Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

SCHOOL OF INTEGRATED
CLIMATE AND EARTH
SYSTEM SCIENCES

MASTER THESIS

A Philosophically Informed Evaluation of Integrated Assessment Models

Felix Schaumann

Faculty: Mathematics, Informatics and Natural Sciences

Program: M.Sc. Integrated Climate System Sciences

Matriculation Number: 7254834

First Reviewer: Prof. Dr. Moritz Drupp, Universität Hamburg

Second Reviewer: Dr. Christopher Hedemann, Expertenrat für Klimafragen

Further Supervision: Prof. Dr. Conrad Heilmann, Erasmus University Rotterdam
Prof. Dr. Johanna Baehr, Universität Hamburg

Official Topic

A Philosophy of Science Perspective on Integrated Assessment Modelling

Acknowledgements

First of all, I would like to thank my four supervisors Moritz Drupp, Christopher Hedemann, Conrad Heilmann and Johanna Baehr, as I am really grateful for all the time and energy that you have devoted to this project. In the beginning, I had absolutely no idea of where this was going. While I can't really say that this changed much until the very end, I know with certainty that I could not have gotten this far without your advice, support and continuous encouragement. I truly appreciate that. I would also like to thank all my friends from SICSS for the amazing time that we spent studying and living together. Special thanks to Edu for proofreading this thesis, to Deniz for continuous moral support, to Kilian and Alec for the feeling of a home, and to Chantal for always being there for me and spending the last two years with me. Finally, I want to thank my family – thank you for all the love and unconditional support that you have given me throughout the last quarter of a century. It means a lot to me.

Abstract

Integrated Assessment Models (IAMs) play an important role in climate policy decision making by combining knowledge from various domains into a single modelling framework. However, IAMs have been criticised for simplifying assumptions, reliance on negative emission technologies, as well as for their power of shaping discourses around climate policy. Given these controversies and the importance of IAMs for international climate policy, model evaluation is an important means of analysing how well IAMs perform and what can be expected of them. While different proposals for evaluating IAMs exist, they typically target a specific model type and are mostly reliant on a combination of abstract criteria and concrete evaluation methods. I enrich these perspectives by reviewing approaches from the philosophy of modelling and analysing their applicability to three canonical IAMs: DICE, REMIND, and IMAGE. The heterogeneity of IAMs and the political and ethical dimensions of their applications imply that using any single evaluation criterion can not capture the complexities of IAMs. In order to allow for the inclusion of these aspects into the evaluation procedure, I develop the idea of expectations, which captures the complex web of user aims, modelling purposes and evaluation criteria. Through this lens, I find that DICE is a useful tool for investigating the effects of different assumptions, but should not be expected to provide quantitative guidance. IMAGE, on the other hand, has proven to be suitable for projecting environmental impacts, but should not be expected to analyse questions that require a description of macroeconomic processes. REMIND can be used for an assessment of different theoretically possible mitigation pathways, but should not be expected to provide accurate forecasts. Further, I find that all three IAMs fail to deliver a comprehensive and informative model commentary, i.e. modellers do not sufficiently inform their audience about the appropriate domain of application, critical modelling choices and assumptions, or about how to interpret model results. Expectations for IAMs are often not clearly formulated, due to user aims which are hard to assess and vague purpose statements by modellers. As clearly formulated expectations form the basis of further evaluations of IAMs, I conclude that modellers should place more emphasis on informative model commentaries, with a special focus on the interpretation of IAM results.

Contents

1 Introduction

Integrated Assessment Models (IAMs) are important tools for informing climate policy decision making. These models combine knowledge from various domains into a single framework, with the explicit aim of informing policy and supporting decision making (Parson & Fisher-Vanden, 1997; Rotmans & Asselt, 1996). Generally, IAMs consist of natural science and economics modules, in order to capture both the physical and socioeconomic causes and effects of climate change (Farmer et al., 2015; Weyant, 2017). They are used to develop quantitative scenarios for the Intergovernmental Panel on Climate Change (IPCC; Bruce et al., 1996; IPCC, 2014), to calculate the social cost of carbon for policy appraisal (SCC; Helm, 2005; Nordhaus, 2017), or to investigate the effects of different policy options on mitigation scenarios (Bauer et al., 2020; Bertram et al., 2015; Hof et al., 2009; van Beek et al., 2020). Furthermore, IAMs are also used as tools for exploring the relative importance of a certain issue within a simple model (Dennig et al., 2015; Moore & Diaz, 2015). Through these modelling activities, IAMs have played influential roles in debates about climate targets (Aldy & Stavins, 2020; Dietz et al., 2018; Edenhofer et al., 2010; Nordhaus, 1993; van der Wijst et al., 2021) and in discussions on different mitigation options (Edenhofer et al., 2005; Grubler et al., 2018; Luderer et al., 2012).

From their advent in the 1990s until the present, IAMs have been subject to criticism, with increasingly fervent debates in recent years (Bosetti, 2021; Gambhir et al., 2019; Keen, 2020; Keppo et al., 2021; Pielke & Ritchie, 2021; Pindyck, 2017; Weyant, 2017). The models have been criticised for being based on 'arbitrary' assumptions (Pindyck, 2013a; Rosen & Guenther, 2015), relying on speculative negative emission technologies (Anderson & Peters, 2016; Beck & Oomen, 2021; Haikola et al., 2019; Low & Schäfer, 2020; Workman et al., 2021), and amplifying and legitimising specific climate policy narratives, e.g. of a market-based gradual readjustment of the world economy (Anderson & Jewell, 2019; Asefi-Najafabady et al., 2020; Beck & Krueger, 2016; Ellenbeck & Lilliestam, 2019; McLaren & Markusson, 2020). Yet, defendants of IAMs argue that much of the criticism stems from a misunderstanding about the nature of integrated assessment modelling. In their view, IAMs are tools to systematically explore different futures, without claiming that these results are especially probable or desirable (Anderson & Jewell, 2019; Gambhir, 2019; Peace & Weyant, 2008). Consequently, a model could be unrealistically simplified through strong assumptions and still be considered a useful tool for generating insights – if applied and interpreted properly (Weyant, 2009).

But how can we judge the quality of IAMs and what can reasonably be expected of them? Evaluation of IAMs has garnered much attention in the early days of IAMs, with a focus on the challenges of integrating different academic disciplines and on developing guidelines for 'good practice' in integrated assessment (Morgan & Dowlatabadi, 1996; Parson, 1996; Ravetz, 1997; Risbey et al., 1996). In recent years, the discourse on IAM evaluation has gained traction again (Hamilton et al., 2019; Schwanitz, 2013; Wilson et al., 2021). Now, the focus has been shifted towards conducting IAM evaluation on practice. Many proposals entail rather abstract criteria such as "appropriateness, interpretability, credibility, and relevance" (Wilson et al., 2021, p. 12) on the one hand, and very concrete evaluation methods on the other hand. These proposed methods include historical simulations which are compared against data (Millner & McDermott, 2016; Wilson et al., 2013), diagnostic indicators that capture complex model behaviour in standardised metrics (Harmsen et al., 2021; Kriegler, Petermann et al., 2015), or stylised historical patterns and trends (Schwanitz, 2013). However, the papers acknowledge the difficulties that arise from evaluating models of open systems (Schwanitz, 2013) in a context of structural epistemic and societal uncertainty (Wilson et al., 2017).

Evaluation of IAMs is further complicated by their heterogeneity. Most of the abovementioned proposals for IAM evaluation focus on a specific subset of model types. Earlier works on IAMs usually distinguished *policy evaluation* models from *policy optimisation* models, where the former simulate biophysical impacts of climate change and the latter optimise globally aggregated welfare in light of climate change (Bruce et al., 1996; Mastrandrea, 2010; Nordhaus, 2013; Tol, 2006). In more recent works, authors tend to distinguish between *benefit-cost* IAMs (BC-IAMs) and *detailed-process* IAMs (DP-IAMs). BC-IAMs are highly aggregated models that perform cost-benefit analyses, comparing benefits of avoided climate damages to costs of mitigation policies, in order to determine economically 'optimal' climate policies. DP-IAMs, as the name suggests, have a detailed representation of sectors and processes that are important for climate mitigation, primarily energy and the land use systems. They are mainly used to calculate cost-efficient mitigation pathways for reaching a given climate target, often in the context of the IPCC. Most model evaluation papers have focused on the IPCC-relevant DP-IAMs (Kowarsch, 2016b; Schwanitz, 2013; Wilson et al., 2021). An evaluation account that encompasses all types of IAMs and accounts for their relative strengths and weaknesses is thus missing.

In this thesis, I develop a perspective on model evaluation that can be applied to different categories of IAMs. This perspective is informed by approaches from the philosophy of modelling, through which I aim to shed a light on overlooked aspects of IAM evaluation. A useful starting point for evaluating a model is to analyse how well it represents the system being modelled. Yet, every model misrepresents its target system to some extent, as simplifications and idealisations are an essential feature of modelling itself (Knuuttila, 2009; Mäki, 2020). Consequently, model evaluation has to account for different kinds of idealisations and different kinds of functions that models fulfil – leading to a range of perspectives on what models are, how they are used and what makes them useful. Among those are the concept of epistemic tools (Knuuttila, 2011) and the notion of a model as a combination of structure and stories (Gibbard & Varian, 1978), where stories link the model to the real world (Morgan, 2001). Through the lens of epistemic tools, for example, IAMs can be analysed with respect to how they are constructed and justified, as well as how they are adapted and manipulated in application. Stories, on the other hand, can help illuminating how models are used to learn about the world, e.g. through a baseline scenario telling the story of a world without (further) climate policy, or through a scenario telling the story of a world with ambitious climate policy.

Both the philosophical literature on modelling and the IAM literature on model evaluation can contribute insights into what these models are and into criteria that could be used for evaluating them. However, the heterogeneity of IAMs and the variety of epistemic, political and ethical dimensions that play a role in their application (Beck & Krueger, 2016) make the use of a single perspective with a single evaluation criterion impossible – or, at the very least, unsatisfactory. Instead, I develop the idea of *expectations*, which captures the fact that IAMs are tied up in a complex web of different user needs, different modelling approaches and different criteria for model evaluation. An expectation, as I understand it, is defined by the combination of a purpose in the view of modeller, an aim that a user might have with respect to the model, and an associated evaluation criterion. One such expectation could, for example, consist of the criterion of realisticness, coupled to the modeller's purpose of describing interacting socioeconomic and climate systems, and be held by policymakers demanding robust evidence on which to base their decisions.

In this thesis, I use the perspective of expectations to evaluate three canonical IAMs: DICE (Nordhaus, 2018b), REMIND (Baumstark et al., 2021) and IMAGE (Stehfest et al., 2014). Through that, I analyse relative strengths and weaknesses of the three models and attempt to match them with different types of expectations, respectively. By selecting these three models, I span the range of different IAM types according to the two main classifications as introduced above: DICE stands for cost-benefit models (policy optimisation and BC-IAM), REMIND for cost-effectiveness models (policy optimisation and DP-IAM), and IMAGE for biophysical impact models (policy evaluation and DP-IAM).

Before proceeding to the evaluation of the three IAMs, I will take a closer look at the literature around model evaluation in Chapter 2. For IAMs, which include elements of natural science as well as economic models, the perspectives provided by the philosophy of modelling are especially interesting. I will discuss how models can represent a target system despite their manifold idealisations, what it means for a model to be adequate for its purpose, and how the application of models relies on interactions between modellers, audiences and purposes, among other elements. Subsequently, I will review how the literature on IAM evaluation has changed since the 1990s, how useful existing approaches are, and how the perspectives on model evaluation can help to enrich the notion of expectations for IAMs. In Chapters 3-5, I evaluate DICE, REMIND and IMAGE, respectively. The focus is placed on the relative strengths and weaknesses of these models, by analysing how they represent their target systems, how and for which purposes they are used, and which interpretations modellers suggest for their IAMs. On this basis, I examine possible expectations of the three models and judge the respective IAMs against these. In Chapter 6, I compare findings from the three evaluated IAMs, with the aim of assessing which expectations link best with each model, and how well the respective models clarify what could reasonably be expected of them. Finally, in Chapter 7, I reflect on the performed evaluation and highlight what can be gained by adopting the perspective of expectations. Based on the findings of previous chapters, I conclude with a proposal to place more emphasis on interpreting IAMs and developing better model commentaries, such that model evaluation can be based on clearly formulated expectations.

2 Model Evaluation and Expectations

The term 'model' can have very different meanings – ranging from architecture and fashion all the way to climate science. Even when restricting the scope to science alone, the term can refer to phenomena as varied as the standard model in physics, a model organism in biology, a box model for stocks and flows, or a dynamical computer model. Despite this multiplicity of usages, Mäki (2001) suggests that all models have one thing in common: they "are used to represent something beyond themselves" (p. 9936). A model is thereby a representation of something else, called the *target system* (Frigg & Hartmann, 2020; Knuuttila, 2011). As partial representations of a target system, some authors argue, models are neither true nor false (Bailer-Jones, 2003). Instead, models should resemble their target system closely enough – i.e. be a sufficiently realistic representation – to allow investigations into the real-world features of interest (Mäki, 2005). IAMs, in Mäki's approach to representation, would aim to be sufficiently realistic representations of their target system, e.g. the coupled socioeconomic and climate system on a centennial timescale (see Figure 1 for a conceptual overview of the involved systems). Accordingly, evaluating IAMs would involve checking the degree of fit or similarity between the model and the represented real-world systems.

Analysing models as representations of target systems is complicated, however, by the question of how to determine the success of this representational relationship (Knuuttila, 2011). In other words: when is the model realistic enough? Who decides that and according to which criteria? By necessity, a model is less detailed than its target system, such that the representation involves unrealistic assumptions and simplifications (Bailer-Jones, 2003; Mäki, 1992). How, then, can a highly idealised model with false assumptions still be of help? Some authors argue that models are able to yield insights into their target system despite unrealistic assumptions, because modelling involves isolating certain causal mechanisms of interest (Cartwright, 2009; Mäki, 1992). This method of isolation is akin to scientific experimentation – all causal factors except for the one under investigation are artificially sealed off through deliberately false assumptions (Mäki, 2005). In the context of economic modelling, however, this approach – starting from a target system and then continuously sealing off confounding influences until one gets to the isolated core – has been challenged by Sugden (2000), on the grounds that modellers don't actually proceed like that. Rather, he argues, models in economics should be seen as constructions that behave sufficiently credibly for making inferences from the world of the model to the real world. Both accounts agree, however, that highly idealised models can be used to explain real-world phenomena. In the first view, this is possible because idealised models isolate some core causal mechanism of the considered target system. In the second view, it is possible because models provide a credible account of how the world – or an aspect of it – could be. IAMs rely strongly on idealisations (Staub-Kaminski et al., 2014). Evaluating them would therefore entail to specify which core mechanism of the target system has been isolated, or by which merits the model can be considered a depiction of a credible world.

The focus on models and their representational relationship with a target system leaves important questions open, for example: What is the purpose of the model? Parker (2020) rejects the idea that evaluation could proceed solely by judging "how accurately and completely a model represents a target" (p. 458). Instead, she argues that model evaluation should assess

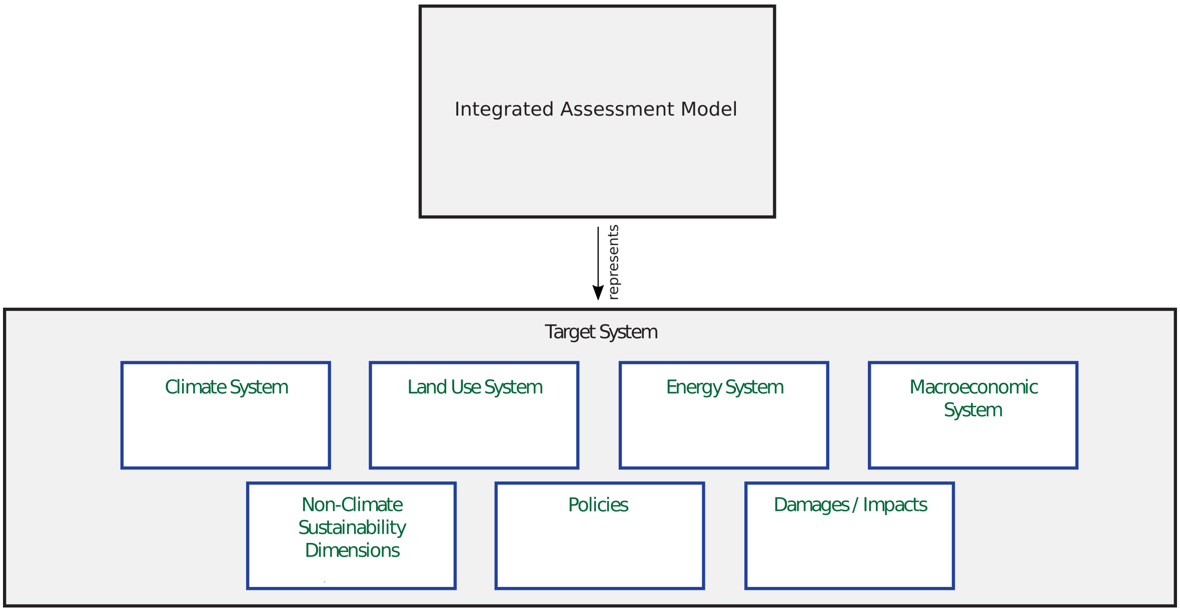


Figure 1: Conceptual framework of an IAM as a representation of its target system, which consists of various subsystems.

model quality only relative to a given purpose – by judging whether the model has properties that make it suitable for this purpose. This *adequacy-for-purpose* approach entails a change of focus – away from evaluating the model as such, towards evaluating its adequacy for a particular purpose – and places more emphasis on the tool-like characteristics of a model. Models can thereby be seen as "representative tools" (Parker, 2009, p. 235), in that they are related to a target system, but the required realisticness of representation is determined by the purpose of the tool (see Figure 2). On a similar note, Knuuttila (2011) argues that seeing models as *epistemic tools* allows for analysing how models are actually used and how they are able to yield knowledge. For IAMs, this implies that concrete purposes and applications have to be considered. These purposes can consist of comparing different mitigation pathways, estimating economically 'optimal' emission pathways, understanding systemic interactions, or highlighting key uncertainties (Weyant, 2017).

IAMs are used for a variety of purposes, which can include analysing 'what is' questions as well as 'what could be' or 'what should be' questions (Schwanitz, 2013, p. 124). Against this backdrop, it is helpful to distinguish different functions that a model can fulfil. Knuuttila and Morgan (2012, p. 73) propose six epistemic functions of economic models: suggesting explanations, carrying out experiments for policy advice, making predictions, deriving solutions to theoretical problems, exploring the limits and range of possible outcomes of the model, and developing theory. Depending on which function the model is supposed to fulfil, it will be expected to adhere to different standards. For instance, a famous instrumentalist position holds that an economic model meant for prediction should not have to worry about unrealistic assumptions, as long as it "yields sufficiently accurate predictions" (Friedman, 1953, p. 15). Conversely, if the model is meant to explain causal mechanisms, Sugden (2000) argues that models should persuade about the credibility of their assumptions, while predictive skill can be considered secondary. Models, as seen in the case of IAMs, can fulfil different functions by being purposefully constructed and manipulated – and exactly this versatility turns them into useful epistemic tools (Knuuttila, 2011). For IAMs, the generation of policy-relevant knowledge is especially important. In order to fulfil this function, models are often used by varying assumptions, exploring the scope of possible policy consequences and analysing trade-offs between policy options

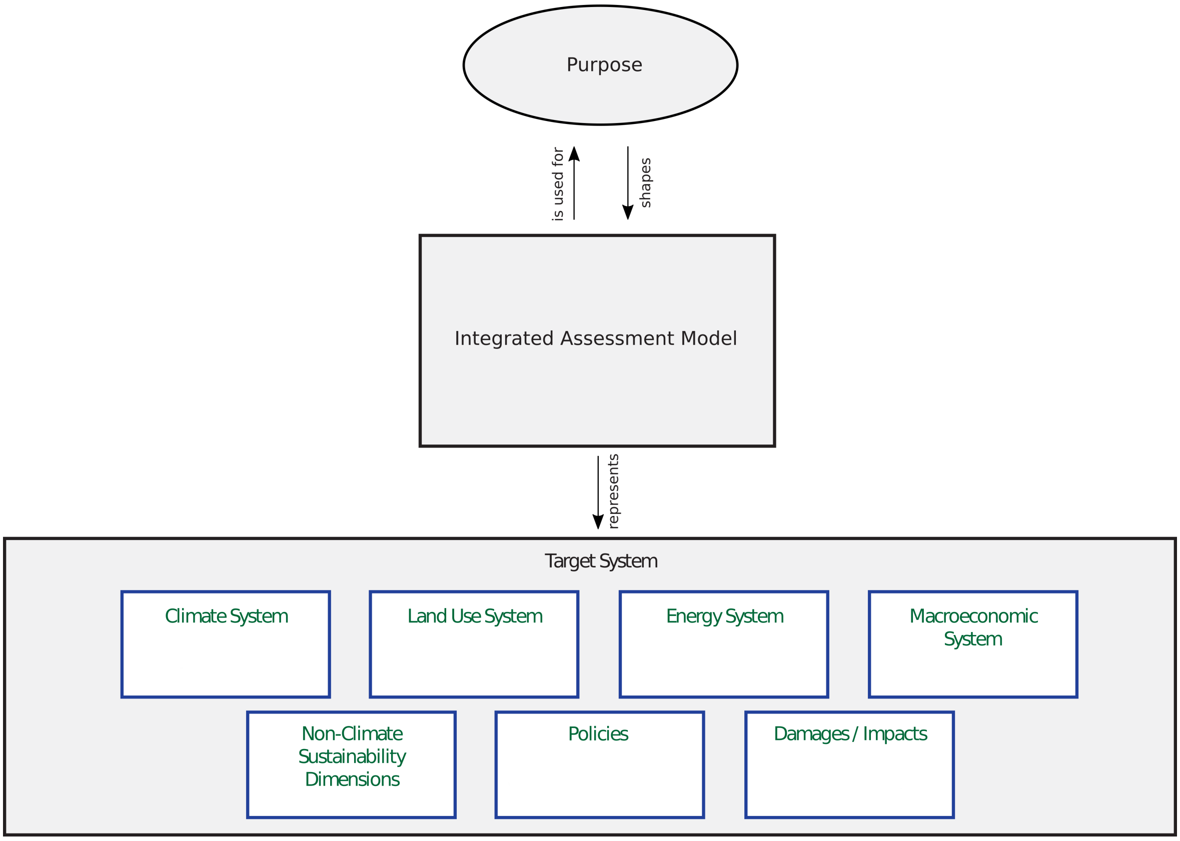
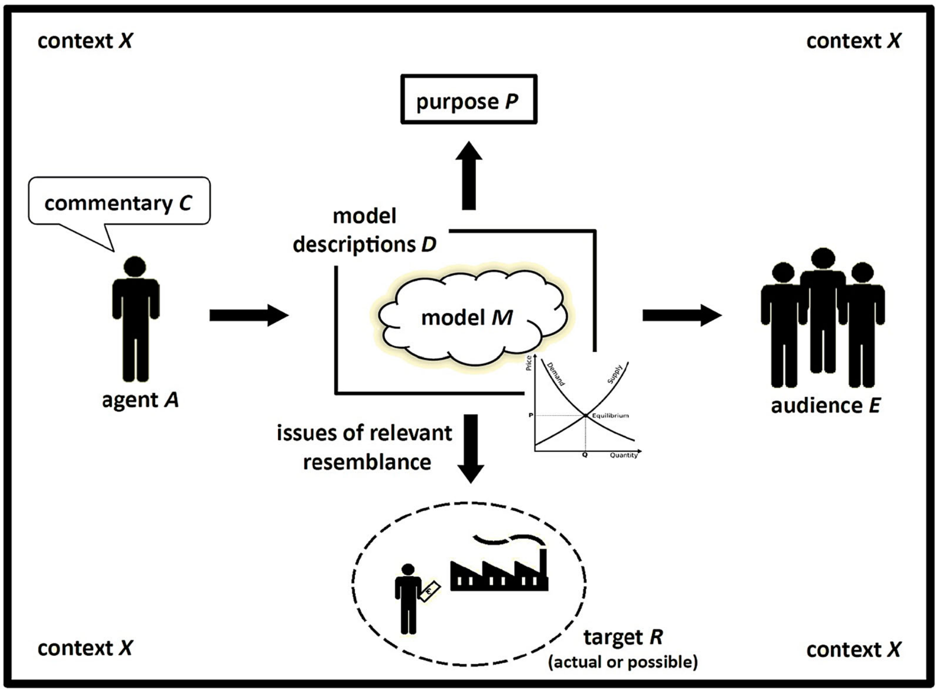


Figure 2: Conceptual framework of an IAM and its target system, in addition to the purpose it is being used for and shaped by.

