989). *Klassische Theoretische Physik.* Springer-Verlag, Berlin. in German. XXXII, XXXV

Hoyt, D. V. and K. H. Schatten (1993). A discussion of plausible solar irradiance variations, 1700-1992. *Journal of Geophysical Research 98(A11)*, 18,895–18,906. 26

Huber, P. J. (1973). The use of Choquet capacities in statistics. *Bull. Internat. Statist. Inst. 45*, 181–188. XLVI

Huber, P. J. (1981). *Robust statistics.* Wiley, New York. XLVIII

Huber, P. J. and V. Strassen (1973). Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist. 1*, 251–263. XLVI

Hurrell, J. (1995). Decadal trends in the North Atlantic Oscillation and relationships to regional temperature and precipitation. *Science 269*, 676–679. 32

Hurrell, J. W., Y. Kushnir, G. Ottersen, and M. Visbeck (2003). *The North Atlantic Oscillation: Climatic Significance and Environmental Impact.* American Geophysical Union. 32

Hurwicz, L. (1951). Optimality criteria for decision making under ignorance. Cowles Commision Discussion Paper. 144

IPCC (2001). *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the IPCC.* Cambridge University Press, Cambridge. 130

IPCC (2004). Workshop on climate sensitivity, Paris, 26-29.7.2004. Technical report, Intergovernmental Panel on Climate Change, Working Group I Technical Support Unit, Boulder, Colorado. 60

Jacobson, M. Z. (2001). Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols. *Nature 409*, 695–697. 25, 91

Jaffray, J. Y. (1989). Linear utility theory for belief functions. *Operations Research Letters 8*, 107–112. 144

Jaffray, J. Y. (1992). Bayesian updating and belief functions. *IEEE Transactions on Systems, Man, and Cybernetics 22*, 1144–1152. 109, 110

Jaffray, J. Y. (1994). Dynamic decision making with belief functions. In R. R. Yager, J. Kacprzyk, and M. Fedrizzi (Eds.), *Advances in the Dempster-Shafer theory of evidence*, pp. 331–352. Wiley, New York. 144

Jones, P. and A. Moberg (2003). Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *Journal of Climate 16*, 206–223. 29

Jones, P., T. J. Osborn, K. R. Briffa, C. Folland, E. B. E.B. Horton, L. V. Alexander, D. E. Parker, and N. A. Rayner (2001). Adjusting for sampling density in grid-box land and ocean surface temperature time series. *Journal of Geophysical Research 106*, 3371–3380. 29, 95

Jones, P. D. (1989). The influence of ENSO on global temperatures. *Climate Monitoring 17*, 80–89. 32

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor (1999). Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics 37*, 173–199. XIII

Joshi, M. M., K. P. Shine, M. Ponater, N. Stuber, R. Sausen, and L. Li (2003). A comparison of climate response to different radiative forcings in three general circulation models: towards an improved metric of climate change. *Climate Dynamics 20*, 843–854. IV

Kass, R. E. and L. Wasserman (1996). Formal rules for selecting prior distributions: A review and annotated bibliography. *Journal of the American Statistical Association 91*, 1343–1370. 4

Kaufmann, R. K. and D. I. Stern (1997). Evidence for human influence on climate from hemispheric temperature relations. *Nature 388*, 39–44. 12, 31

Keeling, C. and T. Whorf (2004). Atmospheric $CO_2$ records from sites in the SIO air sampling network. In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge. 21

Keith, D. W. (1996). When is it appropriate to combine expert judgements? *Climatic Change 33*, 139–143. 3

Kleinen, T. (2005). *Stochastic information and the assessment of climate change*. Ph. D. thesis, University of Potsdam. 146

Knutson and S. T. R. Manabe (1998). Model assessment of decadal variability and trends in the tropical pacific. *Journal of Climate 11*, 2273–2296. 34

Knutti, R., T. F. Stocker, F. Joos, and G. K. Plattner (2002). Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature 416*, 719–723. 11, 46, 56, 57, 58, 60, 65, 68, 70, 74, 81, 87, 88, 90, 91, 92, 102, 121, 124, 151, 152

Knutti, R., T. F. Stocker, F. Joos, and G.-K. Plattner (2003). Probabilistic climate change projections using neural networks. *Climate Dynamics 21*, 257–272. 91

