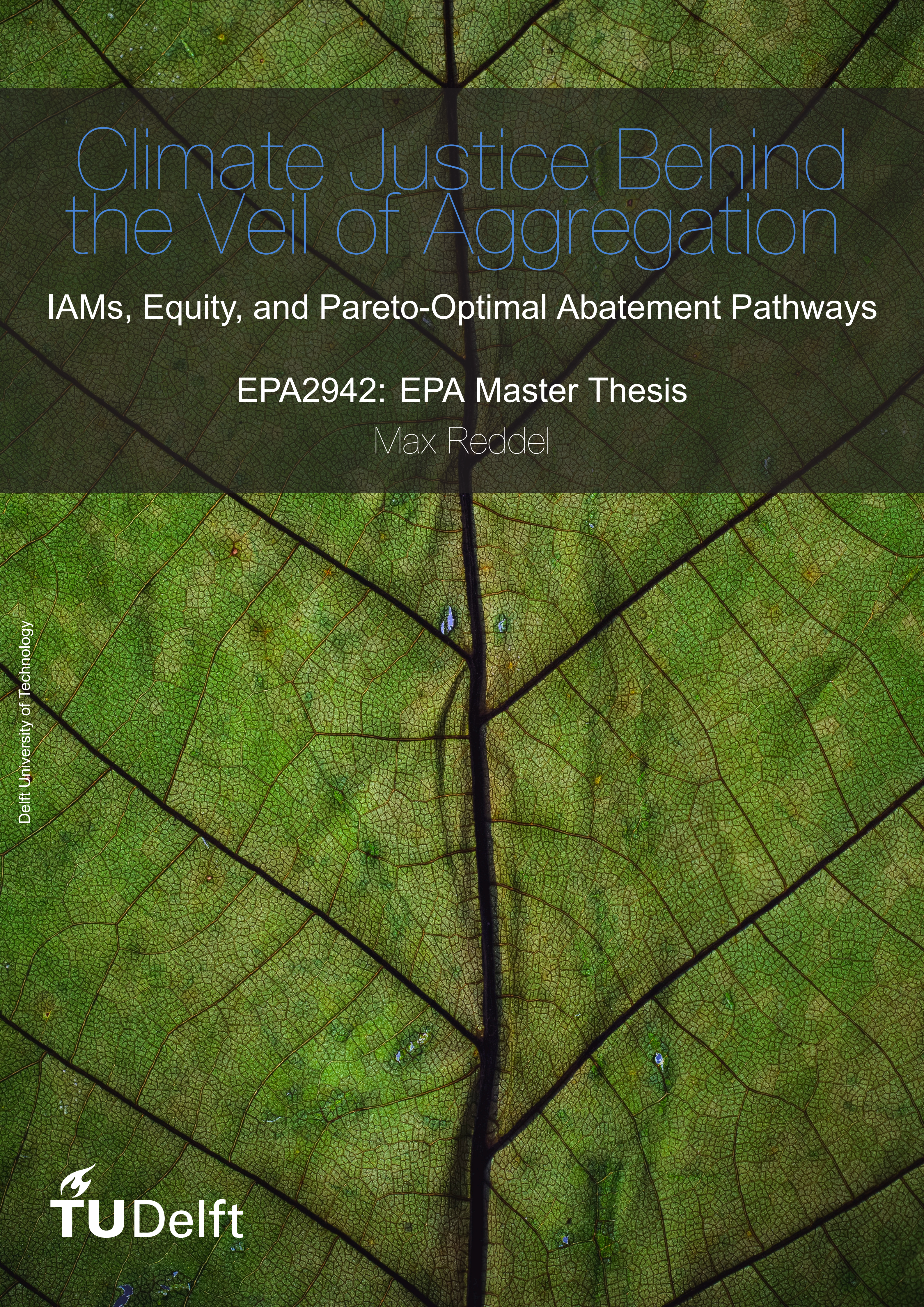


Climate Justice Behind the Veil of Aggregation

IAMs, Equity, and Pareto-Optimal Abatement Pathways

EPA2942: EPA Master Thesis

Max Reddel

Delft University of Technology

Climate Justice Behind the Veil of Aggregation

IAMs, Equity, and Pareto-Optimal Abatement Pathways

by

Max Reddel

Student Number:	5360277

Chair:	Prof. Dr. Ir. Jan Kwakkel
First supervisor:	Dr. Jazmin Zatarain Salazar
Second supervisor:	Dr. S.T.H. (Servaas) Storm