(Knuuttila & Morgan, 2012; Kowarsch, 2016b). At this point, however, IAMs leave the realm of purely epistemic functions and further considerations about political and ethical consequences of model-based policy advice enter the picture (Frank, 2017; Kowarsch, 2016b).

A different perspective on modelling can be taken by considering *stories* as an integral part of models and model application. Gibbard and Varian (1978), for example, define a model as the combination of “a story with a specified structure” (p. 666). Thereby, the structure is made of assumptions and mathematics and can be considered an uninterpreted system, where only the story makes sense of the structural elements. Morgan (2001) builds on this definition by arguing that storytelling is also an important part of using a model. Thereby, a model in use is always accompanied by a story that links it to real world. The link between model world and real world is established by providing narrative answers to specific questions, often ‘what if’ questions. Here, the connection to IAMs is obvious, as they are often characterised as tools for analysing ‘what if’ questions (Anderson & Jewell, 2019; Beck, 2017; Gambhir, 2019; Nordhaus, 2014; Weyant, 2017). Many IAM uses can thus be analysed through the lens of stories. The SSP scenarios, for instance, are a set of five storylines which are quantified by structurally very different IAMs (van Vuuren et al., 2017). Therefore, by looking at the stories told through model use, we can evaluate not only the static properties of what a model is, but also the variable questions that arise during its application.

In order to assess how models operate in a context of different purposes and users, Mäki (2009) developed a framework that captures the dynamics of various elements of the modelling process (see Figure 3). The distinction between the model and its target system forms the basis, but is complemented by several other factors. Most importantly, the framework includes an

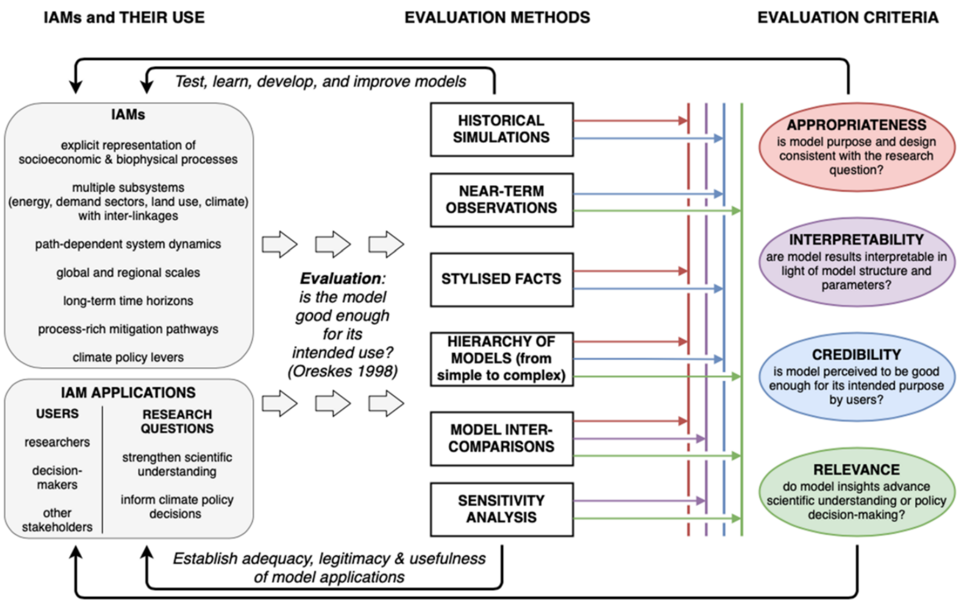


Source: Mäki (2018, p. 221)

Figure 3: Analytical framework of the modelling process by Mäki. An agent A uses a model M to represent a target R for a given purpose P, addressing an audience A and applying a model commentary C that identifies and aligns all other modelling components (Mäki, 2009).

agent who decides to use the model because they consider it adequate for a certain purpose. Additionally, there is an audience that is being addressed by the agent and confronted with the modelling results. It is the responsibility of the modelling agent to provide the audience with a model commentary which is supposed to convey “ideas about how the other components in the modelling endeavour play out their roles in coordination with one another. What is the point of using radically unrealistic assumptions? [...] What’s the proper domain of application of a model? What precise purpose(s) can a given model be used for? What uncertainties are involved in model use?” (Mäki, 2018, p. 222) The inclusion of an agent that chooses and uses the model and of an audience that is addressed, allows to illuminate some of the dynamics around IAMs. Agents, for example, can be big modelling groups, individual researchers that use a given IAM, or the developer of the model. The audience for IAM can vary between policymakers, the scientific community, climate modellers, or the general public. According to this framework, it is the task of the agent to deliver a model commentary that takes the different possible purposes of IAMs and the varying audiences into account. Such a commentary would therefore have to consider several of the previously discussed aspects, such as the credibility of the modelled world, the stories told through modelling, or the political and ethical implications of specific assumptions and idealisations.

The question of IAM evaluation has also concerned modellers themselves, and many views relate to philosophical discussions. Morgan and Dowlatabadi (1996) analyse the target system of IAMs by asking what exactly the ‘climate problem’ is and who the climate decision makers are. Based on that, they evaluate whether the idealisations at the heart of IAMs enable them to capture the main mechanisms of the target system. For example, they argue that the idealisation of modelling the whole world by a single decision maker fails to capture one of the most



Source: Wilson et al. (2021, p. 14)

Figure 4: Framework consisting of evaluation methods which contribute to the testing of IAMs against four evaluation criteria (Wilson et al., 2021).

"fundamental characteristics of the climate problem", such that IAMs "may confuse more than they clarify" (pp. 339f). Parson (1996) reflects more broadly about the criteria that should be used for evaluating IAMs. He argues that IAMs should not be judged by the same standards as the disciplinary models that they consist of, because trying to fulfil all disciplinary standards simultaneously would make the integration into a single model impossible. Instead, evaluation criteria should emerge based on the epistemic purpose of the modelling exercise and on the requirement of delivering useful policy advice. Risbey et al. (1996) also propose that an overall evaluation of IAMs should be dependent on their purpose, especially when considering which factors to include into the model. Additionally, they argue that components of IAMs should be evaluated based on disciplinary standards, while interdisciplinary standards should be developed to evaluate the process of combining different disciplinary building blocks. In this respect, the authors emphasise the distinction between *heuristic tools* and *forecasting tools* and argue that, while IAMs can be used in both ways, an IAM evaluation should always consider whether model results are meant to yield qualitative insights (heuristic tools) or quantitative predictions (forecasting tools). Lastly, Risbey et al. (1996) argue that modellers should document IAM assumptions and explicitly examine their implications, and call for greater diversity of approaches towards IAMs.

More recently, Schwanitz (2013) has revisited the issue of IAM evaluation. She argues that IAMs got increasingly complex as well as increasingly relevant to the policy process, such that more effort should be put into evaluating them. In her view, this evaluation "should be understood as a continuous effort of testing whether the model can fulfill its purpose" (p. 121). Based on this definition, a model is never fully 'validated', but wound up in an iterative process of ongoing model evaluation. Concretely, Schwanitz (2013) proposes the following steps: "setting up an evaluation framework, scrutiny of the conceptual model, code verification and model documentation, model performance tests, uncertainty and sensitivity analysis, documentation of the evaluation process, as well as communication to stakeholders" (p. 125). Wilson et al. (2021) develop another systematic evaluation framework, specifically for DP-IAMs. They justify this focus by emphasising the differences between the two types of IAMs. In their view, DP-IAMs are “complex ‘black boxes”’ (p. 3) that require enhanced interpretability and transparency, while BC-IAMs are widely accessible, but requiring critical discussions of assumptions and general modelling approaches. Implicitly, they thus separate DP-IAMs from any discussions on assumptions and general approaches and limit themselves to technical evaluation approaches. The authors propose a framework consisting of four evaluation criteria (appropriateness, interpretability, credibility, and relevance) which are assessed through six methods: historical simulations, near-term observations, stylised facts, model hierarchies, model inter-comparison projects and sensitivity analysis (see Figure 4). Thus, in combining rather abstract definitions of IAM evaluation with concrete methodologies, Wilson et al. (2021) provide a good overview of existing methods, but little guidance on how IAMs could be evaluated with a less methodological focus.

From the vantage point of pragmatist philosophy, Kowarsch (2016b) undertakes an evaluation of IAM-based literature underlying the IPCC reports. He bases it on a perspective of circular science-policy interaction, according to which possible policy pathways should first be comprehensively mapped, subsequently be evaluated by society, and finally be iteratively readjusted based on the societal judgement (Edenhofer & Kowarsch, 2015). The three evaluation criteria that emerge from this view are: “relevance for the exploration of policy pathways, transparency and diversity of value judgements, and scientific and epistemic quality” (Kowarsch, 2016b, p. 174). The first criterion involves analysing whether IAMs address the actual problems that climate policymakers are faced with and whether IAMs are capable of outlining the implications of different climate policy choices. The second criterion involves analysing whether IAMs incorporate “alternative and disputed ethical viewpoints” (p. 184) including their implications, and whether IAMs are sufficiently transparent about their ethical viewpoints and the reasoning behind them. Lastly, the third criterion involves analysing the credibility and reliability of the scientific material underlying IAMs, with a special focus on the treatment of uncertainties and questions regarding value judgements and objectivity.

In light of the question of what can reasonably be expected of IAMs, how useful are the reviewed approaches to modelling and model evaluation? The literature on IAM evaluation yields a broad range of criteria. Yet, it is not always clear which ones to choose and how to operationalise them. The philosophy of modelling, on the other hand, yields a variety of perspectives on what models are and how they are used for epistemic purposes. Yet, it seems unlikely that there is any one approach that fits perfectly to IAMs. Further, these philosophical accounts of modelling mainly focus on epistemic aspects, while IAMs are fundamentally tied to politics and ethics, in addition to their epistemic characteristics (Beck & Krueger, 2016; Funtowicz & Ravetz, 1994; van der Sluijs, 2002). Therefore, an evaluation of IAMs requires an approach that is able to incorporate the broader sociopolitical context of IAMs as well as its epistemic aspects. This leads me to propose the notion of expectations – what is expected of IAMs, how can we assess whether they live up to it, and which expectations of IAMs are justified in the first place? Answering any of these questions requires some consideration of users holding an expectation, of the purpose that modellers have in mind for the model, and of an evaluation criterion that its performance can be judged against. It is through this triad that I define an expectation: as an evaluation criterion combined with a modeller’s purpose and a model user’s aim. Hereby, I deviate from Mäki’s distinction between a modelling agent and an audience, as the term ‘audience’ does not fully capture the various roles that non-modellers play in the application of IAMs. I call everyone a users, if they have a connection to or an interest in the modelling exercise without running the models themselves. ‘Modeller’ conversely refers to model developers and those researchers that are engaged in the process of model advancement.

The focus on expectations of IAMs as the main unit of analysis allows me to incorporate several of the above-mentioned approaches into the evaluation of DICE, REMIND and IMAGE.

For example, the criterion of realistically representing a target system could be a promising starting point for evaluating a given model. Subsequently, this criterion can be linked to a purpose as specified by the modeller, and to certain aims or demands from model users. Model use could be evaluated through the stories that are told, by assessing the credibility of the modelled worlds, or by analysing how the model operates as an epistemic tool. Again, in order to form an expectation, these criteria subsequently have to be complemented with purposes and user perspectives. By analysing IAMs in light of different approaches to evaluation, possible expectations for them emerge. These can subsequently be critically evaluated against what a given model can actually deliver. Through this procedure, the evaluation of DICE, REMIND and IMAGE does not have to postulate a single set of criteria for all IAMs, but can instead be adapted to the particularities of the respective models.

3 DICE

The Dynamic Integrated Climate-Economy (DICE) model was introduced by William Nordhaus (1992). It is a BC-IAM and can be characterised as a policy optimisation model (Nordhaus & Sztorc, 2013). DICE is one of the most influential IAMs (Aldy & Stavins, 2020) and in 2018, its developer William Nordhaus was awarded the Nobel Memorial Prize in Economics "for integrating climate change into long-run macroeconomic analysis" (Nobel Prize Outreach, 2022). I base the following evaluation on DICE-2016R2, as described in Nordhaus (2018b). For more detailed elaborations on elements that have not changed between model versions, I often turn to the latest full model documentation (Nordhaus & Sztorc, 2013) and the background provided in Nordhaus (2013).

3.1 Target System of DICE

The aim of DICE is to model the whole causal loop of climate change policy, from greenhouse gas (GHG) emissions through global warming and impacts back to political measures that affect GHG emissions (Nordhaus & Sztorc, 2013). The target system of DICE can therefore be conceptualised as a combination of the climate system, the global economy and global climate policy. In a very simplified form, DICE tries to represent all these elements (see Figure 5).

The DICE climate module translates an emissions path[1] into a path describing global mean surface temperature. Subsequently, a damage function translates different levels of global temperature into associated economic damages, by generating a loss ratio relative to a given production level[2]. Finally, the model includes a representation of global macroeconomic quantities such as population or GDP, and a mechanism for deciding on the amount of climate change policy. In DICE, the macroeconomic structure is represented by a neoclassical economic growth model (Nordhaus, 2017), and the decision structure is based on maximising a social welfare function which takes into account both the benefits of avoided climate damages and the costs of climate mitigation. These climate mitigation options are modelled through mitigation cost curves, for which the cost of reducing industrial and energy-related CO_2 emissions is globally aggregated. The cost of pursued mitigation policies and economic damages of climate change finally close the loop by affecting GDP and GHG emissions.

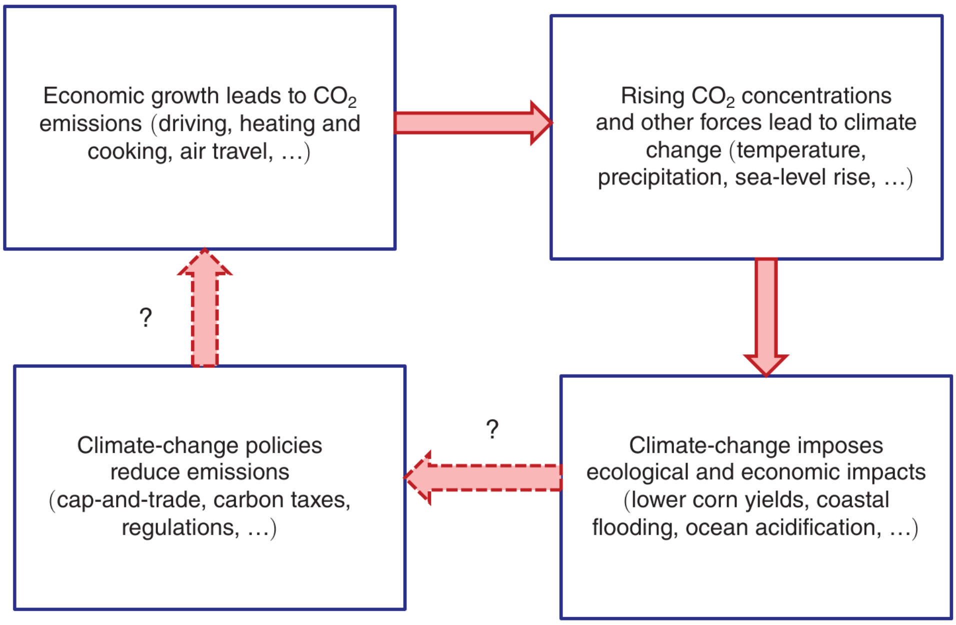
As a starting point for the evaluation of DICE, I will analyse how well it represents its target system. In turn, I look at the climate module, damage function and macroeconomic module of DICE, and critically evaluate the modelling choices taken and, where possible, their effect on model results.

The DICE climate module is purposefully kept parsimonious for the sake of transparency and tractability (Nordhaus & Sztorc, 2013). It consists of a three-box carbon cycle model[3] and equations for translating the atmospheric carbon content into radiative forcing and ultimately

[1] I use 'path' or 'pathway' as equivalent terms for 'time series'.

[2] For simplicity, I use the terms 'production', '(economic) output' and 'GDP' interchangeably.

[3] One box stands for the atmosphere, another for the upper ocean and biosphere, and the third box for the deep ocean.



Source: Nordhaus (2019, p. 1995)

Figure 5: Structure of DICE along the causal loop of climate change policy, as presented by Nordhaus (2019) in his Nobel Prize lecture. The dashed arrows with question marks indicate that the links between climate damages and implemented policies are not fully in place yet – however, they are represented in the DICE model.

into global mean surface temperature. Through this concise formulation, the aim is to capture the main mechanisms at work in the climate system. However, Dietz et al. (2021) have shown that the qualitative dynamics of carbon cycle and temperature are inconsistent with the behaviour of Earth System Models (ESMs)[4]. For one, they find that DICE overestimates the time between GHG emissions and subsequent warming, which makes climate damages appear as to occur further in the future. Secondly, DICE overestimates the absorptive capacity of the oceans – while ESMs project a decrease of the ocean's function as a carbon sink with further climate change, the DICE climate module models an increased carbon sink with further emissions (Dietz et al., 2021; Hall & Behl, 2006). Several studies have adapted the DICE model by introducing formulations of the climate module that are more in accordance with ESMs (Calel & Stainforth, 2017; Dietz et al., 2021; Hänsel et al., 2020). These studies show that changes in the climate system representation can alter the resulting warming level by about half a degree.

The climate sensitivity parameter (describing the amount of long-term warming caused by a doubling of CO_2 concentrations) has been at the centre of several controversies around IAMs (Pindyck, 2013b; Weitzman, 2009b). This is due a combination of its importance for estimating future climate change and its high uncertainty – the parameter range considered likely spans several degrees (IPCC, 2021, p. 7:197; Roe & Baker, 2007). As a consequence, this parameter can change the economically 'optimal' warming level by up to a degree (Gillingham et al., 2018; Glanemann et al., 2020). Weitzman (2009b) has shown that fat tails of the climate sensitivity probability distribution[5] could introduce structural instability into cost-benefit analyses of climate change. On that basis, he argues that the widespread practice of representing uncertain

4 Dietz et al. (2021) compare DICE and other IAMs to a best-fit of the CMIP5 ensemble.

5 'Fat tail' here refers to the behaviour of the probability distribution for high levels of warming, where the associated probability declines extremely slowly (Roe & Baker, 2007). As a consequence, catastrophic high-level warming still has a non-negligible probability.

climate sensitivity by a best-guess estimate makes "IAM-based CBAs [...] especially and unusually misleading" (Weitzman, 2009b, p. 18). On the other hand, Roe and Bauman (2013) caution not to overemphasise the small probability of large warming, as this warming would only ensue on a very long timescale.

Overall, the DICE climate model is successful at capturing the general dynamics of climate change in a small number of equations. However, it suffers from two main shortcomings: the inconsistency of its qualitative dynamics with ESM projections and the disregard of uncertainty in key parameters. As to the first point, DICE can be and has been adapted to include a climate model that behaves more realistically (e.g. Hänsel et al., 2020). Similarly, the exclusion of uncertainty can be tackled through stochastic model formulations (Cai et al., 2015; Cai & Lontzek, 2019; Dietz & Stern, 2015; Lemoine & Traeger, 2016).

The damage function of DICE has been criticised for a long time (Ackerman & Finlayson, 2006; Azar, 1998; Howard & Sterner, 2017; Keen, 2020; Moore & Diaz, 2015; Pezzey, 2019; Weitzman, 2009a). One major point of critique concerns its functional form. DICE uses a quadratic function which is fitted to damage estimates from the literature. The main reason for this functional form seems to be that it is the most simple nonlinear convex function (Nordhaus & Moffat, 2017). However, it is unclear whether a quadratic function is a good representation of actual economic damages from future climate change (Ackerman & Finlayson, 2006; Pindyck, 2017). While Nordhaus and Moffat (2017) argue that the exponent of the polynomial damage function does not matter much and is rather overestimated than underestimated, Weitzman (2012) contends that such a quadratic damage function highly underestimates the impacts of low-probability catastrophic warming. Independent of the functional form, the calibration of the damage function is controversial. Howard and Sterner (2017) demonstrate the difficulties of providing estimates by calculating values which are several times higher than Nordhaus and Moffat (2017) – based on the same set of studies. Nordhaus (2013) recognises the problems in quantifying climate damages by stating that "the usefulness of this approach [is limited] for catastrophic climate change" (p. 1084).

In the context of uncertainty about climate sensitivity, Weitzman (2009b) showed that low-probability, high-impact events significantly alter the results of an expected utility calculation about climate change. This lack of robustness is exacerbated, he finds, by the fact that two equally defensible formulations of a damage function yield wildly different utility losses (Weitzman, 2010). On the one hand, the standard multiplicative formulation, as used in DICE, implicitly assumes substitutability between consumption and climate damages, as might be adequate in a case where impacts mainly affect material wealth. On the other hand, the additive formulation assumes weak substitutability between consumption and damages, as might be adequate when impacts affect ecosystems or health (Bastien-Olvera & Moore, 2021; Sterner & Persson, 2008). The estimation of economic damages from climate change is thus inextricably linked to choices about the degree of substitutability between material wealth and natural capital (Drupp & Hänsel, 2021).

Beyond that, there is a discussion about how to model climate damages in general – as a reduction of economic production at a single point in time only or as a reduction of economic growth which has long-term negative effects on economic production. The DICE model uses the former approach, in which damages occur instantaneously and have no longer-lasting adverse effects on the economy. Yet, Moore and Diaz (2015) argue that damages induced by climate change are likely to have effects on economic growth rates, especially on those of low-income countries. The general question of whether impacts should be modelled as damaging output flows or capital stocks has lead to a big debate (Burke et al., 2015; Kalkuhl & Wenz, 2020; Stern, 2013), and changes to the associated assumptions have been shown to alter the economically 'optimal' level of warming by around one degree (Glanemann et al., 2020; Moore & Diaz, 2015). On top of all the difficulties of providing a realistic representation of future climate damages, there is the uncertain modulating factor of adaptation (de Bruin et al., 2009). Damages from climate change can be reduced through adaptation, depending on societies' adaptive capacities and the amount of resources mobilised for that purpose.

Overall, the damage function of DICE is likely not a very realistic representation of real future climate damages. This is partly due to the immense difficulties of projecting damages into a future with global temperatures for which there is no empirical data, and further exacerbated by degrees of freedom with respect to functional form and by arguably normative assumptions, e.g. about the substitutability between material and natural capital. Recent studies, however, point to an emerging consensus that climate damages will likely be higher than projected by DICE (Burke et al., 2015; Diaz & Moore, 2017; Howard & Sterner, 2017; Kalkuhl & Wenz, 2020; Piontek et al., 2021).

DICE models macroeconomic dynamics through a neoclassical growth model which represents the following dynamics: Output is generated based on capital, labour and technology, and subsequently reduced by damage and mitigation costs (Nordhaus & Sztorc, 2013). The remaining net output is distributed on present-day consumption and investment, where investment leads to higher capital stocks in the future. Labour, which is proportional to global population (Nordhaus, 2018b), is included as an exogenous variable based on UN projections (Nordhaus & Sztorc, 2013). Similarly, technology (total factor productivity, TFP) is calibrated such as to make the model align with projections of global GDP by Christensen et al. (2018). Capital finally emerges endogenously, based on allocation decisions at each time step.