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Foundation of the Theory of Probability, Second English ed.). Chelsea, New York, 1956. 50, XXXIX, XLVI

Koscielny-Bunde, E., A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H.-J. Schellnhuber (1998). Indication of a universal persistence law governing atmospheric variability. *Physical Review Letters 81*, 729–732. 30

Kozine, I. O. and Y. Filimonov (2000). Imprecise reliabilities: experiences and advances. *Reliability Engineering & Systems Safety 67*, 75–83. 52

Kozine, I. O. and V. Krymsky (2003). Reducing uncertainty by imprecise judgements on probability distributions: Application to system reliability. In *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, pp. 335–344. 75

Kriegler, E. and T. Bruckner (2002). Global and Annex I emissions corridors for the 21st century. In A. E. Riazolli and A. J. Jakeman (Eds.), *Integrated Assessment and Decision Support*, Volume 3 of *Proceedings of the First Biennal Meeting of the International Environmental Modelling and Software Society (iEMSs2002)*, Lugano, June 24-27, pp. 151–156. vii, 146

Kriegler, E. and T. Bruckner (2004). Sensitivity analysis of emissions corridors for the 21st century. *Climatic Change 66*, 345–387. vii, 10, 146

Kriegler, E. and H. Held (2003). Transformation of probability families using possibility theory. Working Paper, Potsdam Institute for Climate Impact Research. LII

Kriegler, E. and H. Held (2005). Utilizing belief functions for the estimation of future climate change. *International Journal of Approximate Reasoning 39*, 185–209. 58, 59, 62, 66, 75, 147, 148

Kriegler, E., H. Held, and T. Bruckner (2005). Climate protection strategies under ambiguity about catastrophic consequences. In J. Kropp and J. Scheffran (Eds.), *Decision Making and Risk Management in Sustainability Science (to appear).* Nova Science, New York. 128

Kuznetsov, V. P. (1991). *Interval Statistical Methods* (in Russian). Radio i Svyaz Publ., Moscow. 50

Lange, A. (2000). Decisions on greenhouse gas emissions under uncertainty: The concept of Choquet expected utility maximization. Discussion Paper Nr. 315, Interdisciplinary Institute for Environmental Economics, University of Heidelberg. 147, XLIX

Laplace, P. S. (1812). *Théorie Analytique des Probabilités.* Courcier, Paris. XLVI

Latif, M. and T. P. Barnett (1996). Decadal climate variability over the North Pacific and North America: dynamics and predictability. *Journal of Climate 9*, 2407–2423. 34

Lean, J. (2000). Evolution of the Sun's spectral irradiance since the Maunder Minimum. *Geophysical Research Letters 27(16)*, 2425–2428. 26, 27, 131

Lean, J., J. Beer, and R. Bradley (1995). Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophysical Research Letters 22(23)*, 3195–3198. 26

Lean, J., Y.-M. Wang, and N. R. Sheeley Jr. (2002). The effect of increasing solar activity on the Sun's total and open magnetic flux during multiple cycles: Implications for solar forcing of climate. *Geophysical Research Letters 29(24)*, 2224. 26

Lefohn, A. S., J. D. Husar, and R. B. Husar (1999). Estimating historical anthropogenic global sulfur emissions patterns for the period 1850-1990. *Atmospheric Environment 33*, 3435–3444. 23

Leroy, S. S. (1998). Detecting climate signals: Some Bayesian aspects. *Journal of Climate 11*, 640–651. 11

Levi, I. (1974). On indeterminate probabilities. *Journal of Philosophy 71*, 391–418. 143

Levi, I. (1980). *The Enterprise of Knowledge.* MIT Press, Cambridge, Masachusetts. 7, 50, 143, XXXIX, XL

Levitus, S., J. I. Antonov, T. P. Boyer, and C. Stephens (2000). Warming of the world ocean. *Science 287*, 2225–2229. 82, 94, 95, 96

Levitus, S., J. I. Antonov, J. Wang, T. L. Delworth, K. W. Dixon, and A. J. Broccoli (2001). Anthropogenic warming of earth's climate system. *Science 292*, 267–270. 94, 96

Luce, R. D. and H. Raiffa (1957). *Games and Decisions: Introduction and Critical Survey.* John Wiley and Sons, New York. 144