Institution:	Delft University of Technology
Faculty:	Faculty of Technology, Policy, and Management (TPM)
Degree:	M.Sc. Engineering and Policy Analysis
Specialization:	Modeling, Simulation and Gaming
Project Duration:	Feb, 2022 - August, 2022
Date of Defense:	30 August 2022

The Associated code is available at
https://github.com/max-reddel/PyRICE_2022

TU Delft

Preface

In hindsight, I am happy that I have chosen the master's program *Engineering and Policy Analysis* at the Delft University of Technology. I have sincerely enjoyed my study experience and the acquisition of modeling and simulation skills. This thesis concludes my formal education and provided me with the opportunity to apply these skills on the topic of climate change models and distributive justice.

This project's origin predates the start of the actual thesis. Around one year ago, Jazmin took me under her wing by enabling me to work on an internship at the HIPPO lab. Former students have worked on predecessor versions of the PyRICE model (Kulkarni, 2020; Lingeswaran, 2019; Tjallingii, 2021). Substantially refactoring this model allowed me to dive deep into Integrated Assessment Models and get a foretaste of how working on such a project would be like. This experience gave me also the opportunity to connect to the members of the HIPPO lab and collaborate with them throughout the last academic year. At the end of my internship, it became clear that I wanted to continue working on this topic in the context of my master thesis.

Several courses have prepared me for this project. A notable example is *Macro-economic for Policy Analysis* which was managed by Servaas. This course gave me a first introduction to Integrated Assessment Models in the context of its particular macro-economic theory. I also thank him for his guidance and feedback as my second supervisor. Moreover, the likely most important course that I have taken at TU Delft is *Model-based Decision-making* which was organized by Jan and Jazmin. This course equipped me with a fascination with decision-making under deep uncertainty and the corresponding skills.

Jazmin and Jan have supervised me throughout the course, the internship, and this thesis project. I am deeply thankful for all their time, guidance, and feedback. They have even encouraged me to submit an abstract to the conference *International Environmental Modelling and Software Society* (iEMSs 2022) which has been accepted and at which I have given my first conference presentation. Without their help, I would have not been able to recognize nor seize that opportunity. Finally, I would like to express special gratitude to my wife Felicity who enriched my research experience with countless discussions and valuable feedback throughout this time.

Max Reddel
The Hague, August 2022

Summary

Humanity faces the unprecedented global challenge of climate change. The sheer complexity and uncertainty of this problem renders mere intuitive reasoning insufficient. To aid global climate negotiations, Integrated Assessment Models (IAMs) are used to analyze the interplay between climate and the economy. More specifically, IAMs account for how greenhouse gas emissions affect climate change, how climate change affects economic production, and how economic production affects GHG emissions. We can use IAMs to project trends in emissions and gross domestic product, assess the costs and benefits of climate policies, and estimate the social carbon cost required to achieve stated emissions reduction targets. Although IAMs are central to informing decision-making to avoid catastrophic consequences, policy recommendations resulting from IAMs commonly prompt a very heterogeneous distribution of risks and benefits across the globe. During the recent 2021 United Nations Climate Change Conference (COP26), it became clear that equity is a central issue in the climate action debate. Emerging economies consider currently suggested abatement policies unjust in light of the historical CO_2 generation of high-income countries and the strongly increasing need for energy in low-income countries. The term *double inequality* has been coined to describe the inverse relationship between the distributions of risks and responsibilities. In fact, the regions that are the least responsible for historical and mostly current CO_2 emissions around the world, exhibit the highest degree of vulnerability to climate damages. In order to enable international cooperation and have a shot at meeting the Paris Agreement target, we require policies that promote more equitable mitigation pathways. Equity is therefore an eminently pressing topic, yet most IAM studies largely neglect it due to the implicit use of a utilitarian social welfare function that aggregates risks and benefits over space and time, thus losing sight of distributional consequences.