Importantly, the model operates with only a single commodity which represents the global economy. This means that the model is blind to heterogeneities between or within regions, thereby excluding ethical and political issues of inequality (Asefi-Najafabady et al., 2020; Farmer et al., 2015; Jafino et al., 2021). Further, doubts about DICE's capability of modelling long-term macroeconomic dynamics are raised due to the assumption of exogenous technological change (Acemoglu et al., 2012; Popp, 2004) and the exclusion of energy in the production function (Edenhofer et al., 2005; Keen, 2020). Nonetheless, the neoclassical economic growth model is considered one of the most important models in macroeconomics, due to its ability of giving insights into mechanics of economic growth (Acemoglu, 2009, p. 318). Similarly, Nordhaus and Sztorc (2013) justify the use of this model with the fact that it is "standard to the economic growth literature" (p. 8). Yet, the authors also acknowledge that, due to the "very long time frame" of the model, projections and assumptions are based on "very thin evidence" (p. 8).

In fact, Millner and McDermott (2016) have shown that the DICE model is unable to reproduce 20th-century patterns of economic growth. The authors compare the trajectories for TFP and GDP generated by DICE with historically realised values and conclude that "the version of the neoclassical growth model that DICE relies on could be subject to structural errors on the temporal scales relevant to climate polices [*sic*]" (Millner & McDermott, 2016, p. 4).

3.2 Purposes of DICE

William Nordhaus has always emphasised that DICE is supposed to provide guidance for climate policy (Nordhaus, 1992, 2018a). Mainly, this is done by calculating economically 'optimal' levels of climate change mitigation and global warming, by comparing alternative policy pathways, or by estimating the SCC. Nordhaus grants that this policy guidance can not be based on exact values because of the "highly simplified representations of the complex economic and geophysical realities" (Nordhaus, 2013, p. 1093) that DICE embodies. However, in his view, this does not diminish its value, because IAMs are supposed to act as conceptual frameworks for analysing complex and uncertain systems (Nordhaus, 2014). Consequently, the value of DICE is not given by merits of being an accurate representation of real-world processes, but by its ability to conduct 'if ..., then ...' analyses (Weyant, 2017).

This entails a change of focus from analysing how the model represents to analysing how the model is used. Morgan (2001) suggests that model use can be captured by analysing the stories that are employed to answer 'what if?' questions. Stories, in this sense, are the devices that link models to the world. Evaluating the ability of DICE to conduct 'if ..., then ...' analyses therefore entails an analysis of the story that is being told about how the mathematical structure links to actual climate policy. Further, for an 'if ..., then ...' analysis conducted by DICE to claim authority, it has to provide reasons to believe that causal mechanisms in the model can tell us something about real-world mechanisms. For Sugden (2000), this involves that the model depicts a 'credible world' – which requires assumptions that are internally coherent and in resonance with "what is known about causal processes in the real world" (p. 26).

In his own description of how DICE is used, Nordhaus (2013, p. 1095) mentions several applications: 1) generating consistent projections; 2) calculating impacts of alternative assumptions; 3) estimating uncertainties and their effect on model output; 4) evaluating costs and benefits of alternative policies. Of the six epistemic functions listed by Knuuttila and Morgan (2012), the DICE purposes according to Nordhaus come closest to 'exploring the limits and range of possible outcomes' and 'carrying out experiments for policy advice'. One might ask, however, which properties of DICE warrant that the model is actually adequate for achieving these purposes.

The internal consistency of DICE projections is ensured by the abstract formalism of the model, as the mathematical equations come with basic accounting relations and the components of the model are fully coupled. Therefore, no model quantity is 'lost' and internal contradictions can be excluded.

Due to its compact formulation, the DICE model is relatively accessible and can be scrutinised and modified by different actors. By being completely open-source, anyone can construct own versions of the model or play around with different settings of the standard model. Consequently, there are a range of different versions of DICE[6] and a vast literature that examines the DICE model under different assumptions (Ackerman & Finlayson, 2006; Dietz & Stern, 2015; Glanemann et al., 2020; Grubb et al., 2021; Hänsel et al., 2020; Sterner & Persson, 2008). This property of DICE resonates with Knuuttila's (2011) account of models as epistemic tools, whereby modellers "learn from models by constructing and manipulating them" (p. 267). The concise structure, together with the openly available code, makes DICE into an important vehicle for assessing different assumptions, for academic discussion about them, as well as for estimating the impact of uncertainties.

Finally, is DICE adequate for evaluating costs and benefits of alternative policies? Having analysed how well the climate module, damage function and macroeconomic module represent their target system, DICE does not seem capable of providing accurate assessments of costs and benefits. Nordhaus readily admits the limited ability of DICE to project future conditions, yet also cautions to "recognize that the key issue about the uncertainty about long-term projections

[6] Examples include a regionalised version RICE (Nordhaus, 2010), R&DICE with induced innovation (Nordhaus, 2002), ENTICE with endogenous technological change (Popp, 2004), AD-DICE with adaptation (de Bruin et al., 2009), a stochastic version DSICE (Cai et al., 2012), gro-DICE with climate damages impacting economic growth (Moore & Diaz, 2015), and NICE which includes a inter- and intraregional inequality (Dennig et al., 2015).

is whether they have a large impact upon current policies" (Nordhaus, 2013, p. 1106). Previous analysis suggests that different specifications of climate sensitivities or climate damages, for example, can indeed have a large impact on policy-relevant model output. This does not necessarily mean that DICE is unsuited for evaluating costs and benefits. It implies, however, that a single model run with a certain set of assumptions can not claim much authority. Rather, it is the comparison of different model assumptions and specifications that leads to insights about key dynamics and trade-offs, and only the accumulation of many different model runs can hope to shed a light on plausible ranges for costs and benefits.

On similar terms, Morgan (2001) notes that "it is only by asking questions and telling stories that we explore and demonstrate the full range of features and outcomes" (p. 369) entailed by the model. In this sense, it is a valid epistemic strategy to explore different parameter settings and experiment with different model structures in order to get a grasp of possible stories that DICE could tell about climate change and mitigation policies. The plausibility of a single story, however, must be subject to subsequent scrutiny. For a baseline scenario as modelled by DICE, for example, it is unclear whether it really depicts a credible world of no further climate policy, given the low damage estimate and omission of heterogeneities in the world economy. Overall, DICE can be useful to the evaluation of costs and benefits, but not in any conclusive way. Rather, it serves as a tool for telling a range of more or less credible stories, which can be used as input into a more encompassing debate about costs and benefits of different policy options.

In order to provide useful guidance to policy and achieve the stated goals of DICE, transparency about model assumptions and their effects on model results is key (Weyant, 2014). The recent literature on IAMs emphasises the importance of making code and basic documentation openly available (Bistline et al., 2021; Robertson, 2020; Schwanitz, 2013; Skea et al., 2021) Accordingly, new model versions of DICE are accompanied with updated model documentations, in which changes are outlined. In these as well as in other papers, Nordhaus engages with debates about assumptions on discounting (Nordhaus, 2011), climate damages (Nordhaus & Sztorc, 2013), technological change (Nordhaus, 2013) or the possibility of catastrophic warming (Nordhaus, 2009).

Further, many authors argue that it is not sufficient to make code and documentation transparent. Instead – in line with Mäki's (2018) concept for model commentary – it is considered the responsibility of modellers to proactively make "structural assumptions explicit [...] and communicat[e] value-laden assumptions" (Bistline et al., 2021, p. 11; see also Beck & Krueger, 2016; Kowarsch, 2016b; Skea et al., 2021). At first sight, DICE can be considered to live up to these standards, as Nordhaus frequently positions himself on ethically sensitive and controversial topics. Yet, it is unclear whether that is a response to the frequent criticism of DICE or an intrinsically motivated effort of highlighting ethically sensitive modelling choices. While discounting is discussed at great length, issues such as distributive justice or substitutability of material and natural capital are barely mentioned in the DICE documentation[7]. By being open-source and comprehensive enough to be ran on private computers, DICE enables other researchers to point out critical and ethically sensitive assumptions – as has been done for distributive justice (Ackerman et al., 2009; Budolfson et al., 2017; Dennig et al., 2015; Stanton, 2011) and substitutability and relative prices (Bastien-Olvera & Moore, 2021; Drupp & Hänsel, 2021; Sterner & Persson, 2008). However, the model commentary by the original developer carries special weight in the debate. Therefore, a selective discussion of assumptions, as in the DICE documentation, can be problematic. While DICE offers all the ingredients for an open and transparent discussion of critical assumptions and their implications, Nordhaus himself does

[7] Nordhaus (2013) discusses the simplification of "a single commodity to represent all consumption, investment, and public goods and services" (p. 1093). However, he discusses only the implications for modelling international trade and not the distributive consequences.

not always live up to the standards of providing a comprehensive model commentary.

3.3 Interpreting DICE

So far, I have not analysed the decision structure of DICE. The model decides on the amount of production spent for consumption, investment and for climate change mitigation in each period by maximising a social welfare function. This function is the sum of discounted utilities over the modelling time frame. Utility of a given time step is based on per-capita consumption levels and calculated through a concave function. This implies that additional consumption yields less additional utility as consumption levels increase. The parameter η that determines the curvature of the utility function is often called *inequality aversion*[8] (Nordhaus, 2017). As a consequence, higher consumption levels (which are assumed to be in the future) contribute less to social welfare than lower, present-day consumption levels. However, this parameter is not able to capture all kinds of inequalities that are relevant for climate policy. Because of being idealised such as to only represent one global commodity, DICE cannot account for consumption inequalities between different world regions (Sterner & Persson, 2008)[9].

On top of being valued less on the basis of inequality aversion, future utilities are also directly discounted through a parameter δ, the *rate of pure time preference*. Utilities in future time periods are thereby valued lower than present-day utilities purely because of their temporal distance. In the context of climate change, discounting choices are particularly important because of the long timescales involved. While the costs of climate mitigation are mainly to be borne by earlier generations, the benefits of mitigated climate change will fall primarily on later generations. Thereby, the relative weights which are placed utilities in the near as opposed to the far future, are a crucial parameter with large effects in the results of IAMs (Drupp et al., 2018; Emmerling et al., 2019; Heal, 2017; Hof et al., 2008; Nordhaus, 2014; Stern, 2008).

In the DICE model, δ and η are determined on the basis of the Ramsey rule $\rho = \delta + \eta g$, where ρ is the social discount rate and g the growth rate of consumption[10] (Drupp et al., 2018). Thereby, for every point in time, the social discount rate ρ describes the total rate at which consumption is discounted. DICE calibrates ρ such that it corresponds to real-world interest rates (Nordhaus, 2011). In practice, this means that ρ is set to be equal to an observed market interest rate – the determination of which is not obvious (Frisch, 2013; Giglio et al., 2015; Kowarsch, 2016a) – and subsequently the two parameters δ and η are chosen such that the Ramsey rule holds. Resulting from this are the DICE parameter values $\delta = 1.5\%$ and $\eta = 1.45$, which in conjunction with growth rate assumptions lead to a social discount rate of around 4.25% per year[11] (Nordhaus, 2018b). The approach of choosing δ and η such that the social discount rate reflects observed rates of return on capital, is often referred to as 'descriptive', in contrast to an alternative, 'prescriptive' approach that views these modelling choices as normative parameters requiring ethical justifications (Arrow et al., 1996; Azar, 1998). This terminology goes back to the second

[8] The parameter η is also referred to as 'intertemporal elasticity of substitution' (Heal, 2017).

[9] In contrast to DICE, the regionalised model RICE is able to account for differences between world regions. Nevertheless, RICE treats spatial inequality differently to intertemporal inequality, such that it effectively exhibits no spatial inequality aversion. A further interpretation of η is risk aversion, i.e. inequality aversion with respects to different possible states of an uncertain world. DICE, being deterministic does not include this (Nordhaus, 2013). However, there are extensions of DICE that separate between risk aversion and temporal inequality aversion (Ackerman et al., 2013).

[10] The Ramsey rule in this form is only valid in a deterministic setting (Drupp et al., 2018), and in situations where "the project under evaluation is marginal to the path of future consumption" (Dietz et al., 2008, p. 13). Nonetheless, it is often used in climate economics, e.g. in DICE.

[11] Due to variable growth projections, the social discount rates is also non-constant. The 4.25% figure refers to the average rate throughout this century. As economic growth is projected to fall over time (Christensen et al., 2018), the social discount rate is even higher in the near future.

assessment report of the IPCC (Bruce et al., 1996) and is widely used (e.g. Beck & Krueger, 2016; Kelleher, 2017; Nordhaus, 2013; Weyant, 2014), despite large problems with separating descriptive from prescriptive aspects in time discounting (Heilmann, 2017). Proponents of the descriptive approach claim that it results in a more realistic representation of market realities, because parameters are chosen such that the model "generate[s] savings rates and rates of return on capital that are consistent with observations" (Nordhaus & Sztorc, 2013, p. 37). The proponents of the prescriptive approach, on the other hand, argue that a descriptive approach would only make sense in a marginal context, where considered impacts do not have an effect on future consumption (Dietz et al., 2008) and in a situation where the distribution of consumption losses can be disregarded (Frisch, 2018; Sterner & Persson, 2008)[12]. In the context of climate change, where decisions have to be taken by society as a whole, Stern (2008) argues that the parameters δ and η should be chosen based on ethical considerations.

These two approaches to discounting reveal two very different interpretations of the whole modelling endeavour of DICE. The descriptive interpretation, as propagated by William Nordhaus, sees IAMs as "a description of how economies and real-world decision makers (consumers, firms, and governments) actually behave" (Nordhaus, 2013, p. 1110). This account is based on the notion that optimisation can compute the equilibrium of a market economy, provided there are no externalities[13]. Consequently, the descriptive interpretation of DICE sees the maximisation of the social welfare function as "an algorithm for finding the outcome of efficient competitive markets" (Nordhaus & Sztorc, 2013, p. 1111), which does not necessarily have compelling normative properties. Instead, the welfare maximisation of DICE is interpreted as a projection of how countries with the same preferences as today's countries would act if the GHG externality was fully priced and thereby internalised (Kelleher, 2019).

The prescriptive interpretation, on the other hand, sees IAMs as tools for ranking different pathways according to the criterion of maximal welfare (Kelleher, 2022). Importantly, this approach is mathematically identical – it uses the same social welfare function. Formally, the proponents of the prescriptive approach to discounting only choose the two determining parameters of the social welfare function differently, based on strictly normative considerations (Dietz et al., 2008; Stern, 2008). However, beyond choosing different parameters, they have a fundamentally different view of the modelling endeavour as such. According to the prescriptive interpretation, the DICE model is not a tool for computing market equilibria. Instead, optimising the social welfare function is seen as a way of normatively comparing and ranking alternative policy pathways (Kelleher, 2022).

Following the framework of Mäki (see Figure 3), it is the modeller's job to provide an informative model commentary that includes guidance on how to interpret the model. In light of the two different interpretations of DICE, Nordhaus himself is not always consistent in his model commentary. In the most recent model documentation, he describes the social welfare function as "ranking different paths of consumption" and being "affected by two central normative parameters" δ and η (Nordhaus & Sztorc, 2013, p. 6). This resonates strongly with the prescriptive interpretation of DICE. On the other hand, he insists that the baseline scenario should be an attempt at projecting economic and environmental variables from a positive[14] perspective (p. 8). Taken together, the Nordhaus interpretation of DICE seems to consist of descriptively

[12] A further argument states that investments in climate change mitigation can only sensibly be compared to very long-term assets, for which there is little conclusive evidence on interest rates (Giglio et al., 2015; Kowarsch, 2016a). However, these rates are most likely considerably lower than those assumed by DICE.

[13] An externality is an unintended impact of an economic agent's decision on another agent's utility or profit (Perman et al., 2003, p. 134). In the case of climate change, every emission of GHGs reduces the utility of people that (will) suffer from global warming and it is therefore called a negative externality.

[14] For simplicity, I treat 'positive' and 'descriptive', as well as 'normative' and 'prescriptive' synonymously, respectively.

projecting a baseline scenario and then prescriptively ranking different pathways with climate policy (Kelleher, 2022). While this might be considered a minor detail within the model documentation, it has very real consequences for the interpretation of model output, and thereby for the climate policy debate. Nordhaus, in his communication of DICE assumptions and result, imposes prescriptive interpretations on a descriptive modelling exercise – which leads Kelleher (2019) to conclude that it is "very unclear why he feels entitled to assign the label 'optimal' to the outcome of his quasi-predictive exercise and to associate that outcome with 'idealized' and 'economically desirable' policy" (p. 101).

3.4 Expectations for DICE

Evaluating DICE has revealed several possible expectations for it. The first expectation is linked to the purpose of calculating costs and benefits of climate policy pathways. An associated evaluation criterion would be sufficient forecasting skill to accurately project mitigation costs and avoided climate damages. This expectation is linked to model users aiming to obtain quantitative policy guidance from DICE. These users could be policymakers, or institutions like the Interagency Working Group (IWG) of the United States, which estimates the SCC to be used by all government bodies (Metcalf & Stock, 2017). Implicitly, Pindyck (2017) voices this expectation when claiming that IAMs "have no empirical (or even theoretical) grounding and thus [...] cannot be used to provide any kind of reliable quantitative policy guidance" (p. 103). Pindyck concludes that "economists should not claim that IAMs can forecast climate change and its impact or that IAMs can tell us the magnitude of the SCC" (p. 112). These quotes reveal the expectation that an IAM should provide quantitative guidance on the SCC and accurate forecasts to an audience of policymakers, while being empirically (and theoretically) grounded. For several reasons, DICE can not fully live up to this. First of all, forecasting and quantitative accuracy do not form part of its stated goals (Nordhaus, 2013). Consequently, DICE prioritises tractability and transparency over detail (Nordhaus, 2011). Secondly, the target system encompasses the world economy, the climate system and climate decision-making – all of it on a centennial timescale. Due to the complex nature of the target and associated uncertainties, accurate forecasting seems hardly achievable in principle.

The second expectation is linked to the purpose of learning about the behaviour of the model and its target system, for example by calculating impacts of alternative assumptions. An associated evaluation evaluation could be the performance of the model as an epistemic tool – how easily can it be manipulated and what aspects of the model facilitate learning about its behaviour? (Knuuttila, 2011) Another criterion – combining credible worlds in the sense of Sugden (2000) with stories in the sense of Morgan (2001) – could be the credibility of stories being told through model use. This second expectation of DICE could be raised by model users who want to understand qualitative dynamics around climate policy or investigate the importance of different mechanisms or issues. Users could for example be academics with an epistemic interest in the model results. However, they could also be political institutions or thinktanks who use the model to highlight the relevance of a certain issue linked to their political agenda (e.g. the WWF: Johnson et al., 2020). Often, this expectation is voiced in terms of 'providing insight' (Bauer et al., 2020; Botzen & van den Bergh, 2014; DeCanio, 2005; Leimbach, Bauer, Baumstark, Lueken et al., 2010). While this phrasing is a useful term as a contrast to 'providing numbers' or accurate forecasts, it is too vague to be easily translated into an evaluation criterion. As Huntington et al. (1982) state, "when models become viewed as tools more for developing insights than for forecasting numbers, [...] an assessment of how the models are used is as important as understanding their structures" (p. 450). Yet, criteria for evaluating model use are challenging to derive. In the analysis of DICE, I have focused on the concrete manipulability in the sense of an epistemic tool, on the type of stories that the model can tell, and on the credibility of the possible worlds it depicts. I was able to show that DICE does indeed provide a modelling framework that works well as an epistemic tool, through its tractability and open-source code. It is able to tell a range of different stories about the world, but the interpretation of them is often ambiguous. Finally, for a modelled counterfactual world to be credible, there has to be some significant similarity between the model world and the real world (Sugden, 2000). Whether the similarity can be judged significant will obviously depend on the respective application, such that it can not be evaluated for the DICE model as such. I note, though, that through this aspect of similarity, questions of realisticness enter the picture again – claiming that a model is used ‘only’ for generating insights does not entirely liberate it from concerns about its fit to the real world.

Finally, there is an expectation that is linked to the evaluation criterion of transparency. For modellers, this “serves the purpose of mitigating misinterpretations or errors by users of model results” (Schwanitz, 2013, p. 125), whereas users may be “more concerned about high-level model design, key technical assumptions, and critical modeling practices” (Skea et al., 2021, p. 3). DICE satisfies basic requirements of transparency, like the availability of model code and documentation. Increasingly, though, there are calls for what Bistline et al. (2021) coined *deep transparency* – which describes the communication of potentially value-laden assumptions and modelling choices. When using IAMs in the context of policy advice, modellers should thus provide model commentaries which include these critical assumptions as well as coherent interpretations about the way in which model results connect to the real world. Concerning the communication of critical assumptions, DICE is somewhat ambiguous – the model documentations include discussions of discounting and damage functions, but not of inequality (Dennig et al., 2015) or substitutability between material wealth and natural capital (Bastien-Olvera & Moore, 2021; Neumayer, 1999; Nordhaus & Sztorc, 2013). With respect to the interpretation of model results, my analysis has revealed two contrasting perspectives on DICE’s decision structure – the descriptive and the prescriptive interpretation. Each one is internally consistent, but it demands different parameter choices for the social welfare function and fundamentally different conceptions of the whole modelling endeavour. I found that DICE, in its standard formulation, does not separate clearly between the two interpretations. As a result, it paints the image of a descriptive model with a prescriptive application, which makes its model commentary very ambiguous.

What can be learned from the analysis of DICE along different evaluation criteria and emerging expectations? First, the analysis of how DICE represents its target system highlights the huge challenges of modelling in the context of climate policy, while also shedding light on several questionable modelling choices embodied in DICE. Second, an analysis of how DICE is used shows how its tractability and manipulability enable a broad discourse about modelling assumptions and alternative formulations, such that DICE in its various forms can be used to tell many different stories about global climate change (Beck, 2018). Third, I was able to show that the model commentary about DICE is often vague and ambiguous. Partly, this is due to discussing assumptions selectively, and partly through inconsistencies in the interpretation of modelling result. Overall, DICE can be described as a model with limitations in its representation of the target system, which can nevertheless be useful to the academic and policy discussion through its easy manipulability, provided that modellers are consistent and comprehensive in their model commentary.

4 REMIND

The REgional Model of INvestments and Development (REMIND) has been introduced by Leimbach, Bauer, Baumstark and Edenhofer (2010). It is based on the single-region model MIND (Edenhofer et al., 2005), which places special emphasis on the modelling of endogenous technological change in energy technologies. REMIND has been used in the context of the IPCC (IPCC, 2014) and it is one of the five IAMs that are used to quantify an SSP marker scenario (Kriegler et al., 2017). I base the following evaluation on REMIND 2.1 as documented in (Baumstark et al., 2021), unless otherwise specified. Details on older model versions are primarily referring to REMIND 1.6 (Luderer et al., 2015).