Luo, W. and W. F. Caselton (1997). Using Dempster-Shafer theory to repesent climate change uncertainties. *Journal of Environmental Management 49*, 73–93. 147, LI

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis (1997). A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society 78*, 1069–1079. 34

Mantua, N. J. and R. H. Steven (2002). The Pacific Decadal Oscillation. *Journal of Oceanography 58*, 35–44. 34

Maraun, D., H. W. Rust, and J. Timmer (2004). Tempting long-memory - on the interpretation of DFA results. *Nonlinear Processes in Geophysics 11*, 495–503. 30

Marshall, S., R. J. Oglesby, J. W. Larson, and B. Saltzmann (1994). A comparison of GCM sensitivity to changes in $CO_2$ and solar luminosity. *Geophysical Research Letters 21*, 2487–2490. IV

Mears, C. A., M. C. Schabel, and F. J. Wentz (2003). A reanalysis of the MSU channel 2 tropospheric temperature record. *Journal of Climate 16*, 3650–3664. VII

Milankovitch, M. (1941). Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem. *R. Serbian Acad. Spec. Publ. 132 33*, 633. I

Morgan, M. G. and D. W. Keith (1995). Subjective judgments by climate experts. *Environmental Policy Analysis 29*, 468–476. 3, 55, 56

Mukerji, S. and J. M. Tallon (2004). An overview of economic applications of David Schmeidler's models of decision making under uncertainty. In *Uncertainty in Economic Theory. A collection of essays in honor of David Schmeidler's 65th birthday.* Routledge. XLIX

Munk, W. (1966). Abyssal recipes. *Deep Sea Research 13*, 707–730. 94, 95, IX, XXIX

Murphy, J. M. (1995). Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part III: Analysis of global mean response using simple models. *Journal of Climate 8*, 496–514. 10, 18, V, VI, XXVII

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 uncerainties in a large ensemble of climate change simulations. *Nature 430*, 768–772. 51, 56, 59, 60, 65, 68, 70, 74, 81, 87, 88, 121, 124

Myhre, G., A. Myhre, and F. Stordal (2001). Historical evolution of radiative forcing of climate. *Atmospheric Environment 35*, 2361–2373. 25

Nakićenović, N. and R. Swart (2000). *Emissions Scenarios. Special Report of the IPCC.* Cambridge University Press, Cambridge. 127, 130, 131

Nordhaus, W. D. (1994). *Managing the Global Commons: The Economics of Climate Change.* MIT Press, Cambridge (MA). 10

Paté-Cornell, E. (1997). Different levels of treatment of uncertainty in risk analysis and aggregation of expert opinions. In S. J. Hassol and J. Katzenberger (Eds.), *Elements of Change 1996.* Aspen Global Change Institute. 3

Peixoto, J. P. and A. H. Oort (1992). *Physics of Climate.* Springer-Verlag, New York. 31, I, II, XI, XIV

Penner, J. E. (2001). Aerosols, their direct and indirect effects. In J. Houghton and Y. Ding (Eds.), *Climate Change 2001: The Scientific Basis*, Contribution of Working Group I to the Third Assessment Report of the IPCC, pp. 289–348. Cambridge University Press, Cambridge. 23

Petoukhov, V., A. Ganopolski, V. Brovkin, M. Claussen, A. Eliseev, C. Kubatzki, and S. Rahmstorf (2000). CLIMBER-2: A climate system model of intermediate complexity. Part I: Model description and performance for present climate. *Climate Dynamics 16*, 1–17. XIX

Petschel-Held, G., H. J. Schellnhuber, T. Bruckner, F. L. Tóth, and K. Hasselmann (1999). The tolerable windows approach: Theoretical and methodological foundations. *Climatic Change 41*, 303–331. 146

Ramachandran, S., V. Ramaswamy, G. L. Stenchikov, and A. Robock (2000). Radiative impact of the Mount Pinatubo volcanic eruption: Lower stratospheric response. *Journal of Geophysical Research 105*, 24,409–24,429. 28

Ramaswamy, V. (2001). Radiative forcing of climate change. In J. Houghton and Y. Ding (Eds.), *Climate Change 2001: The Scientific Basis*, pp. 289–348. Cambridge University Press, Cambridge. 11, 21, 22, 23, 25, 26, 54, 81, 154

Ramsey, F. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and other Logical Essays*, pp. 156–198. Routledge and Kegan Paul, London. 50