In order to account for distributional justice, we transform the RICE model into a simulation model and embed it in a many-objective simulation-optimization setup such that we can find Pareto-optimal climate mitigation pathways for different problem formulations. Next to using four ethical premises (rooted in utilitarianism, sufficientarianism, egalitarianism, and prioritarianism), we also direct particular attention to the disaggregation of utility and disutility within each of these ethical premises. The reason for this disaggregation is based on the incommensurability of these two. Usually, IAMs maximize aggregate variables such as welfare. If we also consider the minimization of welfare loss, which is based on economic damages as one of the objectives, we can enable a potentially fairer distribution of not only consumption but economic damages. We argue that we can find climate justice behind the veil of aggregation. What we mean by this is that more equitable policy recommendations are obscured and lie hidden behind a bulwark of highly aggregated variables. If we look beyond this obstruction by the means of disaggregation, we are better equipped to find climate justice. In order to get to the bottom of this, we ask the following question:

Research Question

How are Pareto-optimal climate abatement pathways affected by the disaggregation of utility and disutility in alternative ethical problem formulations when using an integrated assessment model under deep uncertainty?

To answer this question, we use a framework that is called *multi-scenario multi-objective robust decision-making*. For each of the eight problem formulations (4 ethical premises x 2 levels of aggregation), we use a multi-objective evolutionary algorithm to find Pareto-optimal policies. We reevaluate their performances under uncertainty by comparing their climate abatement pathways across the problem formulations. On a high-level, we can summarize our key findings as:

1. dominance of aggregation levels over ethical premises
2. correlation between low welfare and high welfare loss
3. general dominance of egalitarian aggregated Pareto-optimal policies
4. shared misery via egalitarian disaggregated Pareto-optimal policies

The effect of disaggregating utility and disutility is stronger than originally expected. Using disaggregated problem formulations yields substantially different pathways even within the same ethical premise. These results are promising as we could transfer these insights to other more complex IAMs such as IMAGE and MESSAGE. Overall, this could be also good news for the equity debate. Using alternative ethical premises and disaggregating incommensurate objectives such as utility and disutility can offer alternative policy recommendations and resulting climate abatement pathways which could in turn enable more equity. What we likely need now, is a stronger dialogue between the modelers and policy analysts on the one side and the stakeholders and decision-makers on the other side. The latter ones should not just blindly trust in the *magical* outputs of a model but they need to be involved to decide what problem formulations we need to use as there is no correct way to frame a complex real-world problem. As unmitigated climate damages exhibit an independent impact on a region's well-being, we could render IAMs more useful for climate policy if we a) acknowledge that the classical notion of welfare is obsolete, b) use a multi-objective approach, and c) let the decision-makers decide how they want to trade-off the various objectives in post. In this manner, we could use IAMs to advance into the direction of enabling a transition of more climate justice.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
CID	Complexity-Invariant Distant
DICE	Dynamic Integrated Climate-Economy
DMDU	Decision-Making Under Deep Uncertainty
GHG	Greenhouse Gas Emissions
GWP	Gross World Product
IAM	Integrated Assessment Model
KPI	Key Performance Indicator
IPCC	Intergovernmental Panel on Climate Change
MOEA	Many-Objective Evolutionary Algorithm
MORDM	Many-Objective Robust Decision-Making
MS-MORDM	Multi-Scenario MORDM
RICE	Regional Integrated Climate and Economy

Problem Formulations

Symbol	Definition
E_A	Egalitarian Aggregated
E_D	Egalitarian Disaggregated
P_A	Prioritarian Aggregated
P_D	Prioritarian Disaggregated
S_A	Sufficientarian Aggregated
S_D	Sufficientarian Disaggregated
U_A	Utilitarian Aggregated
U_D	Utilitarian Disaggregated