4.1 Target System of REMIND

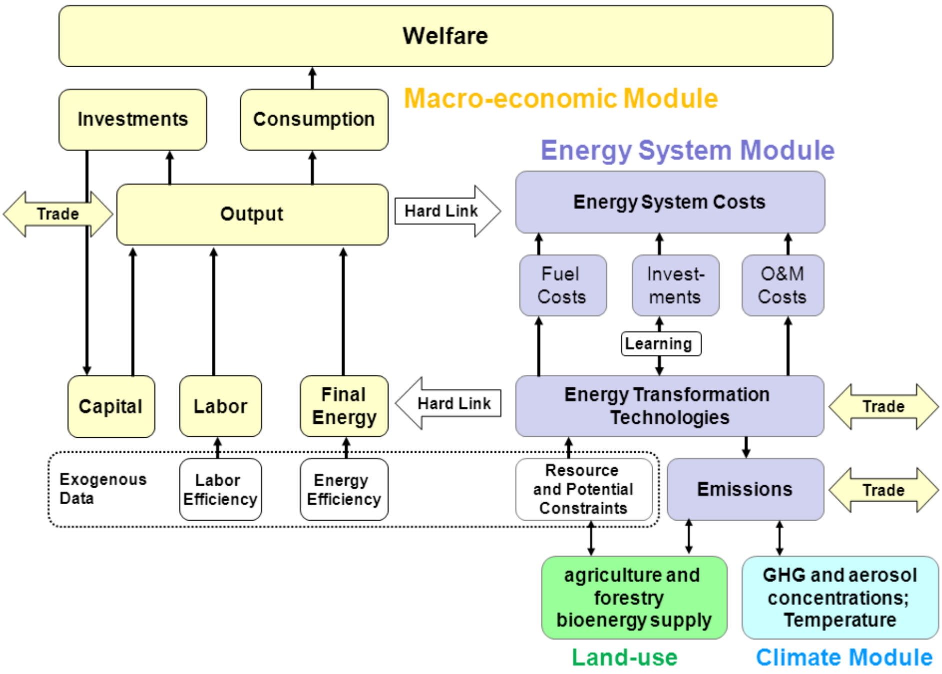
REMIND is a hybrid model, combining a top-down macroeconomic model with a bottom-up energy system model (Baumstark et al., 2021; Leimbach, Bauer, Baumstark & Edenhofer, 2010). In addition to these macroeconomic and an energy system modules, it includes a land-use module and a climate module[1] (see Figure 6).

By placing emphasis on the energy system, REMIND has a detailed representation of the biggest anthropogenic GHG sources. Together with its disaggregated macroeconomic system representation, it is thereby able to model dynamics of the "global energy-economy-emissions system", with a strong focus on climate change mitigation options (Baumstark et al., 2021, p. 6571). In order to evaluate REMIND, an obvious starting point is thus to analyse how well it represents the energy-economy-emissions system.

As the successor of the MIND model (Edenhofer et al., 2005), REMIND explicitly models technological learning and prides itself with an especially high resolution of different energy carriers and conversion technologies (Leimbach, Bauer, Baumstark & Edenhofer, 2010). Its energy system module is based on cost optimisation, such that the model chooses the energy technologies that meet macroeconomically prescribed energy demand at least total cost. Naturally, energy prices for different technologies play a crucial role in this type of model. REMIND calibrates each region's energy system parameters to data by the International Energy Agency (IEA), and generates future energy costs based on technology-specific assumptions (Luderer et al., 2015). For fossil fuels, the model assumes rising extraction costs as low-cost deposits get depleted. For renewables, on the other hand, learning rates are implemented, such that wind and photovoltaic (PV) investment costs decrease with 12% and 20% respectively for every doubling of total installed capacity (Krey et al., 2019). Therefore, the more renewable energy facilities are deployed, the cheaper they get – a dynamic that leads to nonlinear effects and path dependencies which are missing in many other IAMs.

Nonetheless, REMIND has repeatedly been found to project too conservative developments of

[1] While REMIND features a simple climate module and a simple land use module of its own, it is mostly run in conjunction with MAgPIE (Dietrich et al., 2020) for agriculture and land use modelling and MAGICC6 (Meinshausen et al., 2011) for climate modelling. The simple modules within REMIND are calibrated to emulate those two models, respectively.



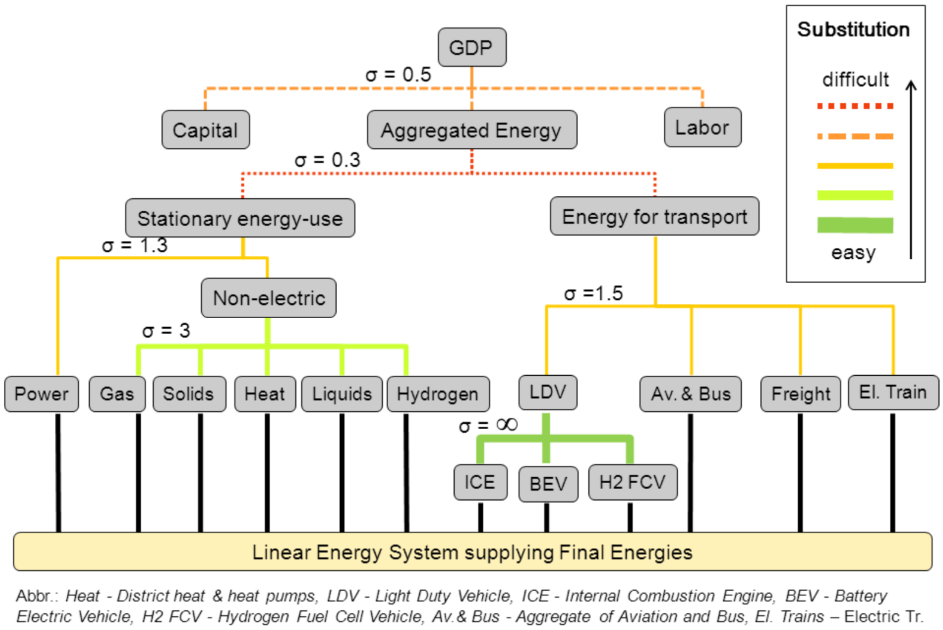
Source: Luderer et al. (2015, p. 4)

Figure 6: Structure of REMIND. The macroeconomic and the energy system module are hard-linked, such that the macroeconomic module determines energy demand, which is covered by the energy system module and whose costs are fed back to the macroeconomic module. Trade can occur in general goods, energy carriers or emission permits.

renewable energies, and especially PV installations (Creutzig et al., 2017; Ives et al., 2021; Wilson et al., 2013). While the assumed learning rate of PV of 20% was actually close to the historically realised rate of 22.5% (Creutzig et al., 2017), modellers typically also assume floor costs, i.e. prices at which the cost decline levels off. These floor costs for PV have repeatedly been overtaken by reality throughout the last decade, such that they had to be corrected downwards (Ives et al., 2021). Ives and colleagues blame the complexity and size of DP-IAMs for this misjudgement, claiming that these models have been disaggregated so much that it is hard to understand which assumption is driving the results, thereby making the models less transparent and harder to evaluate. The difficulty of making appropriate technological assumptions is further illustrated by Luderer et al. (2012), where REMIND projects the future transport energy mix as consisting to a large extent of liquefied coal in combination with carbon capture and storage (CCS), while completely disregarding electric vehicles – modelling choices that already ten years later seem incomprehensible. To the credit of REMIND modellers, though, the model's assumptions are frequently updated, such that new developments in PV technology and electric vehicles have swiftly been implemented into newer model versions (Creutzig et al., 2017; Luderer et al., 2015).

A different challenge to REMIND's ability to represent the real-world energy system lies in the critique of Trutnevyte (2016), which finds that energy models based on cost optimisation perform poorly at reproducing historical patterns of energy transition. To make the cost-optimised scenarios more realistic, REMIND introduces cost-markups for fast scaling-up of investment, for storage and for grid integration, among other adjustments (Baumstark et al., 2021; Luderer et al., 2015). It remains unclear, however, to which extent these modifications contribute to REMIND's ability of representing future dynamics in the energy system.

The macroeconomic structure of REMIND is, similarly to DICE, based on a neoclassical



Source: Luderer et al. (2015, p. 12, original quality)

Figure 7: Production structure of REMIND: Different production factors and energy carriers, coupled at each level with constant elasticities of substitution (CES) σ, as indicated by the colour scheme.

growth model. A major difference to DICE, however, is that REMIND is a regionalised model, such that it has production functions for each of the eleven regions and incorporates trade flows between those (Leimbach, Bauer, Baumstark, Lueken et al., 2010). The production functions are based on the three partly substitutable factors capital, labour and energy. The addition of energy as a production factor distinguishes REMIND from DICE, as does the choice of the more general, nested structure with constant elasticities of substitution (CES). This means that the production factor energy branches into its component parts with different elasticities of substitution at different levels of branching, all the way down to the level of final energy carriers (see Figure 7). This structure has been criticised by Kaya et al. (2017) because of failing to match historically observed patterns. The authors argue that, in CES modelling, elasticities have to be assumed without much empirical basis. At the same time, these choices are very influential on final modelling results, as evidenced by Truong (2009), who finds that a formulation with non-constant elasticities of substitution can reduce modelled mitigation costs by about half. Further, Kaya et al. (2017) find that CES functions tend to "propagate the status quo of energy shares" (p. 8). For example, they criticise that REMIND artificially constrains substitution of non-electric stationary energy with power-based alternatives (third branching in Figure 7), which ultimately biases the model towards bioenergy.

In each region, resulting economic output is distributed on consumption, investment into the capital stock or the energy system, or on trade (Luderer et al., 2015). In this sense, the macroeconomic structure of REMIND is very similar to the DICE model. Thereby, doubts about the appropriateness of using an economic growth model for climate policy, as raised by Millner and McDermott (2016), equally apply. REMIND circumvents the problem of having to rely on projections from its own growth model by calibrating its efficiency parameters such that it aligns with exogenous scenarios, such as OECD's[2] economic growth projections (Dellink

[2] OECD stands for 'Organisation for Economic Co-operation and Development' and it represents a group of 38 high-income countries.

et al., 2017; Luderer et al., 2015).

Overall, REMIND's energy system and macroeconomic modules aim to provide a detailed and disaggregated representation of the energy-economy-emissions system. However, the analysis revealed that the disaggregation leads to many free parameters that require techno-economic assumptions – each associated with its own uncertainties. Further, economic and technological assumptions also interact. Assumed elasticities of substitution, for example, play an influential role in determining which energy technology receives how much investment, which in turn alters energy costs. Yet, despite the strong dependence of its result on specific modelling choices, REMIND has the advantage of representing a comprehensive range of different energy technologies, which, in conjunction with the separation into regions, makes it potentially very useful for modelling mitigation pathways.

4.2 Using REMIND

On its own website, REMIND is described by two paragraphs of text. The first one starts by introducing REMIND as "a numerical model that represents the future evolution of the world economies" (PIK, 2022) – this aspect of representing its target system has been investigated in the previous section. The second paragraph describes REMIND as a model that "aims to help policy and other decision makers to plan ahead" – analysing how this is achieved will be the purpose of this section.

As a fundamental prerequisite for supporting long-term planning, Baumstark et al. (2021) repeatedly stress the ability of "investigating internally consistent transformation pathways" (pp. 6572, 6579). This internal consistency of the pathways modelled by REMIND is guaranteed by the fact that it is a fully coupled model – modelling the economy in general rather than partial equilibrium. That means that all subsystems develop in constant interaction, guaranteeing that one single solution of the optimisation problem is found across all economic sectors (Bauer et al., 2008; Nikas et al., 2019).

The internally consistent transformation pathways are subsequently compared to each other, in order to explore synergies and trade-offs. Before discussing this comparison of pathways, however, it is useful to take a closer look at how they are generated, and how this is shaped by REMIND's model structure. The macroeconomic module of the original REMIND model had no representation of economic damages from climate change (Luderer et al., 2015). Thereby, the only way in which GDP is reduced is through investment into climate mitigation – which means that REMIND can not be used to conduct cost-benefit analyses[3]. Instead, REMIND is mostly used for cost-effectiveness analysis. This means that global temperature change is exogenously constrained through the introduction of a climate target, and the model performs a constrained optimisation for calculating the most cost-efficient pathways that reaches the target. A typical research question would, for example, investigate the techno-economic feasibility of reaching the 2 degree target and identify the cheapest way of doing so (Baumstark et al., 2021).

This leads to two types of pathways: the baseline path which is derived by unconstrained optimisation and policy pathways derived from constrained optimisation. Each of these pathways can be understood as a credible world in the sense of Sugden (2000). By assuming a situation without climate policy (or damages), the baseline path is certainly not realistic, but it aims to depict a credible, counterfactual world – like a thought experiment asking what would happen in a fantasy world without climate change. Policy pathways can similarly be understood to

[3] Since 2021, REMIND allows for the inclusion of a damage function (Schultes et al., 2021). I will talk about this below and, for now, describe the way in which REMIND used to and still mostly does operate.

model credible worlds under certain assumptions about technology, socioeconomic development, international cooperation or political ambition. In contrast to the baseline path, there is a large variety of different policy pathways, each exploring the consequences of a certain set of assumptions[4]. Thereby, every policy pathway can also be understood as a story told through REMIND. As Morgan (2001) claims that "we only fully understand our model when we have identified all the specific stories that it can encompass or tell about the world" (p. 380), REMIND could be evaluated as to what kind of policy pathways it is able to and actually does model – and which stories it is not able to tell.

By comparing these two types of pathways, or types of stories told by REMIND, policy and technology options for climate mitigation can be explored. This comparison can be used to derive the total mitigation costs of achieving a certain target, or for assessing the scale of required investments in specific mitigation technologies. However, the comparison of policy pathways against no-policy baselines has been subject to criticism (Burgess et al., 2020; Grant et al., 2020; Stern, 2016). After all, baseline scenarios which project economic dynamics in a world without mitigation also operate in a completely unrealistic world without damages from ongoing climate change[5]. Therefore, the baseline scenario will always appear cheaper, as it is subject to neither mitigation costs nor damage costs. As a consequence, cost-effectiveness analysis by design excludes the possibility that rapid decarbonisation might overall be cheaper than continuing present-day trends, as suggested by some recent papers (Grubb et al., 2021; Kelleher, 2019; Way et al., 2021) – this being a story that REMIND is not able to tell.

A second problem that arises from the omission of climate damages concerns the modelled policy pathways themselves. While being constrained by a temperature target, they also operate in a counterfactual world without climate change. However, it is not hard to imagine how ongoing climate change could affect mitigation, e.g. economically through altered investment patterns or biophysically by reducing bioenergy options. To account for these influences, REMIND has recently complemented its modelling framework with a damage module, which is based on regional climate indicators and is able to incorporate several damage functions from the literature (Schultes et al., 2021). The inclusion of damage costs into the calculation of mitigation pathways, they find, roughly doubles the resulting constrained-'optimal' carbon price. Before the introduction of its damage module, REMIND was thus restricted to telling stories without any climate damage, both for the baseline and the policy pathways. This illustrates the importance of model structure in determining both the questions that can be asked and the answers that can be obtained through the model.

Through intertemporal optimisation, REMIND implicitly assumes that the 'social planner' who is allocating resources has perfect knowledge about all future developments, including costs of and substitution possibilities between technologies. It also assumes efficient markets – no unemployment, no overproduction, no underinvestment – and other highly idealising assumptions (Staub-Kaminski et al., 2014). Most of these idealisations are not actually fulfilled in reality (Sanstad & Greening, 1998). Therefore, results of these model runs are often considered first-best scenarios that illustrate a theoretical best-case (Baumstark et al., 2021). In this terminology, second-best scenarios are those that include different types of 'imperfections', aiming to be a better representation of reality. As the model documentation of REMIND puts it, its central strength is "its ability to calculate first-best mitigation strategies that provide benchmark development pathways against which mitigation scenarios under sub-optimal settings can

[4] The SSP scenarios consist of five different baseline pathways (Riahi et al., 2017); thereby REMIND can model more than a single baseline scenario. Generally, however, there is always several policy pathways for each baseline scenario.

[5] In the words of the modellers: "In the reference scenario [...] we simulate a development as if climate change has no economically and socially important effects" (Leimbach, Bauer, Baumstark & Edenhofer, 2010, p. 161). Since Schultes et al. (2021), REMIND is capable of modelling a development with climate impacts.

be compared" (Luderer et al., 2015, p. 35).

Besides enhancing tractability, these idealisations also fulfil specific epistemic functions. By devising idealised benchmark scenarios first and subsequently introducing limitations (such as the unavailability of a technology, lack of international cooperation, non-functioning markets, ...), the relative effect of idealising assumptions can be estimated (Bauer et al., 2012; Leimbach & Bauer, 2021). Assuming perfect foresight is, while completely unrealistic, helpful for a policy advice tool because it guarantees that long-term consequences get appropriate weight in the model's decision mechanism (Luderer et al., 2012). REMIND, for example, places special emphasis on dynamics of technological learning, which will make renewable energy very cheap once it has received sufficient investment (Creutzig et al., 2017). If the model's decision mechanism was more myopic, these projected long-term dynamics would not be fully accounted for – leading to less cost-efficient pathways. As a consequence of these idealisations, REMIND is a useful epistemic tool that is well-positioned to explore effects of different assumptions. Conversely, however, REMIND modellers also acknowledge that the idealisations introduce a "distinct normative component" (Baumstark et al., 2021, p. 6593). That is, modelling how a perfectly knowledgeable 'social planner' would allocate resources can not serve as a descriptive projection of future developments, but is instead seen as a normative benchmark scenario that is preferable to scenarios with less idealised assumptions.

4.3 Transparency of REMIND

To some extent, normative elements are unavoidable when modelling climate change – a global issue that is deeply connected to social and ethical questions (Stern, 2008; van der Sluijs, 2002). It is against this background that Bistline et al. (2021) call for IAMs to proactively make the consequences of value-laden modelling choices explicit. Ellenbeck and Lilliestam (2019) emphasise that IAMs – whether they like it or not – are also powerful political tools with strong influence on the climate policy discourse. Consequently, they encourage integrated assessment modellers to document not only *what* was done, but also *why*, such that justifications and theoretical backgrounds of modelling choices come to light. Having analysed the nuanced ways in which REMIND acts as an epistemic tool, by combining descriptive representations with strategic idealisations, I will now evaluate how well it fares with respect to the documentation and communication of its critical assumptions.

Before getting to the transparency of critical assumptions in REMIND, some background on its decision structure is useful. In REMIND, all regions have their own social welfare functions, which are aggregated in a weighted sum to yield global welfare – the quantity that the model maximises. As in DICE, social welfare functions in REMIND have the parameters δ for the rate of pure time preference and η for inequality aversion (Luderer et al., 2015). The model comes with two different algorithms for determining this aggregation – and at the same time global trade patterns: the Negishi approach and the Nash approach. Both fulfil the constraint of having market-clearing equilibrium prices for all traded goods and a so-called intertemporal budget constraint. The latter is imposed to make sure that, over the whole time span, balances of payment for each region are zero. The Nash algorithm calculates global equilibrium prices by maximising regional welfare for each region separately. It is therefore considered a non-cooperative solution (Luderer et al., 2015). The Negishi approach first imposes global equilibrium prices and then adjusts a set of welfare weights such that each region fulfils its intertemporal budget constraint. This is considered a cooperative solution, as the welfare maximisation is performed from a global point of view. The Negishi weights have the effect that utility is valued more for some regions than for others, in order to avoid redistribution of wealth from richer to poorer regions. REMIND has an inequality aversion parameter of $\eta = 1$, such that an algorithm tasked with maximising global welfare will allocate more future consumption to poorer regions, thereby leading to redistribution. Negishi weights and the regional budget constraints are means of ensuring that this redistribution does not occur in the model – essentially a mechanism of cementing current global inequalities (Stanton, 2011). REMIND modellers justify this method with its achievement of Pareto-optimality (Leimbach, Bauer, Baumstark, Lueken et al., 2010), i.e. situations where nobody could be made better-off without making someone else worse-off. However, the Pareto-optimal outcome is blind to preexisting inequalities and therefore readily accepts to perpetuate them (Nelson, 2008) – against the intuitions of inequality aversion as encoded in the utility function[6].

In a comparison of different DP-IAMs regarding the transparency of assumptions on bioenergy with carbon capture and storage (BECCS), Butnar et al. (2020, p. 11) come to the conclusion that REMIND performs well at documenting "wider system settings such as general discount rate, carbon pricing regime, or availability of other NETs [negative emission technologies]." On the other hand, they find it "difficult to separate transparency from completeness (i.e. what the IAMs do not include or is implicit)" (p. 11). Indeed, the model documentation of REMIND barely mentions excluded factors (Baumstark et al., 2021) – an omission also noted by Ellenbeck and Lilliestam (2019). Sensitivity analyses are partly referenced to in REMIND's model description. For example, Giannousakis et al. (2021) analyse the effect of different technological assumptions on model output. However, the sensitivity of models to non-technological assumptions, such as the discount rate, is not reported. While the model documentation described mechanics and structure of REMIND in clear and comprehensible language, there is not much discussion of robustness, uncertainties, omissions, or ethical and political implications. This can be exemplified through the examples of discounting, learning rates and distributive aspects.

The documentation of REMIND 1.6 states that the model assumes a rate of pure time preference and an inequality aversion of 3% and 1, respectively. The justification is that these values yield a social discount rate of 5-6%, which is taken to be in line with interest rates observed on capital markets (Luderer et al., 2015). By this reasoning, REMIND endorses the so-called descriptive approach to discounting, even though the justification of this choice and its ethical implications are not elaborated on. In the newest documentation of REMIND 2.1, information on discounting is even more sparse. There are no numerical values for crucial parameters and no indication of the reasoning for choosing them is given. Given that Emmerling et al. (2019) show substantial impacts of changes to the social discount rate on the timing of emission reductions and on the amount of negative emission technologies, this is a serious omission.

Learning-by-doing effects, modelled through learning rates and floor costs, have been shown to have a significant impact on mitigation scenarios (Creutzig et al., 2017; Edenhofer et al., 2005), and both the reasoning and the associated parameters have been outlined in Luderer et al. (2015). Surprisingly, the newest model documentation only mentions learning-by-doing effects, but does not elaborate on either the parameters, the empirical foundation or the influence of these effects on model output. Given that learning-by-doing effects recently led Way et al. (2021) to suggest that a fast energy transition might be cheaper than a slower one, more context on these technology assumptions would be desirable.

Aspects of distribution play no role in REMIND's model descriptions. Neither the implications of omitting intraregional inequalities by choosing a single representative agent for each

[6] There are arguments for separating intertemporal and spatial η into two distinct values – mainly because of evidence that many people exhibit different implicit aversions to different dimensions of inequality (Dietz et al., 2008; Hepburn & Beckerman, 2007).

region (Budolfson et al., 2017), nor the interregional distributive effects encoded in Negishi weights (Stanton, 2011) are discussed. Given the increased calls to deepen transparency about ethically and politically sensitive modelling assumptions (Bistline et al., 2021; Kowarsch, 2016b; Skea et al., 2021), it is hard to understand why the REMIND documentation places so little emphasis on these topics.

What the authors of REMIND do place great emphasis on, is the distinction between predictive and exploratory modelling. On several occasions in the document they stress that "these self-consistent scenarios are not to be understood as forecasts but rather as projections that depend on a broad set of assumptions" (Baumstark et al., 2021, p. 6579). Through these qualifications about what can reasonably be expected of the model, the authors ensure transparency and avoid misunderstandings. All the more, though, it is difficult to understand why the "specific assumptions" that future projections are conditional on, get so little attention in REMIND's model description.

4.4 Expectations for REMIND

Expectations for REMIND can be clustered into three ideal types. The first one, as voiced on its own website, is linked to REMIND providing a representation of the world economy, with a focus on energy and climate aspects (PIK, 2022). This expectation entails a descriptive modelling strategy, whereby long-term developments and trade dynamics are supposed to approximate real-world dynamics. Evaluation criteria would therefore be measures of fit or similarity between REMIND and future macroeconomic behaviour. Users holding this expectation could have interests in the economic aspects of climate mitigation and reliable and realistic cost projections, for example energy companies or government agencies. For the energy sector, REMIND does indeed provide a detailed representation of more than 50 technologies (Luderer et al., 2015). Similarly, it is able to provide a detailed representation of climate and land-use dynamics when coupled to MAGICC6 and MAgPIE. Nevertheless, projections have shown to been very sensitive to assumptions such as learning rates or elasticities of substitutions. Consequently, REMIND could be considered a realistic representation of energy, land use and climate systems only in the sense that it includes most relevant elements and processes, but not in the sense that it is able to reliably project those into the future. REMIND's macroeconomic representation is more idealised – it includes eleven regions and trade among those, but no further heterogeneities in the form of income groups per region or economic sectors. Instead, REMIND relies heavily on exogenous pathways for socioeconomic parameters, mainly from SSP scenarios (Soergel et al., 2021). Thereby, the model can indirectly be used to study future developments in the world economy, but its own dynamic representation of socioeconomic parameters is limited.