Raper, S. C. B. and U. Cubasch (1996). Emulation of the results from a coupled general circulation model using a simple climate model. *Geophysical Research Letters 23*, 1107–1110. 18, VIII, XIX, XX

Raper, S. C. B., J. M. Gregory, and T. J. Osborn (2001). Use of an upwelling-diffusion balance climate model to simulate and diagnose A/OGCM results. *Climate Dynamics 17*, 601–613. 20, IX, XVIII, XX, XXX

Raper, S. C. B., J. M. Gregory, and R. J. Stouffer (2002). The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *Journal of Climate 15*, 124–130. 101

Raper, S. C. B., T. M. L. Wigley, and W. R. A. (1996). Global sea-level rise: Past and future. In D. Milliman and B. U. Haq (Eds.), *Sea-Level Rise and Coastal Subsidence*, pp. 11–46. Kluwer Academic Publisher. 21

Regan, H. M., S. Ferson, and D. Berleant (2004). Equivalence of methods for uncertainty propagation of real-valued random variables. *International Journal of Approximate Reasoning 36*, 1–30. 64

Robock, A. (2000). Volcanic eruptions and climate. *Reviews of Geophysics 38*, 191–219. 27

Ropelewski, C. and P. Jones (1987). An extension of the Tahiti-Darwin Southern Oscillation Index. *Monthly Weather Review 115*, 2161–2165. 31, 41

Rotmans, J. (1990). *IMAGE: An Integrated Model to Assess the Greenhouse Effect.* Dordrecht: Kluwer Academics. 1

Rotstayn, L. D. and J. E. Penner (2001). Indirect aerosol forcing, quasi forcing, and climate response. *Journal of Climate 14*, 2960–2975. IV

Rowntree, P. R. (1998). Global average climate forcing and temperature response since 1750. *International Journal of Climatology 18*, 355–377. VIII

Savage, L. J. (1954). *The Foundations of Statistics.* John Wiley & Sons, New York. 6, 144, 145

Schellnhuber, H.-J. (1998). Earth Systems Analysis - The Scope of the Challenge. In H.-J. Schellnhuber and V. Wenzel (Eds.), *Earth System Analysis - Integrating Science for Sustainability*, pp. 3–195. Springer, Heidelberg. 145

Schellnhuber, H.-J. (1999). 'Earth Systems Analysis' and the second Copernican revolution. *Nature 402*, C19–C23. 145

Schellnhuber, H.-J. and G. Yohe (1997). Comprehending the economic and social dimensions of climate change by integrated assessment. In *Proceedings of the Conference on the World Climate Research Programme: Achievements, Benefits, and Challenges, 26-28 August 1997, Geneva.* 1

Schervish, M. J., T. Seidenfeld, J. B. Kadane, and I. Levi (2003). Extensions of expected utility theory and some limitations of pairwise comparisons. In *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications*, pp. 496–510. 143, 144

Schlesinger, M. E., N. G. Andronova, B. Entwistle, A. Ghanem, N. Ramankutty, W. Wang, and F. Yang (1997). Modeling and simulation of climate and climate change. In *Proceedings of the International School of Physics Enrico Fermi CXXXIII*, Amsterdam, pp. 389–429. IOS Press. 58, IX, X, XI, XVIII, XXVII

Schlesinger, M. E. and X. Jiang (1990). Simple model representation of atmosphere-ocean GCMs and estimation of the time scale of $CO_2$ induced climate change. *Journal of Climate 3*, 1297–1315. 10

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica 57*, 571–587. 7, 144, XLVI

Schneider, S. H. (2001). What is dangerous climate change? *Nature 411*, 17–19. 3, 55

Schneider von Deimling, T., H. Held, A. Ganopolski, and S. Rahmstorf (2005). Climate sensitivity range derived from large ensemble simulations of glacial climate constrained by proxy data. *Climate Dynamics*, submitted. 18, 81, 153, XI, XIX, XX

Seidenfeld, T., M. J. Schervish, and J. B. Kadane (1995). A representation of partially ordered preferences. *Annals of Statistics 23*, 2168–2217. 141, XXXIX

Seidenfeld, T. and L. Wasserman (1993). Dilation for sets of probabilities. *The Annals of Statistics 21*, 1139–1154. 106, 122