A second expectation is linked to REMIND providing policymakers with insights on the synergies and trade-offs between different mitigation pathways, in order to facilitate long-term planning. Naturally, this expectation is often held by actors from the policy process. Yet, also people who are academically involved in the topic of climate mitigation could have an interest to develop further understanding of the qualitative system behaviours around climate mitigation. Evaluation criteria linked to this expectation are connected to REMIND's ability to analyse interactions of different factors and assumptions in a systematic manner. Emphasis could be placed on consistency, tractability, or flexibility of the model to be adapted for different kinds of analyses. Due to REMIND's high resolution of technologies, it is capable of projecting a broad variety of different mitigation pathways. It can generate alternative scenarios either by running with different exogenous assumptions (e.g. on economic growth or population; Kriegler et al., 2017) or by introducing targeted changes to certain model components, such as allowing for trade imbalances (Leimbach & Bauer, 2021), restricting the use of certain technologies (Bauer et al., 2012), or restricting the spread of regional carbon prices while allowing for financial transfers (Bauer et al., 2020). Thereby, REMIND can generate a large variety of internally consistent pathways and serve as an epistemic tool. However, in contrast to DICE, the model is so complex that it is mainly used by its developers and not easily adapted by other researchers. Further, the quantification of synergies and trade-offs for long-term planning is a challenge to REMIND, as its policy-relevant outputs are sensitive to technological assumptions (Creutzig et al., 2017), assumptions about negative emissions (Giannousakis et al., 2021) and the discount rate (Emmerling et al., 2019) – it is thus restricted to providing qualitative insights into mitigation pathways.

Lastly, there is an expectation of REMIND which stems from its importance for the climate policy process and its discursive power (Cointe et al., 2019; Ellenbeck & Lilliestam, 2019), and concerns the model commentary. This expectation could be held by model users from the policy process, by other researchers from within the IPCC process, or by the general public and media. On behalf of modellers, this expectation is linked to the purpose of mitigating misinterpretations and making the model more accessible to the public (Schwanitz, 2013). It entails efforts towards being transparent about REMIND's underlying assumptions, the paradigms ingrained in the modelling framework, and the possible consequences of alternative modelling choices. Associated evaluation criteria are transparency and value diversity (Kowarsch, 2016b). While REMIND provides extensive model documentation, detailed justifications for modelling choices are sparse, and critical assumptions are often neither highlighted nor compared to alternatives. Particular room for improvement of REMIND's model commentary lies in the further delineation between descriptive and normative elements. The model clearly states its descriptive ambition, and notes the distinctly normative elements to scenario projections. However, REMIND modellers do not provide much guidance on where the boundary between those two domains lies – whereby many model elements are left ambiguously underspecified with respect to their purpose.

The evaluation of REMIND along different criteria and emerging expectations revealed two interesting characteristics of the model. First, questions of representation and criteria regarding model use are tightly coupled in the case of REMIND. The model is used to produce and compare different 'credible worlds'. In order to provide valuable insights into possible mitigation pathways, it thus needs both a certain degree of realisticness and a certain degree of flexibility. These two aspects can complement each other; the disaggregated and detailed energy system module of REMIND, for example, can make the model both more realistic and more suitable to model various policy pathways. However, the two aspects can also interfere with each other; idealisations like a 'social planner' with perfect foresight contribute to REMIND's suitability as an epistemic tool for producing policy-relevant output, but they do so at the expense of a more realistic representation of macroeconomic dynamics. Second, the model commentary about REMIND is often ambiguous and fragmentary. Its model description does not provide much insight into omitted factors, alternative modelling choices or implications of critical assumptions. Thereby, it also fails to clarify potential conflicts between the above-mentioned aims of descriptively representing a target system while being flexible enough to model a large variety of pathways. Overall, REMIND has the potential of both modelling mitigation pathways in great detail and assessing alternative policy options, but it is fraught with uncertainties in its representation of technology and the economy, which are exacerbated by a lack of clarity and guidance on how to interpret model results.

5 IMAGE

The Integrated Model to Assess the Greenhouse Effect (IMAGE) was originally introduced by Rotmans (1990). Over the course of its history, its scope has broadened to include more environmental problems, and accordingly the name was changed to 'Integrated Model to Assess the Global Environment' (Stehfest et al., 2014). IMAGE has been frequently used on the context of the IPCC (Bruce et al., 1996; IPCC, 2014), but also in assessments of other international bodies such as the UN Environment Programme, the OECD, or the European Commission (Stehfest et al., 2014). I base the following evaluation on the version IMAGE 3.0 as documented in Stehfest et al. (2014).

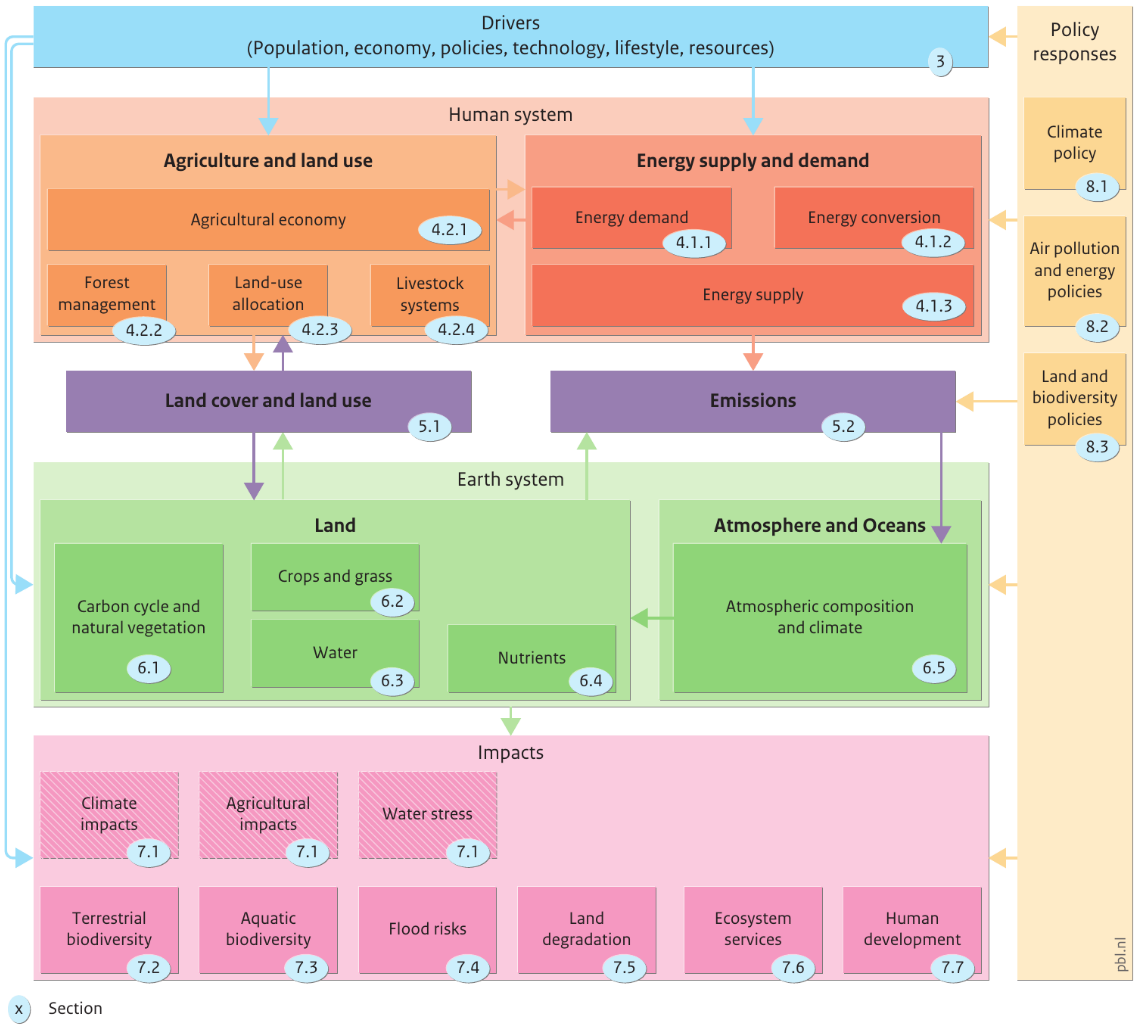
5.1 Target System of IMAGE

The stated aim of IMAGE is to provide a "comprehensive integrated modelling framework of interacting human and natural systems" (Stehfest et al., 2014, p. 17). It is structured along the causal chain of climate change and other global environmental problems, by starting with socioeconomic drivers, modelling their impact on the natural environment and finally the effect on human development. As Figure 8 shows, IMAGE consists of two main components, the *Human system* and the *Earth system*. The Human system comprises the agriculture and land-use sector as well as the energy system, while the Earth system consists of land components and a climate module. The two systems are linked through land use and emissions, which allow for feedbacks between climate developments, agriculture and the energy sector.

IMAGE can thus be analysed as a representation of the Human and Earth system, with a focus on those subsystems that play important causal roles in global sustainability problems like climate change. This representational core is complemented with drivers, impacts and policy responses. Drivers are exogenous model inputs that provide projections of population, economic indicators, policy assumptions and other socioeconomic parameters. Impacts are detailed model outputs, with separate components for different global environmental challenges respectively. Policy responses are assumptions that affect every other component of IMAGE, such that impacts can be compared for situations with different policy configurations.

The single IMAGE modules, each representing a certain subsystem or process, are developed separately from each other and can vary in terms of spatial or temporal resolutions (Stehfest et al., 2014). For model evaluation, these components are primarily assessed in isolation through sensitivity analyses and the introduction of alternative model formulations. Further, the performance of IMAGE as a whole has frequently been assessed through model intercomparison projects (Kriegler, Riahi et al., 2015; Luderer et al., 2018; McCollum et al., 2018; Tavoni et al., 2015; Vrontisi et al., 2018).These structured comparisons of different IAMs allow for drawing inferences about specific tendencies or emerging properties of a model, as compared to others. For example, IMAGE has been shown to respond strongly to carbon prices, while projecting comparatively low mitigation costs (Kriegler, Petermann et al., 2015).

In comparison to other IAMs, IMAGE has a very detailed representation of biophysical pro-



Source: Stehfest et al. (2014, p. 18)

Figure 8: Structure of IMAGE 3.0. The Human system and the Earth system form the modelling core, exogenous drivers provide socioeconomic and political pathways, policy responses alter individual model components as well as drivers, and finally environmental and human impacts are calculated.

cesses as well as environmental and human development indicators (Stehfest et al., 2014). As such, it is well positioned to analyse land use dynamics and interactions with other sustainability issues such as biodiversity (Doelman et al., 2018; Molotoks et al., 2018; van Vuuren et al., 2015). Conversely, IMAGE has less detail on economic and policy issues than other IAMs. In fact, as can be seen in Figure 8, it has no macroeconomic module at all. Instead, economic processes in the agriculture and energy sector are modelled as part of the respective modules. For example, international trade in fossil fuels is represented within the energy supply submodule. All socioeconomic variables which are not part of the agriculture or energy sector, are exogenously prescribed by the drivers.

As an exemplary component of IMAGE, I will briefly analyse its energy system and the way in which it represents its target system. The energy system module of IMAGE is called TIMER (PBL, 2005). Similar to IMAGE as a whole, TIMER is a simulation model (Stehfest et al., 2014). This means that the system behaviour is projected into the future based on a set of deterministic algorithms and the initial state of each variable. In order to be in accordance with observed trends, the model is calibrated against historical data from 1971-2015 (van Vuuren et al., 2021). TIMER models energy demand based on exogenous economic variables and a bottom-up description of different economic sectors. For every sector, it assumes a similar evolution of energy intensity over time – an initial increase as the sector develops and finally a decline as the sector grows more mature (PBL, 2005, pp. 25f). TIMER also accounts for technological developments by introducing autonomous energy efficiency increases dependent on the economic growth rate (van Ruijven et al., 2010). In several uncertainty analyses of TIMER, those two assumptions have been shown to have the highest influence on model output (Risbey et al., 2005; Van der Sluijs et al., 2002; van Vuuren et al., 2008).

After having simulated energy demand, TIMER calculates energy conversion and supply (Stehfest et al., 2014). The resulting energy mix is dominated by two main mechanisms: resource depletion and technological change. The former increases costs over time, especially for non-renewable resources, while the latter reduced costs. This cost reduction is, as in REMIND, modelled through learning curves, whereby costs decline as a power function of cumulative capacity. In contrast to REMIND, TIMER models learning-by-doing dynamics for almost all energy sources. Renewable sources, however, exhibit the strongest learning effects. The value of the learning rate strongly influences long-term price trends, such that this parameter is found to be among the most influential on TIMER's results (Edenhofer et al., 2010; Risbey et al., 2005). Overall, TIMER represents the energy system with high technological detail and is judged to be moderately robust to changes in assumptions: one of the modellers thinks that results "could probably be changed by a factor of two or so without much tinkering with parameter values, but not by a factor of ten without requiring implausible changes to the model" (Risbey et al., 2005, p. 66).

Being a simulation model, decisions in TIMER are not taken under the assumption of perfect foresight. This implies that the trajectories simulated by TIMER are not necessarily intertemporally 'optimal' (Stehfest et al., 2014). Further, TIMER is not able to "examine macroeconomic consequences of mitigation strategies, such as GDP losses" (Stehfest et al., 2014, p. 74), because it is not directly linked to a macroeconomic module. In fact, as a macroeconomic module is absent from IMAGE in general, this statement is valid for the whole modelling framework. Thereby, monetary feedbacks between components are not well represented, which risks to disregard environmental impacts of technological investments (Pauliuk et al., 2017).

5.2 Using IMAGE

According to the IMAGE developers, the model has three main objectives (Stehfest et al., 2014, p. 14): 1) analysing large-scale and long-term interactions between human development and the natural environment; 2) indicating key interlinkages and associated levels of uncertainty; 3) identifying response strategies to global environmental change based on an assessment of options for mitigation and adaption. While having a representation of key human and environmental processes is crucial for these objectives, process representation alone is not sufficient for explaining how, for example, response strategies are identified. In order to understand that, it is helpful to look at a central narrative that permeates the IMAGE documentation. The model is structured around two types of stories: a baseline scenario that projects a world without "deliberate, drastic changes changes in prevailing [...] developments", and policy scenarios that include measures to prevent unwanted impacts (Stehfest et al., 2014, p. 14). Comparing the impacts of the baseline with those of a policy scenario is one way in which IMAGE helps to identify response strategies. Further, it often features qualitatively different policy scenarios in order to explore the range of possible pathways for achieving certain goals. For example, IMAGE is used to analyse trade-offs and synergies between different ways of achieving climate and biodiversity targets, by projecting three different policy and socioeconomic scenarios called 'Global technology', 'Decentralised solutions' and 'Consumption change' (van Vuuren et al., 2015). While all three pathways achieve the targets, they do so in different manners, which allows for a broad assessment of possible policy options and associated impacts.

Each scenario run aims to describe a certain possible climate change trajectory (van Vuuren, Isaac et al., 2011). Scenario modelling can thus be analysed in terms of 'credible worlds' (Sugden, 2000) – each projection aims to depict a stylised and contingent, but plausible future development. For a modelled world to be credible, it has to be coherent in its assumptions, and these in turns have to cohere with intuitions about causal processes in the real world (Sugden, 2000). This means that not any set of assumptions is equally valuable for learning about the real world. If IMAGE wants to tell compelling stories about possible future worlds, it would not be sufficient to model any combination of scenario assumptions. Instead, there has to be some form of judgement in selecting scenario drivers that are both coherent among themselves and with respect to real-world developments. IMAGE chooses drivers based on an exercise where brief and consistent scenario storylines are formulated first, in order to ensure coherence for the subsequent implementation of assumptions in different model components (Stehfest et al., 2014, p. 36). To the extent that they are able to capture the relevant causal processes for a research question, IMAGE scenarios can thus be treated as credible worlds. This ability to project credible scenarios based on different socioeconomic and policy assumptions makes IMAGE a useful tool for global sustainability-related assessments. In the context of climate change, these assumptions are mainly provided by RCP or SSP scenarios (Riahi et al., 2017; van Vuuren, Edmonds et al., 2011). The RCP scenarios span four different emission pathways, while the SSPs outline five narratives of global socioeconomic development. IAMs are then used to quantitatively project the implications of respective scenario assumptions.

Since the different SSPs are fully specified socioeconomic narratives, they can fully take the role of determining overarching dynamics, while the core IMAGE model simulates associated agriculture, energy, land use and climate processes, including feedbacks among them. As part of the IPCC scenario process, IAMs often conduct a form of cost-effectiveness analysis – by constraining the model such as to stay below a certain climate target. Most models do this by imposing a constraint on the optimisation routine. IMAGE, however, does not feature an optimisation mechanism. It is therefore complemented by the FAIR model (den Elzen & Lucas, 2005), which is able to force the IMAGE model onto an emissions pathway that is consistent with a given exogenous target. The FAIR model is a decision support tool that calculates the cheapest path towards a target, while taking into account different types of GHGs (Stehfest et al., 2014). The mitigation costs that these calculations are based on, come from IMAGE, to which FAIR is coupled. Similar to IMAGE, FAIR features different regions, such that it can analyse effects of regional climate targets and emissions trading under different effort-sharing regimes (Hof et al., 2009). Further, FAIR is used to derive economic variables for different IMAGE pathways. It can, for example, estimate aggregated mitigation and damage costs (Hof et al., 2008) or analyse the effects of adaptation on climate policy (Hof et al., 2010). The inclusion of FAIR into the IMAGE framework is thus an interesting case that shows the flexibility of the modelling framework. While IMAGE, as a simulation model, has no way of assessing minimum-cost strategies and other macroeconomic issues, modellers found a way of extending it such that it is able to fulfil its purposes within the IPCC process.

5.3 Transparency of IMAGE

As seen above, the core of IMAGE aims to provide a representation of biophysical processes of global change and energy and agriculture systems. Based on different socioeconomic drivers and policy assumptions, it is used to analyse policy options and impacts of different possible pathways. In the IPCC process, it is used to quantify SSP scenarios and perform climate policy analyses together with the FAIR model. Each of these applications comes with different requirements, and it is on the modelling agent to provide a model commentary that explicates the model's different roles and the resulting consequences (Mäki, 2009).

The release of IMAGE 3.0 has been accompanied by an extensive model documentation in the form of a book (Stehfest et al., 2014). What sets is apart from other documentations, e.g. DICE's or REMIND's, is the fact that it is both very structured and accessible. The whole modelling framework is introduced in non-technical language, including comparisons to other IAMs and background on the history and applications of IMAGE. Then, as can be seen in Figure 8, each component is described in isolation. These descriptions all follow the same structure, consisting of: 'introduction', 'model description', 'policy issues', 'data, uncertainties and limitations', 'key publications' and 'input/output table'. Thereby, every model component is contextualised, described in detail and critically discussed. The inclusion of policy issues in every module is structured along IMAGE's two main stories: baseline developments and policy interventions. This guarantees that, for IMAGE as a whole, justifications for modelling choices within every component are provided, and policy-relevant alternatives are outlined.

The documentation of IMAGE is also revealing of what IMAGE modellers consider the model to be: a detailed description of human and environmental processes, which analyses the effects of exogenous socioeconomic drivers and differentiates between baseline developments and policy interventions. While IMAGE can be used to support the analysis of normative futures, it is not perceived to be normative itself – the over 350 pages long model documentation does not contain the words 'ethical', 'moral', 'prescriptive' or 'normative'[1]. In light of the fact that this documentation also includes the climate policy model FAIR, which for example allocates emission permits and mitigation costs across time and space, one would expect it to at least acknowledge the ethical importance of the modelled phenomena.

5.4 Expectations for IMAGE

The main expectation for IMAGE is linked to the portrayal of IAMs in its model documentation, as "describ[ing] the key processes in the interaction of human development and the natural environment" (Stehfest et al., 2014, p. 14). Related to that are evaluation criteria of realisticness and completeness, in that an IAM should include everything that is 'key' in a sufficiently realistic description. Users holding this expectations could be institutions or actors that need detailed projections of both environmental and social processes – such as governments, planning agencies or reinsurance companies. Relevant users who could expect a complete and sufficiently realistic description of its target system are also international institutions, such as the OECD, which used IMAGE to develop an environmental baseline scenario (OECD, 2012). IMAGE, as can be seen from its structure in Figure 8, incorporates a large amount of environmental processes and impacts. The level of detail of IMAGE is described as "intermediate complexity" (Stehfest et al., 2014, p. 14) – it is thus not able to resolve fine-grained spatial or temporal dynamics. One of

[1] To be precise, 'normative' does occur once, but not in the context of IMAGE. It is used to describe the Global Energy Assessment to which it contributed. Further, 'norms and values' are discussed, but as part of socioeconomic drivers, not as part of IMAGE itself.

the main weaknesses of IMAGE is the omission of economic processes that involve several model components. Thereby, consistency in monetary flows is not automatically guaranteed, which can lead to a failure of capturing the emission effects related to shifting investments (Pauliuk et al., 2017). The expectation of describing key processes of human development and the natural environment is thus only partly fulfilled – while a large amount of environmental processes are accounted for, processes of human development are limited to agriculture, land use and the energy system, disregarding further macroeconomic effects.

A second expectation for IMAGE is connected to serving as a useful tool for policy advice. In its model documentation, this is phrased as: "IMAGE can be used as a tool to construct long-term scenarios and is often deployed to feed policy analysis". (Stehfest et al., 2014, p. 9). The model users in this case are easily identified as belonging to the policy process. Associated evaluation criteria are more difficult to establish. A look at how IMAGE is used for policy advice reveals that it mainly works through scenario projections: the impacts of a range of different policy scenarios are compared to the impacts of a baseline scenario (Stehfest et al., 2014, p. 15). A key evaluation criterion is thus the ability of reliably assessing the merits and downsides of scenarios. Since IMAGE does not aggregate its results into a single monetary metric, this means that the individual projections of environmental impacts, such as temperature rise, water scarcity or biodiversity loss must be reliable enough to form a basis for decision-making. Interestingly, this second expectation is thereby not fundamentally different from the first expectation which is connected to a realistic representation of the target system. Additionally, though, being able to project many alternative scenarios also requires a certain flexibility that enables the model to be readily adapted. For IMAGE, this is facilitated through its clear separation of the representational core, consisting of Human and Earth system, on the one hand, and socioeconomic drivers and policy assumptions on the other hand. For a particular scenario projection, the representational core does not usually have to be modified and most scenario assumptions can be included as part of drivers and policies. Thereby, IMAGE lends itself well to projecting impacts of different scenarios. However, this conversely implies that scenarios have to be previously specified for the whole modelling time frame and interactions between the representational core and socioeconomic and policy variables can not be assessed.

Lastly, IMAGE is faced with an expectation linked to providing a comprehensive and informative model commentary. This expectation can be held by any user who is trying to make sense of IMAGE results – thereby mainly people connected to the policy process or from within academia. Further, as IMAGE is one of the models used in IPCC reports (IPCC, 2014), a wide range of users have an interest in understanding how it works, for example journalists or citizens. An associated evaluation criterion is transparency, or deep transparency as defined by Bistline et al. (2021). As analysed above, IMAGE has a very comprehensive and well-structured model documentation. The scope and objectives are introduced, and baseline and policy developments for every model component are outlined, respectively. Further, each element of the modelling framework comes with a discussion of data sources, limitations and uncertainties. Thus, IMAGE performs very well with respect to this expectation – with the caveat of not providing any guidance on identifying potentially normative elements of the modelling exercise.

Evaluating IMAGE has highlighted two aspects that set it apart from other IAMs. First, the model has a well-defined target system, consisting mainly of agriculture and land use systems, the energy system and the climate system. In order to represent these systems, IMAGE resorts to simulation based on exogenous drivers, has a decisively descriptive aim and a high resolution both spatially and in terms of processes. Thereby, it is structurally comparable to climate models from the natural sciences. In fact, in an overview publication about Earth System Models of Intermediate Complexity (EMICs, Claussen et al., 2002), IMAGE was cited as an example. Second, IMAGE is interesting in that the requirements for its use as a policy-relevant tool align largely with its aim of realistically representing a target system. Thereby, it is able to provide a consistent account and model commentary about its own working and the interpretation of its results. This is facilitated by the fact that most IMAGE applications rely on projecting the impacts of specific scenarios. In the IPCC context, however, it is often used to conduct cost-effectiveness analyses by coupling it to the FAIR model, which introduces the risk of losing consistency with respect to the interpretation of simulation and optimisation, and descriptive and normative purposes, respectively. Overall, IMAGE can be characterised as a comprehensive model of human impacts on the environment, which is able to analyse different policy options and flexible enough to be used for climate change mitigation analysis, but not suitable for analysing macroeconomic processes.

6 Discussion

6.1 Target System

All three analysed IAMs have – more or less explicitly – referenced the target system they aim to represent. Often, this is expressed through language of causality. DICE aims to capture the whole causal loop from emissions to policy that again affects emissions, and IMAGE aims to capture the causal chain of global environmental change, from socioeconomic drivers to environmental and human development impacts. This distinction between a causal loop and a causal chain is reflected by their respective model structures, with circular causality in the policy optimisation model DICE and more linear causality in the policy evaluation model IMAGE. REMIND, on the other hand, describes its target system – without alluding to a flow of causality – as "the global energy–economy–emissions system" (Baumstark et al., 2021, p. 6571).

The main dividing line in how the three IAMs represent their targets is the level of detail. While DICE tends to model processes with only a few equations, REMIND and IMAGE have detailed representations of energy and land use sectors – the main emission sources and thereby biggest levers for mitigation. DICE, on the other and, represents these sectors with limited detail and in a very stylised manner. Here, they are modelled through simple mitigation cost curves that do not distinguish different technologies and do not exhibit temporal dynamics (Gillingham & Stock, 2018; Grubb et al., 2021). REMIND and IMAGE both represent different energy technologies in great detail and provide technology-specific assumptions on endogenously determined future cost developments. These assumptions have also been subject to criticism, e.g. for introducing many more degrees of freedom (Ives et al., 2021), for underestimating the cost decline of solar energy (Creutzig et al., 2017) or for relying heavily on BECCS (Lenzi et al., 2018). Nevertheless, the ability of representing a large amount of mitigation technologies and their endogenous cost trajectory – however uncertain – clearly sets REMIND and IMAGE apart from DICE.

A second dividing line separates IMAGE, with its focus on biophysical processes and simulation, from REMIND and DICE, both based on economic growth models and intertemporal optimisation. Despite this division, all three models face challenges in representing the economic aspects of their target system. DICE and REMIND suffer from the inability of the growth model to capture long-term dynamics (Millner & McDermott, 2016) and from the many idealisations that this framework requires (Staub-Kaminski et al., 2014). IMAGE, on the other hand, suffers from the fact that it does not capture macroeconomic feedback at all – economic processes in IMAGE are represented either within a model component or as externally provided relationships. Deep challenges to the three IAMs are further posed by the representation of economic damages from climate change. DICE is the only model that has included damage representations from the beginning, but its modelling choices are subject to serious doubt (Diaz & Moore, 2017). I have not evaluated the performance of REMIND's recently incorporated damage module, but it faces the same empirical and theoretical challenges as DICE's. However, the representation of economic dynamics is also impacted by the omission of climate damages – as was the case for REMIND until 2021 and is still the case for IMAGE. While IMAGE does include climate impacts, these do not translate into a reduction of GDP, which is provided exogenously.

6.2 Purpose and Use

The three IAMs share some essential features in the way in which they are used. The concept of pathway is key to all models, especially the duality between a baseline path and a policy path. This setup allows for the comparison of climate inaction and climate action. However, this is done in terms of different metrics for the three IAMs. DICE and REMIND both compare pathways in terms of their effect on global welfare. Additionally, REMIND often compares further quantities related to technologies and regions, whereas IMAGE mainly compares pathways in terms of their impacts on the environment and human development. The perspective of models as structure and story (Gibbard & Varian, 1978) has proven to be a useful lens for capturing these features of IAMs. The baseline pathway and the policy pathway can both be understood as a story told through the model. According to Morgan (2001), the function of a story is to link the model to the world. In that sense, IAMs are related to the world of climate policy through these two stories: one in which no (further) climate policy is undertaken and one in which certain climate policies are implemented. Further, IAMs perform the role of epistemic tools by being easily manipulated to incorporate different assumptions or constraints, which leads to a whole range of additional pathways. Thereby, it is not only one policy pathway that is compared to a baseline path, but different policy pathways which can be compared among each other in order to determine the relative effects of different assumptions and mechanisms.

The applications of the three IAMs are nonetheless very different. REMIND and IMAGE are often used within the IPCC scenario framework consisting of SSP and RCP scenarios. As such, they are employed to provide comprehensive and consistent projections of possible futures, as well as to analyse specific policy options and interactions among them. Thereby, they tell stories about high and low challenges to mitigation and adaptation respectively, which provide an important framing for further research use or policy advice (van Vuuren et al., 2017). DICE, on the other hand, operates on a much more aggregated and simplified level. It is often adapted by other researchers in order to investigate the effect of specific assumptions or processes on ‘optimal’ levels of warming or the SCC. Based on its open-source model code, a large collection of adapted DICE models has emerged. The aggregate nature of DICE and the fact that it can easily modified by other researchers, are thus key elements of how the model operates as an epistemic tool. This property of DICE also clearly distinguishes it from REMIND and IMAGE, which are both developed at larger research institutions and not normally adapted by other researchers.

6.3 Interpretations and Transparency

While all three IAMs are able to produce a wide range of climate mitigation pathways, the interpretation of these model results is not always obvious. For example, do these pathways depict the world as it is, as it could be, or as it should be? I have analysed how literature around DICE is torn between two different interpretations of its modelling framework, as it is partly aiming to descriptively represent real-world dynamics, while also providing normative guidance on ‘optimal’ climate policy. Similar complications arise for REMIND, where the interpretation of a descriptive representation of the world economy under climate change is coupled to the interpretation of idealised scenarios that serve as benchmarks. The interpretation of IMAGE is more consistent in this respect; it is unequivocally taken to be a descriptive representation of the Human and Earth System, where normative elements enter through exogenous scenarios and the climate policy model FAIR. The documentation of FAIR, however, is also not entirely clear about how its modelling is to be interpreted. To some extent, all three IAMs suffer from a form of conflation of descriptive and normative modelling elements. IMAGE, of the three models, is the most consistent in its interpretation, by separating the descriptive core from socioeconomic and policy aspects – and by outsourcing its interpretation problems to the FAIR model.

The transparency of critical assumptions has been an important element of my model evaluation. More broadly, Mäki's (2009, 2018) concept of a model commentary captures the way in which modellers communicate about their models, about what they do, and about how to interpret the results. I find that, while all three IAMs come with regularly updated model descriptions, critical assumptions with ethical dimensions are only discussed for DICE – and even here not in all cases. REMIND, by being structurally similar, is confronted with many of the same modelling choices that DICE discusses, yet they are often not acknowledged by its model commentary. On top of giving guidance on the justification and effect of specific assumptions, a model commentary should clarify the role and purpose of the model. Part of this is an account of how the model should be interpreted. As seen in the previous paragraph, DICE and REMIND both fail to provide a consistent account about the model's interpretation – which is a serious shortcoming to any model commentary.

6.4 Expectations

Each of the three IAMs was evaluated according to three expectations. These expectations followed similar patterns for each model, but were adapted to specificities that emerged from the previous analysis along different evaluation criteria and perspectives on modelling. This assessment and evaluation of expectations for DICE, REMIND and IMAGE is thus not without limitations. A different emphasis in the analysis of each IAM would likely have lead to slightly different expectations. Further, some elements of the expectation concept are hard to pin down, most notably the aims of model users. By basing the thesis largely on academic literature, these perspectives are more abundant than those from non-academic model users. A last complicating factor in the formulation of distinct expectations for each of the three models lies in the discourse around IAMs. In much of the literature discussing the merits and problems of IAMs, this is done so under the general term of 'IAMs', which complicates the assessment of expectations which are more specific to a certain type of IAMs. For these reasons, the analysed expectations should be regarded as illustrative examples, rather than a result of comprehensive expectation mapping. Nonetheless, the comparison of these illustrative expectations for all three IAMs allows me to draw some inferences both about the respective models and the expectations of them.

The first expectation for the three IAMs varies between them, but it shares a certain descriptive element. For DICE, this expectation is linked to calculating the costs and benefits of different climate policy options through quantitative forecasts. The model was found not to be able to live up to this – however, it is also not the aim of DICE to provide reliable quantitative forecasts. For REMIND, the expectation is about representing future developments of the world economy. It was found to partially live up to it, through its detailed energy system module and regionalised macroeconomic module. However, the macroeconomic representation is still very aggregated and REMIND is not capable of capturing the temporal evolution of socioeconomic parameters endogenously, but relies mainly on exogenous projections. For IMAGE, the first expectation is linked to describing key processes of interacting environmental and human systems. The model was found to partially fulfil this – in that it provides a very detailed representation of environmental, energy-related and agricultural processes, while lacking a representation of overarching socioeconomic processes. A comparison of the three models reveals that DICE is expected to represent aggregated costs and benefits, while REMIND and IMAGE are rather expected to provide detailed process representations.

The respective second expectation of the evaluated IAMs is connected to model purposes and their usefulness as epistemic tools. For DICE, this expectation encompasses the purpose of learning about the represented system through targeted manipulations of the model. It was found to be able of fulfilling this, due to its tractability and open-source code, which make it into a flexible epistemic tool and led to a range of variations of the DICE model. For REMIND, this expectation is connected to the assessment of mitigation pathways as well as synergies and trade-offs between them. It is found to live up to this through its flexible modelling framework and its detailed representation of mitigation options. However, REMIND is only capable of yielding qualitative insights, as its quantitative projections are fraught with uncertainties. For IMAGE, the second expectation is similarly connected to the provision of policy-relevant information through the assessment of different scenarios. The model is found to be capable of living up this expectation, as it is designed to project the impacts of different scenario assumptions. However, the separation between scenario drivers and resulting projections implies that IMAGE can not account for socioeconomic processes affecting scenario drivers. A comparison of the three IAMs shows that they are all expected to be flexible enough to run a variety of pathways and thereby tell a variety of stories that highlight aspects of their target systems.

The third expectation is similar for all three IAMs, and it is linked to transparency and the provision of a comprehensive and informative model commentary. DICE was found to fulfil this expectation partially, in that it discusses at least some critical modelling choices and their implications. REMIND fails at providing sufficient transparency about critical assumptions, whereas IMAGE is successful at giving a comprehensive assessment of its critical modelling choices. However, all three IAMs have difficulties with respect to a comprehensive model commentary that includes the question of whether the models and their output should be interpreted descriptively or normatively – and where to draw the boundary between respective aspects.

Overall, evaluating expectations for the analysed models reveals some high-level commonalities: All three IAMs are unable to provide quantitative forecasts of their target system, but they are flexible enough to yield qualitative insights by modelling many different pathways. At the same time, they have room for improvement in terms of their model commentary, both for highlighting critical assumptions and for providing guidance on descriptive and normative model interpretations.

7 Conclusion

Having analysed DICE, REMIND and IMAGE according to the expectations that they are faced with, what can be said about the usefulness of this approach? The main motivation for developing it lies in the fact that there are a variety of possible evaluation criteria, where none seem apt for evaluating more than certain aspects of IAMs. Through the notion of expectations, I was able to couple these criteria to specific modelling purposes and user perspectives, and thereby evaluate IAMs in a more comprehensive way. The flexibility of this approach also means that there is no straightforward way of determining an expectation. Whether all expectations have been mapped, or whether the ones that have been analysed are the most prominent ones, is difficult to establish. However, I was able to obtain an illustrative set of expectations for each model, by analysing how the model works, how its purposes are portrayed by the modellers and how this relates to certain demands from model users. And it is through analysing this interaction between models, purposes and users that the most interesting findings of this thesis emerged.

Based on the previous analysis of each model, I return to the question of what can reasonably be expected of IAMs. DICE has proven to be a useful tool for investigating the relative effects of different assumptions on model output. It can therefore be expected to yield insights about specific aspects of climate policy, but should not be asked to provide quantitative guidance. IMAGE, on the other hand, has proven to be suitable for projecting the environmental impacts of different socioeconomic and policy pathways. It can therefore be expected to contribute to a better understanding of global environmental change and the impacts of certain policy choices, but should not be used for analysing questions that require a description of macroeconomic processes. REMIND shares features of both DICE and IMAGE, and it can be expected to investigate effects of different assumptions on technology, economic indicators or climate policy, while providing detailed pathways of mitigation technologies. Yet, the precise output about different regions and technologies should not be seen as a forecast, but rather as an assessment of different theoretically possible scenarios.

The evaluation of IAMs has also shown how the elements that constitute an expectation – purpose, user and criterion – can be rather general and ambiguous. On the one hand, this is a challenge for the expectations approach as such. On the other hand, it reveals something very important about IAMs themselves, and about how they are used. I was able to show that, while the analysed IAMs are well-documented in general, they are not always clear about the model purpose and the interpretation of its results. Yet, questions about whether a model is seen as modelling the world from a descriptive or a normative standpoint, deeply affect the expectations for it. As users partly base their view of the model on the way in which it is presented by the developer, model commentary is particularly important to the formulation of clear expectations. In an ideal world, models and the expectations of them would match. In reality, however, expectations of IAMs are not always met with according capabilities on behalf of the models. While working to improve IAMs would be one possible way of reacting to this mismatch, I suggest to take an alternative route. Rather than solely focusing on making the models align with expectations, much could be gained by adjusting expectations to what the models can actually deliver. If IAM users were able to voice their needs more clearly, and if modellers were to provide a better model commentary about what the models can deliver and how that is to be interpreted, evaluation could be based on a much more solid footing. Altogether, evaluating IAMs along possible expectations of them has revealed their respective strengths and weaknesses and provides the basis for further discussions about what IAMs are capable of, what we can expect of them, and how these questions are communicated. This last point captures the biggest room for improvement around IAMs: Expectations can only be formulated clearly on the basis of an informative model commentary – and we can only evaluate a model if we know what to expect of it.

Bibliography

Acemoglu, D. (2009). *Introduction to modern economic growth*. Princeton University Press. https://press.princeton.edu/books/hardcover/9780691132921/introduction-to-modern-economic-growth

Acemoglu, D., Aghion, P., Bursztyn, L. & Hemous, D. (2012). The environment and directed technical change. *American Economic Review*, *102*(1), 131–66. https://doi.org/10.1257/aer.102.1.131

Ackerman, F., DeCanio, S., Howarth, R. & Sheeran, K. (2009). Limitations of integrated assessment models of climate change. *Climatic Change*, *95*, 297–315. https://doi.org/10.1007/s10584-009-9570-x

Ackerman, F. & Finlayson, I. J. (2006). *The economics of inaction on climate change: A sensitivity analysis* (GDAE Working Papers No. 06-07). GDAE, Tufts University. https://ideas.repec.org/p/dae/daepap/06-07.html

Ackerman, F., Stanton, E. A. & Bueno, R. (2013). Epstein-Zin utility in DICE: Is risk aversion irrelevant to climate policy? *Environmental & Resource Economics*, *56*(1), 73–84. https://doi.org/10.1007/s10640-013-9645-z

Aldy, J. E. & Stavins, R. N. (2020). Rolling the dice in the corridors of power: William Nordhaus's impacts on climate change policy. *Climate Change Economics*, *11*(04), 2040001. https://doi.org/10.1142/S2010007820400011

Anderson, K. & Jewell, J. (2019). Debating the bedrock of climate-change mitigation scenarios. *Nature*, *573*(7774), 348–349. https://doi.org/10.1038/d41586-019-02744-9

Anderson, K. & Peters, G. (2016). The trouble with negative emissions. *Science*, *354*(6309), 182–183. https://doi.org/10.1126/science.aah4567

Arrow, K. J., Cline, W. R., Maler, K. G., Munasinghe, M., Squitieri, R. & Stiglitz, J. E. (1996). Intertemporal equity, discounting, and economic efficiency. In J. Bruce, H. Lee & E. Haites (Eds.), *Climate Change 1995: Economic and Social Dimensions of Climate Change, Contribution of Working Group III to the Second Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://www.ipcc.ch/report/ar2/wg3/

Asefi-Najafabady, S., Villegas-Ortiz, L. & Morgan, J. (2020). The failure of Integrated Assessment Models as a response to 'climate emergency' and ecological breakdown: The Emperor has no clothes. *Globalizations*, *0*(0), 1–11. https://doi.org/10.1080/14747731.2020.1853958

Azar, C. A. (1998). Are optimal CO_2 emissions really optimal? *Environmental & Resource Economics*, *11*, 301–315. https://doi.org/10.1023/A:1008235326513

Bailer-Jones, D. M. (2003). When scientific models represent. *International Studies in the Philosophy of Science*, *17*(1), 59–74. https://doi.org/10.1080/02698590305238

Bastien-Olvera, B. A. & Moore, F. C. (2021). Use and non-use value of nature and the social cost of carbon. *Nature Sustainability*, *4*(2), 101–108. https://doi.org/10.1038/s41893-020-00615-0

Bauer, N., Baumstark, L. & Leimbach, M. (2012). The REMIND-R model: the role of renewables in the low-carbon transformation-first-best vs. second-best worlds. *Climatic Change*, *114*(1, SI), 145–168. https://doi.org/10.1007/s10584-011-0129-2

Bauer, N., Bertram, C., Schultes, A., Klein, D., Luderer, G., Kriegler, E., Popp, A. & Edenhofer, O. (2020). Quantification of an efficiency-sovereignty trade-off in climate policy. *Nature*, *588*(7837), 261+. https://doi.org/10.1038/s41586-020-2982-5

Bauer, N., Edenhofer, O. & Kypreos, S. (2008). Linking energy system and macroeconomic growth models. *Computational Management Science*, *5*(1), 95–117. https://doi.org/10.1007/s10287-007-0042-3

Baumstark, L., Bauer, N., Benke, F., Bertram, C., Bi, S., Gong, C. C., Dietrich, J. P., Dirnaichner, A., Giannousakis, A., Hilaire, J., Klein, D., Koch, J., Leimbach, M., Levesque, A., Madeddu, S., Malik, A., Merfort, A., Merfort, L., Odenweller, A., . . . Luderer, G. (2021). REMIND2.1: Transformation and innovation dynamics of the energy-economic system within climate and sustainability limits. *Geoscientific Model Development*, *14*(10), 6571–6603. https://doi.org/10.5194/gmd-14-6571-2021

Beck, M. (2017). *Modeling climate change impacts at the science-policy boundary* (Doctoral dissertation). University of Waterloo. UWSpace. http://hdl.handle.net/10012/11229

Beck, M. (2018). Telling stories with models and making policy with stories: An exploration. *Climate Policy*, *18*(7), 928–941. https://doi.org/10.1080/14693062.2017.1404439

Beck, M. & Krueger, T. (2016). The epistemic, ethical, and political dimensions of uncertainty in integrated assessment modeling. *WIREs Climate Change*, *7*(5), 627–645. https://doi.org/10.1002/wcc.415

Beck, S. & Oomen, J. (2021). Imagining the corridor of climate mitigation – What is at stake in IPCC's politics of anticipation? *Environmental Science & Policy*, *123*, 169–178. https://doi.org/10.1016/j.envsci.2021.05.011

Bertram, C., Luderer, G., Pietzcker, R. C., Schmid, E., Kriegler, E. & Edenhofer, O. (2015). Complementing carbon prices with technology policies to keep climate targets within reach. *Nature Climate Change*, *5*(3), 235–239. https://doi.org/10.1038/nclimate2514

Bistline, J., Budolfson, M. & Francis, B. (2021). Deepening transparency about value-laden assumptions in energy and environmental modelling: Improving best practices for both modellers and non-modellers. *Climate Policy*, *21*(1), 1–15. https://doi.org/10.1080/14693062.2020.1781048

Bosetti, V. (2021). Integrated assessment models for climate change. https://doi.org/10.1093/acrefore/9780190625979.013.572

Botzen, W. J. W. & van den Bergh, J. C. J. M. (2014). Specifications of Social Welfare in Economic Studies of Climate Policy: Overview of Criteria and Related Policy Insights. *Environmental and Resource Economics*, *58*(1), 1–33. https://doi.org/10.1007/s10640-013-9738-8

Bruce, J., Lee, H. & Haites, E. (Eds.). (1996). *Climate change 1995: Economic and social dimensions of climate change, contribution of working group III to the second assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Budolfson, M., Dennig, F., Fleurbaey, M., Siebert, A. & Socolow, R. H. (2017). The comparative importance for optimal climate policy of discounting, inequalities and catastrophes. *Climatic Change*, *145*(3), 481–494. https://doi.org/10.1007/s10584-017-2094-x

Burgess, M. G., Ritchie, J., Shapland, J. & Pielke, R. (2020). IPCC baseline scenarios have over-projected CO2 emissions and economic growth. *Environmental Research Letters*, *16*(1), 014016. https://doi.org/10.1088/1748-9326/abcdd2

Burke, M., Hsiang, S. M. & Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, *527*(7577), 235–239. https://doi.org/10.1038/nature15725

Butnar, I., Li, P.-H., Strachan, N., Pereira, J. P., Gambhir, A. & Smith, P. (2020). A deep dive into the modelling assumptions for biomass with carbon capture and storage (BECCS): A transparency exercise. *Environmental Research Letters*, *15*(8), 084008. https://doi.org/10.1088/1748-9326/ab5c3e

Cai, Y., Judd, K. L., Lenton, T. M., Lontzek, T. S. & Narita, D. (2015). Environmental tipping points significantly affect the cost-benefit assessment of climate policies. *Proceedings of the National Academy of Sciences*, *112*(15), 4606–4611. https://doi.org/10.1073/pnas.1503890112

Cai, Y., Judd, K. L. & Lontzek, T. S. (2012). *DSICE: A dynamic stochastic integrated model of climate and economy* (tech. rep. No. ID 1992674). RDCEP. Rochester, NY. https://doi.org/10.2139/ssrn.1992674

Cai, Y. & Lontzek, T. S. (2019). The social cost of carbon with economic and climate risks. *Journal of Political Economy*, *127*(6), 2684–2734. https://doi.org/10.1086/701890

Calel, R. & Stainforth, D. A. (2017). On the physics of three integrated assessment models. *Bulletin of the American Meteorological Society*, *98*(6), 1199–1216. https://doi.org/10.1175/BAMS-D-16-0034.1

Cartwright, N. (2009). If no capacities then no credible worlds. but can models reveal capacities? *Erkenntnis*, *70*(1), 45–58. https://doi.org/10.1007/s10670-008-9136-8

Christensen, P., Gillingham, K. & Nordhaus, W. (2018). Uncertainty in forecasts of long-run economic growth. *Proceedings of the National Academy of Sciences*, *115*(21), 5409–5414. https://doi.org/10.1073/pnas.1713628115

Claussen, M., Mysak, L., Weaver, A., Crucifix, M., Fichefet, T., Loutre, M.-F., Weber, S., Alcamo, J., Alexeev, V., Berger, A., Calov, R., Ganopolski, A., Goosse, H., Lohmann, G., Lunkeit, F., Mokhov, I., Petoukhov, V., Stone, P. & Wang, Z. (2002). Earth system models of intermediate complexity: Closing the gap in the spectrum of climate system models. *Climate Dynamics*, *18*(7), 579–586. https://doi.org/10.1007/s00382-001-0200-1