Sellers, W. D. (1969). A global climate model based on the energy balance of the earth-atmosphere system. *Journal of Applied Meteorology 8*, 396–400. 10

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

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton: Princeton U. Press. 55, XLIX, L, LXVI

Shapley, L. S. (1971). Cores of convex games. *International Journal of Game Theory 1*, 11–26. XLVI

Shine, K. P., J. Cook, E. J. Highwood, and M. M. Joshi (2003). An alternative to radiative forcing for estimating the relative importance of climate change mechanisms. *Geophysical Research Letters 30*, 2047. V

Simmons, H. L., S. R. Jayne, L. C. S. Laurent, and A. J. Weaver (2004). Tidally driven mixing in a numerical model of the ocean circulation. *Ocean Modelling 6*, 245–263. 94, XXIX

Smets, P. and R. Kennes (1994). The transferable belief model. *Artificial Intelligence 66*, 191–234. LI

Smith, J., H.-J. Schellnhuber, and M. Q. Mirza (2001). Lines of evidence for vulnerability to climate change: A synthesis. In O. Canziani and J. McCarthy (Eds.), *Climate Change 2001: Impacts, Adaptation and Vulnerability*, Contribution of Working Group II to the Third Assessment Report of the IPCC, pp. 914–967. Cambridge University Press, Cambridge. 1

Smith, R. L., T. M. L. Wigley, and B. D. Santer (2003). A bivariate time series approach to anthropogenic trend detection in hemispheric mean temperatures. *Journal of Climate 16*, 1228–1240. 12, 31, 32

Smith, S. J., H. Pitcher, and T. M. L. Wigley (2001). Global and regional anthropogenic sulfur dioxide emissions. *Global and Planetary Change 29*, 99–119. 23

Sokolov, A. P., C. E. Forest, and P. H. Stone (2003). Comparing oceanic heat uptake in aogcm transient climate change experiments. *Journal of Climate 16*, 1573–1582. VI

Sokolov, A. P. and P. H. Stone (1998). A flexible climate model for use in integrated assessments. *Climate Dynamics 14*, 291–303. 56, VI

Stainforth, D. A., T. Aina, C. Christensen, M. Collins, N. Fauli, D. J. Frame, K. J. A., 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 predicitions of the climate response to rising levels of greenhouse gases. *Nature 433*, 403–406. 51, 60, 81

Stott, P. A. (2001). Attribution of twentieth century temperature change to natural and anthropogenic forcing. *Climate Dynamics 17*, 1–21. 12

Stott, P. A. and J. A. Kettleborough (2002). Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature 416*, 723–726. 3

Strat, T. M. (1990). Decision analysis using belief functions. *Int. Journal of Approx. Reasoning 4*, 391–418. 144

Stuber, N., R. Sausen, and M. Ponater (2001). Stratosphere adjusted radiative forcing calculations in a comprehensive climate model. *Theoretical and Applied Climatology 68*, 125–135. IV

Sugeno, M. (1974). *Theory of Fuzzy Integrals and Its Applications.* Ph. D. thesis, Tokyo Institute of Technology, Tokyo, Japan. XLVIII

Tett, S. F. B. (2002). Estimation of natural and anthropogenic contributions to 20th century temperature change. *Journal of Geophysical Research 107*, doi:10.1029/2000JD000028. 12

Tett, S. F. B., P. A. Stott, M. R. Allen, W. I. Ingram, and J. F. B. Michell (1999). Causes of twentieth century temperature change near the Earth's surface. *Nature 399*, 569–572. 12

Thompson, D. W. J., S. Lee, and M. Baldwin (2003). Atmospheric processes governing the northern hemisphere annular mode/north atlantic oscillation. In J. W. Hurrell, Y. Kushnir, G. Ottersen, and M. Visbeck (Eds.), *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, Geophysical Monograph Series, Vol. 134, pp. 81–112. American Geophysical Union. 32

Tol, R. S. J. and A. F. De Vos (1998). A Bayesian statistical analysis of the enhanced greenhouse effect. *Climatic Change 38*, 87–112. 60

Tonon, F. (2004). Using random set theory to propagate uncertainty through a mechanical system. *Reliability Engineering and System Safety 85*, 169–181. 65

Tonon, F., A. Bernardini, and A. Mammino (2000). Reliability analysis of rock mass response by means of random set theory. *Reliability Engineering and System Safety 70*, 263–282. 134