Cointe, B., Cassen, C. & Nadaï, A. (2019). Organising Policy-Relevant Knowledge for Climate Action: *Science & Technology Studies*, *32*(4), 36–57. https://doi.org/10.23987/sts.65031

Creutzig, F., Agoston, P., Goldschmidt, J. C., Luderer, G., Nemet, G. & Pietzcker, R. C. (2017). The underestimated potential of solar energy to mitigate climate change. *Nature Energy*, *2*(9), 1–9. https://doi.org/10.1038/nenergy.2017.140

de Bruin, K. C., Dellink, R. B. & Tol, R. S. J. (2009). AD-DICE: an implementation of adaptation in the DICE model. *Climatic Change*, *95*(1-2), 63–81. https://doi.org/10.1007/s10584-008-9535-5

DeCanio, S. J. (2005). Descriptive or Conceptual Models? Contributions of Economics to the Climate Policy Debate. *International Environmental Agreements: Politics, Law and Economics*, *5*(4), 415–427. https://doi.org/10.1007/s10784-005-8333-z

Dellink, R., Chateau, J., Lanzi, E. & Magné, B. (2017). Long-term economic growth projections in the shared socioeconomic pathways. *Global Environmental Change*, *42*, 200–214. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2015.06.004

den Elzen, M. G. J. & Lucas, P. L. (2005). The FAIR model: A tool to analyse environmental and costs implications of regimes of future commitments. *Environmental Modeling & Assessment*, *10*(2), 115–134. https://doi.org/10.1007/s10666-005-4647-z

Dennig, F., Budolfson, M. B., Fleurbaey, M., Siebert, A. & Socolow, R. H. (2015). Inequality, climate impacts on the future poor, and carbon prices. *Proceedings of the National Academy of Sciences*, *112*(52), 15827–15832. https://doi.org/10.1073/pnas.1513967112

Diaz, D. & Moore, F. (2017). Quantifying the economic risks of climate change. *Nature Climate Change*, *7*(11), 774–782. https://doi.org/10.1038/nclimate3411

Dietrich, J. P., Bodirsky, B. L., Weindl, I., Humpenöder, F., Stevanovic, M., Kreidenweis, U., Wang, X., Karstens, K., Mishra, A., Beier, F. D., Molina Bacca, E. J., Klein, D., Ambrósio, G., Araujo, E., Biewald, A., Lotze-Campen, H. & Popp, A. (2020, September 15). MAgPIE - an open source land-use modeling framework - version 4.3.0. https://doi.org/10.5281/zenodo.1418752

Dietz, S., Bowen, A., Doda, B., Gambhir, A. & Warren, R. (2018). The economics of 1.5°C climate change. *Annual Review of Environment and Resources*, *43*(1), 455–480. https://doi.org/10.1146/annurev-environ-102017-025817

Dietz, S., Hepburn, C. & Stern, N. (2008). Economics, ethics and climate change. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199239979.003.0020

Dietz, S. & Stern, N. (2015). Endogenous Growth, Convexity of Damage and Climate Risk: How Nordhaus' Framework Supports Deep Cuts in Carbon Emissions. *The Economic Journal*, *125*(583), 574–620. https://doi.org/10.1111/ecoj.12188

Dietz, S., van der Ploeg, F., Rezai, A. & Venmans, F. (2021). Are Economists Getting Climate Dynamics Right and Does It Matter? *Journal of the Association of Environmental and Resource Economists*, *8*(5), 895–921. https://doi.org/10.1086/713977

Doelman, J. C., Stehfest, E., Tabeau, A., van Meijl, H., Lassaletta, L., Gernaat, D. E. H. J., Hermans, K., Harmsen, M., Daioglou, V., Biemans, H., van der Sluis, S. & van Vuuren, D. P. (2018). Exploring SSP land-use dynamics using the IMAGE model: Regional and gridded scenarios of land-use change and land-based climate change mitigation. *Global Environmental Change - Human and Policy Dimensions*, *48*, 119–135. https://doi.org/10.1016/j.gloenvcha.2017.11.014

Drupp, M. A., Freeman, M. C., Groom, B. & Nesje, F. (2018). Discounting Disentangled. *American Economic Journal: Economic Policy*, *10*(4), 109–134. https://doi.org/10.1257/pol.20160240

Drupp, M. A. & Hänsel, M. C. (2021). Relative prices and climate policy: How the scarcity of nonmarket goods drives policy evaluation. *American Economic Journal: Economic Policy*, *13*(1), 168–201. https://doi.org/10.1257/pol.20180760

Edenhofer, O., Bauer, N. & Kriegler, E. (2005). The impact of technological change on climate protection and welfare: Insights from the model MIND. *Ecological Economics*, *54*(2), 277–292. https://doi.org/10.1016/j.ecolecon.2004.12.030

Edenhofer, O., Knopf, B., Barker, T., Baumstark, L., Bellevrat, E., Château, B., Criqui, P., Isaac, M., Kitous, A., Kypreos, S., Leimbach, M., Lessmann, K., Magne, B., Scrieciu, S., Turton, H. & van Vuuren, D. (2010). The economics of low stabilisation: Exploring its implications for mitigation costs and strategies. *Energy Journal -Cambridge Ma*, *31*.

Edenhofer, O. & Kowarsch, M. (2015). Cartography of pathways: A new model for environmental policy assessments. *Environmental Science & Policy*, *51*, 56–64. https://doi.org/https://doi.org/10.1016/j.envsci.2015.03.017

Ellenbeck, S. & Lilliestam, J. (2019). How modelers construct energy costs: Discursive elements in Energy System and Integrated Assessment Models. *Energy Research & Social Science*, *47*, 69–77. https://doi.org/10.1016/j.erss.2018.08.021

Emmerling, J., Drouet, L., van der Wijst, K.-I., van Vuuren, D. P., Bosetti, V. & Tavoni, M. (2019). The role of the discount rate for emission pathways and negative emissions. *14*(10), 104008. https://doi.org/10.1088/1748-9326/ab3cc9

Farmer, J. D., Hepburn, C., Mealy, P. & Teytelboym, A. (2015). A third wave in the economics of climate change. *Environmental and Resource Economics*, *62*(2), 329–357. https://doi.org/10.1007/s10640-015-9965-2

Frank, D. M. (2017). Ethics of the scientist qua policy advisor: Inductive risk, uncertainty, and catastrophe in climate economics. *Synthese*, *196*(8), 3123–3138. https://doi.org/10.1007/s11229-017-1617-3

Friedman, M. (1953). The methodology of positive economics. In M. Friedman (Ed.), *Essays in positive economics* (pp. 3–43). University of Chicago Press.

Frigg, R. & Hartmann, S. (2020). Models in Science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2020). Metaphysics Research Lab, Stanford University. Retrieved January 18, 2022, from https://plato.stanford.edu/entries/models-science/

Frisch, M. (2013). Modeling climate policies: A critical look at integrated assessment models. *Philosophy & Technology*, *26*(2), 117–137. https://doi.org/10.1007/s13347-013-0099-6

Frisch, M. (2018). Modeling climate policies: The social cost of carbon and uncertainties in climate predictions. In E. A. Lloyd & E. Winsberg (Eds.), *Climate modelling* (pp. 413–448). Springer International Publishing. https://doi.org/10.1007/978-3-319-65058-6_14

Funtowicz, S. O. & Ravetz, J. R. (1994). Uncertainty, complexity and post-normal science. *Environmental Toxicology and Chemistry*, *13*(12), 1881–1885. https://doi.org/10.1002/etc.5620131203

Gambhir, A., Butnar, I., Li, P.-H., Smith, P. & Strachan, N. (2019). A review of criticisms of integrated assessment models and proposed approaches to address these, through the lens of BECCS. *Energies*, *12*(9), 1747. https://doi.org/10.3390/en12091747

Gambhir, A. (2019). Planning a Low-Carbon Energy Transition: What Can and Can't the Models Tell Us? *Joule*, *3*(8), 1795–1798. https://doi.org/10.1016/j.joule.2019.07.016

Giannousakis, A., Hilaire, J., Nemet, G. F., Luderer, G., Pietzcker, R. C., Rodrigues, R., Baumstark, L. & Kriegler, E. (2021). How uncertainty in technology costs and carbon dioxide removal availability affect climate mitigation pathways. *Energy*, *216*, 119253. https://doi.org/10.1016/j.energy.2020.119253

Gibbard, A. & Varian, H. R. (1978). Economic Models. *The Journal of Philosophy*, *75*(11), 664–677. Retrieved January 20, 2022, from https://www.jstor.org/stable/2025484

Giglio, S., Maggiori, M. & Stroebel, J. (2015). Very long-run discount rates. *The Quarterly Journal of Economics*, *130*(1), 1–53. https://doi.org/10.1093/qje/qju036

Gillingham, K., Nordhaus, W., Anthoff, D., Blanford, G., Bosetti, V., Christensen, P., McJeon, H. & Reilly, J. (2018). Modeling Uncertainty in Integrated Assessment of Climate Change: A Multimodel Comparison. *Journal of the Association of Environmental and Resource Economists*, *5*(4), 791–826. https://doi.org/10.1086/698910

Gillingham, K. & Stock, J. H. (2018). The cost of reducing greenhouse gas emissions. *Journal of Economic Perspectives*, *32*(4), 53–72. https://doi.org/10.1257/jep.32.4.53

Glanemann, N., Willner, S. N. & Levermann, A. (2020). Paris Climate Agreement passes the cost-benefit test. *Nature Communications*, *11*(1), 110. https://doi.org/10.1038/s41467-019-13961-1

Grant, N., Hawkes, A., Napp, T. & Gambhir, A. (2020). The appropriate use of reference scenarios in mitigation analysis. *Nature Climate Change*, *10*(7), 605–610. https://doi.org/10.1038/s41558-020-0826-9

Grubb, M., Wieners, C. & Yang, P. (2021). Modeling myths: On DICE and dynamic realism in integrated assessment models of climate change mitigation. *WIREs Climate Change*, *12*(3), e698. https://doi.org/https://doi.org/10.1002/wcc.698

Grubler, A., Wilson, C., Bento, N., Boza-Kiss, B., Krey, V., McCollum, D. L., Rao, N. D., Riahi, K., Rogelj, J., De Stercke, S., Cullen, J., Frank, S., Fricko, O., Guo, F., Gidden, M., Havlík, P., Huppmann, D., Kiesewetter, G., Rafaj, P., ... Valin, H. (2018). A low energy demand scenario for meeting the 1.5 °C target and sustainable development goals without negative emission technologies. *Nature Energy*, *3*(6), 515–527. https://doi.org/10.1038/s41560-018-0172-6

Haikola, S., Hansson, A. & Fridahl, M. (2019). Map-makers and navigators of politicised terrain: Expert understandings of epistemological uncertainty in integrated assessment modelling of bioenergy with carbon capture and storage. *Futures*, *114*, 102472. https://doi.org/10.1016/j.futures.2019.102472

Hall, D. C. & Behl, R. J. (2006). Integrating economic analysis and the science of climate instability. *Ecological Economics*, *57*(3), 442–465. https://doi.org/10.1016/j.ecolecon.2005.05.001

Hamilton, S. H., Fu, B., Guillaume, J. H., Badham, J., Elsawah, S., Gober, P., Hunt, R. J., Iwanaga, T., Jakeman, A. J., Ames, D. P., Curtis, A., Hill, M. C., Pierce, S. A. & Zare, F. (2019). A framework for characterising and evaluating the effectiveness of environmental modelling. *Environmental Modelling & Software*, *118*, 83–98. https://doi.org/https://doi.org/10.1016/j.envsoft.2019.04.008

Hänsel, M., Drupp, M. A., Johannson, D., Nesje, F., Azar, C. A., Freeman, M., Groom, B. & Sterner, T. (2020). Climate economics support for the un climate targets. *Nature Climate Change*. https://doi.org/10.1038/s41558-020-0833-x

Harmsen, M., Kriegler, E., van Vuuren, D. P., Wijst, K.-I. v. d., Luderer, G., Cui, R., Dessens, O., Drouet, L., Emmerling, J., Morris, J. F., Fosse, F., Fragkiadakis, D., Fragkiadakis, K., Fragkos, P., Fricko, O., Fujimori, S., Gernaat, D., Guivarch, C., Iyer, G., ... Zakeri, B. (2021). Integrated assessment model diagnostics: Key indicators and model evolution. *Environmental Research Letters*, *16*(5), 054046. https://doi.org/10.1088/1748-9326/abf964

Heal, G. (2017). The economics of the climate. *Journal of Economic Literature*, *55*(3), 1046–63. https://doi.org/10.1257/jel.20151335

Heilmann, C. (2017). Values in time discounting. *Science and Engineering Ethics*, *23*. https://doi.org/10.1007/s11948-017-9950-y

Helm, D. (Ed.). (2005). *Climate change policy*. Oxford: Oxford University Press. https://global.oup.com/academic/product/climate-change-policy-9780199281459?cc=de&lang=en&#

Hepburn, C. & Beckerman, W. (2007). Ethics of the discount rate in the Stern review on the economics of climate change. *World Economics*, *8*, 187–210.

Hof, A. F., de Bruin, K. C., Dellink, R. B., den Elzen, M. G. J. & van Vuuren, D. P. (2009). The effect of different mitigation strategies on international financing of adaptation. *Environmental Science & Policy*, *12*(7), 832–843. https://doi.org/10.1016/j.envsci.2009.08.007

Hof, A. F., den Elzen, M. G. J. & van Vuuren, D. P. (2008). Analysing the costs and benefits of climate policy: Value judgements and scientific uncertainties. *Global Environmental Change*, *18*(3), 412–424. https://doi.org/10.1016/j.gloenvcha.2008.04.004

Hof, A. F., den Elzen, M. G. J. & van Vuuren, D. P. (2010). Including adaptation costs and climate change damages in evaluating post-2012 burden-sharing regimes. *Mitigation and Adaptation Strategies for Global Change*, *15*(1), 19–40. https://doi.org/10.1007/s11027-009-9201-x

Howard, P. H. & Sterner, T. (2017). Few and not so far between: A meta-analysis of climate damage estimates. *Environmental & Resource Economics*, *68*(1, SI), 197–225. https://doi.org/10.1007/s10640-017-0166-z

Huntington, H. G., Weyant, J. P. & Sweeney, J. L. (1982). Modeling for insights, not numbers: The experiences of the energy modeling forum. *Omega*, *10*(5), 449–462. https://doi.org/10.1016/0305-0483(82)90002-0

IPCC. (2014). *Climate change 2014: Mitigation of climate change. contribution of working group III to the fifth assessment report of the intergovernmental panel on climate change* (O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel & J. Minx, Eds.). Cambridge University Press, Cambridge, United Kingdom; New York, NY, USA.

IPCC. (2021). *Climate change 2021: The physical science basis. contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change* (Masson-Delmotte, V. and Zhai, P. and Pirani, A. and Connors, S.L. and Péan, C. and Berger, S. and Caud, N. and Chen, Y. and Goldfarb, L. and Gomis, M.I. and Huang, M. and Leitzell, K. and Lonnoy, E. and Matthews, J.B.R. and Maycock, T.K. and Waterfield, T. and Yelekçi, O. and Yu, R. and Zhou, B., Ed.). Cambridge University Press. In Press.

Ives, M., Righetti, L., Schiele, J., Meyer, K. D., Hubble-Rose, L., Teng, F., Kruitwagen, L., Tillmann-Morris, L., Wang, T., Way, R. & Hepburn, C. (2021). *A new perspective on decarbonising the global energy system* (tech. rep. No. 21-04). Oxford: Smith School of Enterprise and the Environment.

Jafino, B. A., Kwakkel, J. H. & Taebi, B. (2021). Enabling assessment of distributive justice through models for climate change planning: A review of recent advances and a research agenda. *WIREs Climate Change*, *12*(4), e721. https://doi.org/https://doi.org/10.1002/wcc.721

Johnson, J., Baldos, U., Hertel, T., Liu, J., Nootenboom, C., Polasky, S. & Roxburgh, T. (2020). *Global futures: Modelling the global economic impacts of environmental change to support policy-making* (tech. rep.). World Wildlife Fund. Retrieved February 1, 2022, from https://www.wwf.org.uk/globalfutures

Kalkuhl, M. & Wenz, L. (2020). The impact of climate conditions on economic production. Evidence from a global panel of regions. *Journal of Environmental Economics and Management*, *103*, 102360. https://doi.org/10.1016/j.jeem.2020.102360

Kaya, A., Csala, D. & Sgouridis, S. (2017). Constant elasticity of substitution functions for energy modeling in general equilibrium integrated assessment models: A critical review and recommendations. *Climatic Change*, *145*(1), 27–40. https://doi.org/10.1007/s10584-017-2077-y

Keen, S. (2020). The appallingly bad neoclassical economics of climate change. *Globalizations*, *0*(0), 1–29. https://doi.org/10.1080/14747731.2020.1807856

Kelleher, J. P. (2017). Descriptive versus prescriptive discounting in climate change policy analysis. *Georgetown Journal of Law and Public Policy*, *15*, 957–977.

Kelleher, J. P. (2019). Reflections on the 2018 Nobel Memorial Prize Awarded to William Nordhaus. *Erasmus Journal for Philosophy and Economics*, *12*(1), 93–107. https://doi.org/10.23941/ejpe.v12i1.415

Kelleher, J. P. (2022). *Optimization and the tasks of climate change economics* (tech. rep.). University of Wisconsin–Madison. https://www.overleaf.com/read/ycxqhrqsntvs

Keppo, I., Butnar, I., Bauer, N., Caspani, M., Edelenbosch, O., Emmerling, J., Fragkos, P., Guivarch, C., Harmsen, M., Lefèvre, J., Gallic, T. L., Leimbach, M., McDowall, W., Mercure, J.-F., Schaeffer, R., Trutnevyte, E. & Wagner, F. (2021). Exploring the possibility space: Taking stock of the diverse capabilities and gaps in integrated assessment models. *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/abe5d8

Knuuttila, T. (2009). Isolating Representations Versus Credible Constructions? Economic Modelling in Theory and Practice. *Erkenntnis*, *70*(1), 59–80. https://doi.org/10.1007/s10670-008-9137-7

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, *42*(2), 262–271. https://doi.org/10.1016/j.shpsa.2010.11.034

Knuuttila, T. & Morgan, M. S. (2012). Models and modelling in economics. In U. Mäki (Ed.), *Handbook of the philosophy of science. volume 13: Philosophy of economics.* Elsevier BV. https://www.elsevier.com/books/philosophy-of-economics/maki/978-0-444-51676-3

Kowarsch, M. (2016a). Ethics in Climate Economics: Balance or Bias? In M. Kowarsch (Ed.), *A pragmatist orientation for the social sciences in climate policy* (pp. 197–227). Springer International Publishing. https://doi.org/10.1007/978-3-319-43281-6_8

Kowarsch, M. (2016b). Understanding and Evaluating the IAM-Based Economics. In M. Kowarsch (Ed.), *A pragmatist orientation for the social sciences in climate policy* (1st ed., pp. 173–195). Springer International Publishing. https://doi.org/10.1007/978-3-319-43281-6_7

Krey, V., Guo, F., Kolp, P., Zhou, W., Schaeffer, R., Awasthy, A., Bertram, C., de Boer, H.-S., Fragkos, P., Fujimori, S., He, C., Iyer, G., Keramidas, K., Köberle, A. C., Oshiro, K., Reis, L. A., Shoai-Tehrani, B., Vishwanathan, S., Capros, P., . . . van Vuuren, D. P. (2019). Looking under the hood: A comparison of techno-economic assumptions across national and global integrated assessment models. *Energy*, *172*, 1254–1267. https://doi.org/10.1016/j.energy.2018.12.131

Kriegler, E., Bauer, N., Popp, A., Humpenoeder, F., Leimbach, M., Strefler, J., Baumstark, L., Bodirsky, B. L., Hilaire, J., Klein, D., Mouratiadou, I., Weindl, I., Bertram, C., Dietrich, J. .-P., Luderer, G., Pehl, M., Pietzcker, R., Piontek, F., Lotze-Campen, H., . . . Edenhofer, O. (2017). Fossil-fueled development (SSP5): An energy and resource intensive scenario for the 21st century. *Global Environmental Change - Human and Policy Dimensions*, *42*, 297–315. https://doi.org/10.1016/j.gloenvcha.2016.05.015

Kriegler, E., Petermann, N., Krey, V., Schwanitz, V. J., Luderer, G., Ashina, S., Bosetti, V., Eom, J., Kitous, A., Méjean, A., Paroussos, L., Sano, F., Turton, H., Wilson, C. & Van Vuuren, D. P. (2015). Diagnostic indicators for integrated assessment models of climate policy. *Technological Forecasting and Social Change*, *90*, 45–61. https://doi.org/https://doi.org/10.1016/j.techfore.2013.09.020

Kriegler, E., Riahi, K., Bosetti, V., Capros, P., Petermann, N., van Vuuren, D. P., Weyant, J. P. & Edenhofer, O. (2015). Introduction to the AMPERE model intercomparison studies on the economics of climate stabilization. *Technological Forecasting and Social Change*, *90*, 1–7. https://doi.org/10.1016/j.techfore.2014.10.012

Leimbach, M. & Bauer, N. (2021). Capital markets and the costs of climate policies. *Environmental Economics and Policy Studies*. https://doi.org/10.1007/s10018-021-00327-5

Leimbach, M., Bauer, N., Baumstark, L. & Edenhofer, O. (2010). Mitigation Costs in a Globalized World: Climate Policy Analysis with REMIND-R. *Environmental Modeling & Assessment*, *15*(3), 155–173. https://doi.org/10.1007/s10666-009-9204-8

Leimbach, M., Bauer, N., Baumstark, L., Lueken, M. & Edenhofer, O. (2010). Technological change and international trade - insights from REMIND-R. *Energy Journal*, *31*(1), 109–136.

Lemoine, D. & Traeger, C. P. (2016). Economics of tipping the climate dominoes. *Nature Climate Change*, *6*(5), 514–519. https://doi.org/10.1038/nclimate2902

Lenzi, D., Lamb, W. F., Hilaire, J., Kowarsch, M. & Minx, J. C. (2018). Don't deploy negative emissions technologies without ethical analysis. *Nature*, *561*(7723), 303–305. https://doi.org/10.1038/d41586-018-06695-5

Low, S. & Schäfer, S. (2020). Is bio-energy carbon capture and storage (BECCS) feasible? The contested authority of integrated assessment modeling. *Energy Research & Social Science*, *60*, 101326. https://doi.org/10.1016/j.erss.2019.101326

Luderer, G., Bosetti, V., Jakob, M., Leimbach, M., Steckel, J., Waisman, H. & Edenhofer, O. (2012). The economics of decarbonizing the energy system—results and insights from the RECIPE model intercomparison. *Climatic Change*, *114*, 1–29. https://doi.org/10.1007/s10584-011-0105-x

Luderer, G., Leimbach, M., Bauer, N., Kriegler, E., Baumstark, L., Bertram, C., Giannousakis, A., Hilaire, J., Klein, D., Levesque, A., Mouratiadou, I., Pehl, M., Pietzcker, R., Piontek, F., Roming, N., Schultes, A., Schwanitz, V. J. & Strefler, J. (2015). *Description of the REMIND model (version 1.6)*. PIK. https://doi.org/10.2139/ssrn.2697070

Luderer, G., Vrontisi, Z., Bertram, C., Edelenbosch, O. Y., Pietzcker, R. C., Rogelj, J., De Boer, H. S., Drouet, L., Emmerling, J., Fricko, O., Fujimori, S., Havlík, P., Iyer, G., Keramidas, K., Kitous, A., Pehl, M., Krey, V., Riahi, K., Saveyn, B., . . . Kriegler, E. (2018). Residual fossil CO2 emissions in 1.5–2 °C pathways. *Nature Climate Change*, *8*(7), 626–633. https://doi.org/10.1038/s41558-018-0198-6

Mäki, U. (2001). Models: Philosophical aspects. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 9931–9937). Pergamon. https://doi.org/10.1016/B0-08-043076-7/01006-8

Mäki, U. (1992). On the method of isolation in economics, idealization iv: Intelligibility in science, edited by craig dilworth. *Poznan Studies in the Philosophy of the Sciences and the Humanities*.

Mäki, U. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology*, *12*(2), 303–315. https://doi.org/10.1080/13501780500086255

Mäki, U. (2009). MISSing the world. Models as Isolations and credible Surrogate Systems. *Erkenntnis*, *70*(1), 29–43. https://doi.org/10.1007/s10670-008-9135-9

Mäki, U. (2018). Rights and wrongs of economic modelling: Refining Rodrik. *Journal of Economic Methodology*, *25*(3), 218–236. https://doi.org/10.1080/1350178X.2018.1488475

Mäki, U. (2020). Puzzled by Idealizations and Understanding Their Functions. *Philosophy of the Social Sciences*, *50*(3), 215–237. https://doi.org/10.1177/0048393120917637

Mastrandrea, M. D. (2010). Representation of climate impacts in integrated assessment models. *Assessing the Benefits of Avoided Climate Change: Cost-Benefit Analysis and Beyond.*

McCollum, D. L., Zhou, W., Bertram, C., de Boer, H.-S., Bosetti, V., Busch, S., Després, J., Drouet, L., Emmerling, J., Fay, M., Fricko, O., Fujimori, S., Gidden, M., Harmsen, M., Huppmann, D., Iyer, G., Krey, V., Kriegler, E., Nicolas, C., ... Riahi, K. (2018). Energy investment needs for fulfilling the Paris Agreement and achieving the Sustainable Development Goals. *Nature Energy*, *3*(7), 589–599. https://doi.org/10.1038/s41560-018-0179-z

McLaren, D. & Markusson, N. (2020). The co-evolution of technological promises, modelling, policies and climate change targets. *Nature Climate Change*, *10*(5), 392–397. https://doi.org/10.1038/s41558-020-0740-1

Meinshausen, M., Raper, S. C. B. & Wigley, T. M. L. (2011). Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration. *Atmospheric Chemistry and Physics*, *11*(4), 1417–1456. https://doi.org/10.5194/acp-11-1417-2011

Metcalf, G. E. & Stock, J. H. (2017). Integrated Assessment Models and the Social Cost of Carbon: A Review and Assessment of U.S. Experience. *Review of Environmental Economics and Policy*, *11*(1), 80–99. https://doi.org/10.1093/reep/rew014

Millner, A. & McDermott, T. K. J. (2016). Model confirmation in climate economics. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1604121113

Molotoks, A., Stehfest, E., Doelman, J., Albanito, F., Fitton, N., Dawson, T. P. & Smith, P. (2018). Global projections of future cropland expansion to 2050 and direct impacts on biodiversity and carbon storage. *Global Change Biology*, *24*(12), 5895–5908. https://doi.org/10.1111/gcb.14459

Moore, F. C. & Diaz, D. B. (2015). Temperature impacts on economic growth warrant stringent mitigation policy. *Nature Climate Change*, *5*(2), 127–131. https://doi.org/10.1038/nclimate2481

Morgan, M. G. & Dowlatabadi, H. (1996). Learning from integrated assessment of climate change. *Climatic Change*, *34*(3-4), 337–368. https://doi.org/10.1007/bf00139297

Morgan, M. S. (2001). Models, stories and the economic world. *Journal of Economic Methodology*, *8*(3), 361–384. https://doi.org/10.1080/13501780110078972

Nelson, J. A. (2008). Economists, value judgments, and climate change: A view from feminist economics. *Ecological Economics*, *65*(3), 441–447. https://doi.org/10.1016/j.ecolecon.2008.01.001

Neumayer, E. (1999). Global warming: Discounting is not the issue, but substitutability is. *Energy Policy*, *27*(1), 33–43. https://doi.org/10.1016/S0301-4215(98)00063-9

Nikas, A., Doukas, H. & Papandreou, A. (2019). A detailed overview and consistent classification of climate-economy models. In H. Doukas, A. Flamos & J. Lieu (Eds.). Springer International Publishing. https://doi.org/10.1007/978-3-030-03152-7_1

Nobel Prize Outreach. (2022). William d. nordhaus - facts - 2018. Retrieved January 30, 2022, from https://www.nobelprize.org/prizes/economic-sciences/2018/nordhaus/facts/

Nordhaus, W. (1992). An optimal transition path for controlling greenhouse gases. *Science*, *258*(5086), 1315–1319. https://doi.org/10.1126/science.258.5086.1315

Nordhaus, W. (1993). Rolling the 'DICE': An optimal transition path for controlling greenhouse gases. *Resource and Energy Economics*, *15*(1), 27–50. https://doi.org/10.1016/0928-7655(93)90017-o

Nordhaus, W. (2002). Modeling induced innovation in climate-change policy. *Technological change and the environment.* Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781936331758-10/modeling-induced-innovation-climate-change-policy-william-nordhaus

Nordhaus, W. (2009). *An analysis of the dismal theorem* (Cowles Foundation Discussion Papers No. 1686). Cowles Foundation for Research in Economics, Yale University.

Nordhaus, W. (2010). Economic aspects of global warming in a post-Copenhagen environment. *Proceedings of the National Academy of Sciences*, *107*(26), 11721–11726. https://doi.org/10.1073/pnas.1005985107

Nordhaus, W. (2011). *On the concept of optimal economic growth* (Cowles Foundation Discussion Papers No. 1839). Cowles Foundation for Research in Economics, Yale University. https://ideas.repec.org/h/eee/hacchp/v1y2013icp1069-1131.html

Nordhaus, W. (2013). Integrated economic and climate modeling. *Handbook of computable general equilibrium modeling* (pp. 1069–1131). Elsevier. https://doi.org/10.1016/B978-0-444-59568-3.00016-X

Nordhaus, W. (2014). Estimates of the Social Cost of Carbon: Concepts and Results from the DICE-2013R Model and Alternative Approaches. *Journal of the Association of Environmental and Resource Economists*, *1*(1/2), 273–312. https://doi.org/10.1086/676035

Nordhaus, W. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1609244114

Nordhaus, W. (2018a). Evolution of modeling of the economics of global warming: Changes in the DICE model, 1992–2017. *Climatic Change*, *148*. https://doi.org/10.1007/s10584-018-2218-y

Nordhaus, W. (2018b). Projections and uncertainties about climate change in an era of minimal climate policies. *American Economic Journal: Economic Policy*, *10*(3), 333–360. https://doi.org/10.1257/pol.20170046

Nordhaus, W. (2019). Climate change: The ultimate challenge for economics. *American Economic Review*, *109*(6), 1991–2014. https://doi.org/10.1257/aer.109.6.1991

Nordhaus, W. & Moffat, A. (2017). *A Survey of Global Impacts of Climate Change: Replication, Survey Methods, and a Statistical Analysis* (Working Paper No. 23646). National Bureau of Economic Research. https://doi.org/10.3386/w23646

Nordhaus, W. & Sztorc, P. (2013). *DICE 2013R: Introduction and User's Manual* (tech. rep.).

OECD. (2012). *OECD environmental outlook to 2050*. https://doi.org/https://doi.org/https://doi.org/10.1787/9789264122246-en

Parker, W. S. (2009). II—confirmation and adequacy-for-purpose in climate modelling. *Aristotelian Society Supplementary Volume*, *83*(1), 233–249. https://doi.org/10.1111/j.1467-8349.2009.00180.x

Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science*, *87*(3), 457–477. https://doi.org/10.1086/708691

Parson, E. A. (1996). Three dilemmas in the integrated assessment of climate change. *Climatic Change*, *34*(3-4), 315–326. https://doi.org/10.1007/bf00139295

Parson, E. A. & Fisher-Vanden, K. (1997). Integrated assessment models of global climate change. *Annual Review of Energy and the Environment*, *22*(1), 589–628. https://doi.org/10.1146/annurev.energy.22.1.589

Pauliuk, S., Arvesen, A., Stadler, K. & Hertwich, E. G. (2017). Industrial ecology in integrated assessment models. *Nature Climate Change*, *7*(1), 13–20. https://doi.org/10.1038/nclimate3148

PBL. (2005). *The Targets IMage Energy Regional (TIMER) Model. Technical Documentation*. Retrieved January 27, 2022, from https://www.pbl.nl/sites/default/files/downloads/461502024.pdf

Peace, J. & Weyant, J. (2008). *Insights not numbers: The appropriate use of economic models* (tech. rep.). Pew Center on Global Climate Change.

Perman, R., Ma, Y., McGilvray, J. & Common, M. (2003). *Natural resource and environmental economics* (3rd ed.). Pearson Education Limited.

Pezzey, J. C. V. (2019). Why the social cost of carbon will always be disputed. *WIREs Climate Change*, *10*(1), e558. https://doi.org/10.1002/wcc.558

Pielke, R. & Ritchie, J. (2021). Distorting the view of our climate future: The misuse and abuse of climate pathways and scenarios. *Energy Research & Social Science*, *72*, 101890. https://doi.org/10.1016/j.erss.2020.101890

PIK. (2022). REMIND website. Retrieved January 26, 2022, from https://www.pik-potsdam.de/en/institute/departments/transformation-pathways/models/remind

Pindyck, R. S. (2013a). Climate change policy: What do the models tell us? *Journal of Economic Literature*, *51*(3), 860–72. https://doi.org/10.1257/jel.51.3.860

Pindyck, R. S. (2013b). The climate policy dilemma. *Review of Environmental Economics and Policy*, *7*(2), 219–237. https://doi.org/10.1093/reep/ret007

Pindyck, R. S. (2017). The use and misuse of models for climate policy. *Review of Environmental Economics and Policy*, *11*(1), 100–114. https://doi.org/10.1093/reep/rew012

Piontek, F., Drouet, L., Emmerling, J., Kompas, T., Méjean, A., Otto, C., Rising, J., Soergel, B., Taconet, N. & Tavoni, M. (2021). Integrated perspective on translating biophysical to economic impacts of climate change. *Nature Climate Change*, *11*(7), 563–572. https://doi.org/10.1038/s41558-021-01065-y

Popp, D. (2004). ENTICE: Endogenous technological change in the DICE model of global warming. *Journal of Environmental Economics and Management*, *48*(1), 742–768. https://doi.org/10.1016/j.jeem.2003.09.002

Ravetz, J. R. (1997). *Integrated environmental assessment forum: Developing guidelines for"good practice"* (WP No. 97-1). ULYSSES. https://www.jvds.nl/ulysses/eWP97-1.pdf

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., ... Tavoni, M. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2016.05.009

Risbey, J., Kandlikar, M. & Patwardhan, A. (1996). Assessing integrated assessments. *Climatic Change*, *34*, 369–395. https://doi.org/10.1007/BF00139298

Risbey, J., van der Sluijs, J., Kloprogge, P., Ravetz, J., Funtowicz, S. & Corral Quintana, S. (2005). Application of a checklist for quality assistance in environmental modelling to an energy model. *Environmental Modeling & Assessment*, *10*(1), 63–79. https://doi.org/10.1007/s10666-004-4267-z

Robertson, S. (2020). Transparency, trust, and integrated assessment models: An ethical consideration for the intergovernmental panel on climate change. *WIREs Climate Change*, *12*(1). https://doi.org/10.1002/wcc.679

Roe, G. H. & Baker, M. B. (2007). Why Is Climate Sensitivity So Unpredictable? *Science*. https://doi.org/10.1126/science.1144735

Roe, G. H. & Bauman, Y. (2013). Climate sensitivity: Should the climate tail wag the policy dog? *Climatic Change*, *117*(4), 647–662. https://doi.org/10.1007/s10584-012-0582-6

Rosen, R. A. & Guenther, E. (2015). The economics of mitigating climate change: What can we know? *Technological Forecasting and Social Change*, *91*, 93–106. https://doi.org/10.1016/j.techfore.2014.01.013

Rotmans, J. (1990). *IMAGE: An integrated model to assess the greenhouse effect*. Springer Netherlands. https://doi.org/10.1007/978-94-009-0691-4

Rotmans, J. & Asselt, M. V. (1996). Integrated assessment: A growing child on its way to maturity. *Climatic Change*, *34*(3-4), 327–336. https://doi.org/10.1007/bf00139296

Sanstad, A. H. & Greening, L. A. (1998). Economic models for climate policy analysis: A critical discussion. *Environmental Modeling & Assessment*, *3*(1), 3–18. https://doi.org/10.1023/A:1019002620369

Schultes, A., Piontek, F., Soergel, B., Rogelj, J., Baumstark, L., Kriegler, E., Edenhofer, O. & Luderer, G. (2021). Economic damages from on-going climate change imply deeper near-term emission cuts. *Environmental Research Letters*, *16*(10), 104053. https://doi.org/10.1088/1748-9326/ac27ce

Schwanitz, V. J. (2013). Evaluating integrated assessment models of global climate change. *Environmental Modelling & Software*, *50*, 120–131. https://doi.org/https://doi.org/10.1016/j.envsoft.2013.09.005

Skea, J., Shukla, P., Khourdajie, A. A. & McCollum, D. (2021). Intergovernmental Panel on Climate Change: Transparency and integrated assessment modeling. *WIREs Climate Change*, *12*(5), e727. https://doi.org/10.1002/wcc.727

Soergel, B., Kriegler, E., Bodirsky, B. L., Bauer, N., Leimbach, M. & Popp, A. (2021). Combining ambitious climate policies with efforts to eradicate poverty. *Nature Communications*, *12*(1), 2342. https://doi.org/10.1038/s41467-021-22315-9

Stanton, E. A. (2011). Negishi welfare weights in integrated assessment models: The mathematics of global inequality. *Climatic Change*, *107*(3), 417–432. https://doi.org/10.1007/s10584-010-9967-6

Staub-Kaminski, I., Zimmer, A., Jakob, M. & Marschinski, R. (2014). Climate policy in practice: A typology of obstacles and implications for integrated assessment modeling. *Climate Change Economics*, *5*(01), 1–30. https://doi.org/10.1142/S2010007814400041

Stehfest, E., van Vuuren, D. P., Kram, T. & Bouwman, L. (2014). *Integrated assessment of global environmental change with IMAGE 3.0. model description and policy applications*. The Hague: PBL Netherlands Environmental Assessment Agency. https://www.pbl.nl/en/publications/integrated-assessment-of-global-environmental-change-with-IMAGE-3.0

Stern, N. (2008). The economics of climate change. *American Economic Review*, *98*(2), 1–37. https://doi.org/10.1257/aer.98.2.1

Stern, N. (2013). The Structure of Economic Modeling of the Potential Impacts of Climate Change: Grafting Gross Underestimation of Risk onto Already Narrow Science Models. *Journal of Economic Literature*, *51*(3), 838–859. https://doi.org/10.1257/jel.51.3.838

Stern, N. (2016). Economics: Current climate models are grossly misleading. *Nature*, *530*(7591), 407–409. https://doi.org/10.1038/530407a

Sterner, T. & Persson, U. M. (2008). An even sterner review: Introducing relative prices into the discounting debate. *Review of Environmental Economics and Policy*, *2*(1), 61–76. https://doi.org/10.1093/reep/rem024

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, *7*(1), 1–31. https://doi.org/10.1080/135017800362220

Tavoni, M., Kriegler, E., Riahi, K., van Vuuren, D. P., Aboumahboub, T., Bowen, A., Calvin, K., Campiglio, E., Kober, T., Jewell, J., Luderer, G., Marangoni, G., McCollum, D., van Sluisveld, M., Zimmer, A. & van der Zwaan, B. (2015). Post-2020 climate agreements in the major economies assessed in the light of global models. *Nature Climate Change*, *5*(2), 119–126. https://doi.org/10.1038/nclimate2475

Tol, R. S. (2006). *Integrated Assessment Modelling* (Working Papers FNU-102). Research unit Sustainability and Global Change, Hamburg University. https://ideas.repec.org/p/sgc/wpaper/102.html

Truong, T. P. (2009). *Constant elasticity of substitution (CES) production function can greatly overestimate the economic costs of climate policies.* (tech. rep.). Institute of Transport and Logistics Studies. http://hdl.handle.net/2123/19477

Trutnevyte, E. (2016). Does cost optimization approximate the real-world energy transition? *Energy*, *106*, 182–193. https://doi.org/10.1016/j.energy.2016.03.038

van Beek, L., Hajer, M., Pelzer, P., van Vuuren, D. & Cassen, C. (2020). Anticipating futures through models: the rise of Integrated Assessment Modelling in the climate science-policy interface since 1970. *Global Environmental Change*, *65*, 102191. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2020.102191

Van der Sluijs, J., Risbey, J., Quintana, S. C. & Ravetz, J. (2002). Uncertainty management in complex models: The NUSAP method. *International Congress on Environmental Modelling and Software*. https://scholarsarchive.byu.edu/iemssconference/2002/all/240

van der Sluijs, J. P. (2002). A way out of the credibility crisis of models used in integrated environmental assessment. *Futures*, *34*(2), 133–146. https://doi.org/10.1016/S0016-3287(01)00051-9

van der Wijst, K.-I., Hof, A. F. & van Vuuren, D. P. (2021). On the optimality of 2°C targets and a decomposition of uncertainty. *Nature Communications*, *12*(1), 2575. https://doi.org/10.1038/s41467-021-22826-5

van Ruijven, B., van der Sluijs, J. P., van Vuuren, D. P., Janssen, P., Heuberger, P. S. C. & de Vries, B. (2010). Uncertainty from Model Calibration: Applying a New Method to Transport Energy Demand Modelling. *Environmental Modeling & Assessment*, *15*(3), 175–188. https://doi.org/10.1007/s10666-009-9200-z van Vuuren, D. P., de Vries, B., Beusen, A. & Heuberger, P. S. C. (2008). Conditional probabilistic estimates of 21st century greenhouse gas emissions based on the storylines of the IPCC-SRES scenarios. *Global Environmental Change*, *18*(4), 635–654. https://doi.org/10.1016/j.gloenvcha.2008.06.001

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J. & Rose, S. K. (2011). The representative concentration pathways: An overview. *Climatic Change*, *109*(1-2), 5–31. https://doi.org/10.1007/s10584-011-0148-z

van Vuuren, D. P., Isaac, M., Kundzewicz, Z. W., Arnell, N., Barker, T., Criqui, P., Berkhout, F., Hilderink, H., Hinkel, J., Hof, A., Kitous, A., Kram, T., Mechler, R. & Scrieciu, S. (2011). The use of scenarios as the basis for combined assessment of climate change mitigation and adaptation. *Global Environmental Change*, *21*(2), 575–591. https://doi.org/10.1016/j.gloenvcha.2010.11.003

van Vuuren, D. P., Kok, M., Lucas, P. L., Prins, A. G., Alkemade, R., van den Berg, M., Bouwrnan, L., van der Esch, S., Jeuken, M., Kram, T. & Stehfest, E. (2015). Pathways to achieve a set of ambitious global sustainability objectives by 2050: Explorations using the IMAGE integrated assessment model. *Technological Forecasting and Social Change*, *98*, 303–323. https://doi.org/10.1016/j.techfore.2015.03.005

van Vuuren, D. P., Riahi, K., Calvin, K., Dellink, R., Emmerling, J., Fujimori, S., Kc, S., Kriegler, E. & O'Neill, B. (2017). The Shared Socio-economic Pathways: Trajectories for human development and global environmental change. *Global Environmental Change*, *42*, 148–152. https://doi.org/10.1016/j.gloenvcha.2016.10.009

van Vuuren, D. P., Stehfest, E., Gernaat, D., Boer, H.-S. d., Daioglou, V., Doelman, J., Edelenbosch, O., Harmsen, M., Zeist, W.-J. v., Berg, M. v. d., Dafnomilis, I., Sluisveld, M. v., Tabeau, A., Vos, L. D., Waal, L. d., Berg, N. v. d., Beusen, A., Bos, A., Biemans, H., … Castillo, V. Z. (2021). The 2021 SSP scenarios of the IMAGE 3.2 model. Retrieved January 13, 2022, from https://eartharxiv.org/repository/view/2759/

Vrontisi, Z., Luderer, G., Saveyn, B., Keramidas, K., Lara, A. R., Baumstark, L., Bertram, C., Boer, H. S. d., Drouet, L., Fragkiadakis, K., Fricko, O., Fujimori, S., Guivarch, C., Kitous, A., Krey, V., Kriegler, E., Broin, E. Ó., Paroussos, L. & Vuuren, D. v. (2018). Enhancing global climate policy ambition towards a 1.5°C stabilization: A short-term multi-model assessment. *Environmental Research Letters*, *13*(4), 044039. https://doi.org/10.1088/1748-9326/aab53e

Way, R., Ives, M., Mealy, P. & Farmer, J. D. (2021). *Empirically grounded technology forecasts and the energy transition* (tech. rep. No. 2021-01). INET Oxford Working Paper. Retrieved November 19, 2021, from https://www.inet.ox.ac.uk/publications/no-2021-01-empirically-grounded-technology-forecasts-and-the-energy-transition/

Weitzman, M. L. (2009a). Additive damages, fat-tailed climate dynamics, and uncertain discounting. *Economics: The Open-Access, Open-Assessment E-Journal*, *3*(2009-39), 1. https://doi.org/10.5018/economics-ejournal.ja.2009-39

Weitzman, M. L. (2009b). On modeling and interpreting the economics of catastrophic climate change. *Review of Economics and Statistics*, *91*(1), 1–19.

Weitzman, M. L. (2010). What is the "damages function" for global warming — and what difference might it make? *Climate Change Economics*, *01*(01), 57–69. https://doi.org/10.1142/S2010007810000042

Weitzman, M. L. (2012). GHG targets as insurance against catastrophic climate damages. *Journal of Public Economic Theory*, *14*(2), 221–244. https://doi.org/10.1111/j.1467-9779.2011.01539.x

Weyant, J. P. (2009). A perspective on integrated assessment. *Climatic Change*, *95*(3), 317–323. https://doi.org/10.1007/s10584-009-9612-4

Weyant, J. P. (2014). Integrated assessment of climate change: State of the literature. *Journal of Benefit-Cost Analysis*, *5*(3), 377–409. https://doi.org/10.1515/jbca-2014-9002

Weyant, J. P. (2017). Some contributions of integrated assessment models of global climate change. *Review of Environmental Economics and Policy*, *11*(1), 115–137. https://doi.org/10.1093/reep/rew018

Wilson, C., Grubler, A., Bauer, N., Krey, V. & Riahi, K. (2013). Future capacity growth of energy technologies: Are scenarios consistent with historical evidence? *Climatic Change*, *118*(2), 381–395. https://doi.org/10.1007/s10584-012-0618-y

Wilson, C., Kriegler, E., van Vuuren, D., Guivarch, C., Frame, D., Krey, V., Osborn, T., Schwanitz, V. & Thompson, E. (2017). *Evaluating process-based integrated assessment models of climate change mitigation* (IIASA Working Paper). IIASA, Laxenburg, Austria. IIASA, Laxenburg, Austria, WP-17-007. http://pure.iiasa.ac.at/id/eprint/14502/

Wilson, C., Guivarch, C., Kriegler, E., van Ruijven, B., van Vuuren, D. P., Krey, V., Schwanitz, V. J. & Thompson, E. L. (2021). Evaluating process-based integrated assessment models of climate change mitigation. *Climatic Change*, *166*(1), 3. https://doi.org/10.1007/s10584-021-03099-9

Workman, M., Darch, G., Dooley, K., Lomax, G., Maltby, J. & Pollitt, H. (2021). Climate policy decision making in contexts of deep uncertainty - from optimisation to robustness. *Environmental Science & Policy*, *120*, 127–137. https://doi.org/https://doi.org/10.1016/j.envsci.2021.03.002

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit im Studiengang M.Sc. Integrated Climate System Sciences selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Arbeit vorher nicht in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Einer Veröffentlichung der vorliegenden Arbeit in der zuständigen Fachbibliothek des Fachbereichs stimme ich zu.

Hamburg, den 7. Februar 2022