Tóth, F. L., T. Bruckner, H. M. Füssel, M. Leimbach, G. Petschel-Held, and H. J. Schellnhuber (1997). The tolerable windows approach to integrated assessments. In *Climate Change and Integrated Assessment Models - Bridging the gaps*, Proceedings of the IPCC Asia-Pacific Workshop on Integrated Assessments Models, Ibaraki, pp. 403–440. Center for Global Environmental Research. 146

United Nations (1992). United Nations Framework Convention on Climate Change. *International Legal Materials 31*, pp. 849. 136

Utkin, L. V. and S. Gurov (1999). Imprecise reliability of general structures. *Knowledge and Information Systems 1*, 459–480. 52

Utkin, L. V. and I. O. Kozine (2001). Different faces of the natural extension. In *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pp. 316–323. XLIII

Utkin, L. V. and I. O. Kozine (2002). Stress-strength reliability models under incomplete information. *International Journal of General Systems 31*, 549–568. 52

von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior.* Princeton Univ. Press, Princeton, New Jersey. 3, 145

von Storch, H. and F. W. Zwiers (1999). *Statistical analysis in climate research.* Cambridge University Press, Cambridge. 11, 30, 35, 36, 41, 43

Walley, P. (1981). Coherent lower (and upper) probabilities. Statistics Research Report 22, University of Warwick, Coventry. XLVII, XLVIII

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London. 5, 7, 49, 50, 54, 63, 69, 73, 75, 76, 77, 100, 104, 105, 117, 141, 142, 143, 147, XXXIX, XLI, XLII, XLIII, XLIV, LV, LVII, LXVI

Walley, P. (1996a). Inference from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society 58*, 3–57. 7

Walley, P. (1996b). Measures of uncertainty in expert systems. *Artificial Intelligence 83*, 1–58. 117, XLII, XLVI

Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning 24*, 125–148. XXXIX, XLVI

Wang, Z. and G. J. Klir (1992). *Fuzzy Mearure Theory.* Plenum Press, New York. XLVII

Wasserman, L. A. and J. B. Kadane (1990). Bayes' theorem for Choquet capacities. *The Annals of Statistics 18*, 1328–1339. 71, 107, 108, 110, 115, LXIX

Watterson, I. G. (2000). Interpretation of simulated global warming using a simple model. *Journal of Climate 13*, 202–215. VI

WBGU (1995). *Scenario for the Derivation of Global $CO_2$ Reduction Targets and Implementation Strategies.* Special Report for the First Conference of the Parties to the FCCC. Bremerhaven: German Advisory Council on Global Change (WBGU). 139

Webster, M. D. and A. P. Sokolov (2000). A methodology for quantifying uncertainty in climate projections. *Climatic Change 46*, 417–446. 56, 60

Weichselberger, K. (2000). The theory of interval probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning 24*, 149–170. 50, XLV, XLVII

Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept* (in German). Physica Verlag, Heidelberg. XLV, XLVII

Wigley, T. M. L. and S. C. B. Raper (1992). Implications for climate and sea level of revised IPCC emissions scenarios. *Nature 357*, 293–300. 10, 21, IX, XI

Wigley, T. M. L. and S. C. B. Raper (2001). Interpretation of high projections for global-mean warming. *Science 293*, 451–454. 60

Wigley, T. M. L. and S. C. B. Raper (2002). Reasons for larger warming projections in the IPCC Third Assessment Report. *Journal of Climate 15*, 2945–2952. 21

Wigley, T. M. L., R. Richels, and J. A. Edmonds (1996). Alternative emissions pathways for stabilizing concentrations. *Nature 379*, 240–243. 10, 136

Wigley, T. M. L. and M. E. Schlesinger (1985). Analytical solution for the effect of increasing $CO_2$ on global mean temperature. *Nature 315*, 649–652. 17, XXXII

Williamson, R. C. and T. Downs (1990). Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning 4*, 89–158. 62, 65, 96

Yager, R. R. (1986). Arithmetic and other operations on Dempster-Shafer structures. *International Journal of Man-Machine Studies 25*, 357–366. 63, 128

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems 1*, 3–29. LII

Zhang, Y., J. Wallace, and D. Battisti (1997). ENSO-like interdecadal variability: 1900-93. *Journal of Climate 10*, 1004–1020. 34