Other Symbols

Symbol	Definition
$\mathcal{PF}$	Set of Problem Formulations
$\mathcal{E}$	Set of Ethical Premises
$\mathcal{A}$	Set of Aggregation Levels

List of Figures

List of Tables

1

Introduction

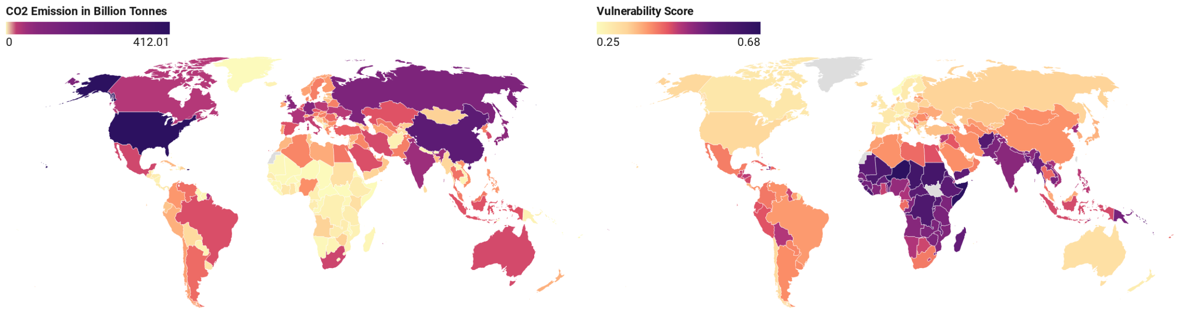
1.1. Context

1.1.1. Climate Change

Humanity faces the unprecedented global challenge of climate change. Climate change exhibits an already perceivable impact. According to the World Meteorological Organization, the year 2021 has measured a mean global temperature which is 1.11°C above the pre-industrial levels (WMO, 2022). Furthermore, 2021 has been the seventh consecutive year in which global atmospheric temperature has been more than 1°C above pre-industrial levels (WMO, 2022). These changes have brought more extreme precipitation patterns causing record droughts especially in Africa and Australia and record flooding especially in China and Argentina (Winsberg, 2018, p. 2). Next to risks to the natural environment, humans are affected in many sectors such as agriculture, forestry, water management, energy, tourism, and health. According to the European Commission, there have been direct economic losses of at least €90 billion between 1980 and 2011 (European Commission, 2022). A look at current developments suggests future exacerbation. A rise of 2.7°C is considered conceivable by the end of the 21st century (Masson-Delmotte et al., 2021), but even smaller temperature increases can already bring about catastrophic repercussions. In the long-term (up to 2100), the IPCC expects risks with a high degree of confidence on issues such as a loss in biodiversity, grand damages to ecosystems, availability of water resources, extraordinarily high economic damages, and the loss of many human lives (IPCC, 2022, pp. 16-17). The impact of climate change depends strongly on a plethora of factors – some that we can control through policy implementation and some factors are external that come with high degrees of uncertainty. According to Bostrom and Cirkovic (2011), the severity of a risk can be characterized by three variables: (a) scope, (b) intensity, and (c) probability. In the case of climate change, the scope is a global and trans-generational one, while the intensity might be very high. The probability of climate change varies as described above depending on myriad of factors. In the same book, Frame and Allen conclude their analysis by carefully stating that climate change might be a dangerous or even catastrophic global risk (Bostrom and Cirkovic, 2011, p. 282). While the stakes are high, solving the problem of climate change is not a simple one. On the contrary, climate change turns out to be a *super wicked problem* as for example our time is running out and there is no central authority who would address this issue (Levin et al., 2012). Global collaboration is the only means we have and this turns out to be a non-trivial endeavor. Mitigating the impacts of climate change is an international grand challenge, to say the least.

1.1.2. Equity

And as if this situation would not be enough, we also face a massive problem with respect to equity. Although, great strides have been made in the last decades to fight global extreme poverty (Roser and Ortiz-Ospina, 2013), climate change might enable a regression to a state of more extreme poverty (Chancel, 2020; Masson-Delmotte et al., 2018). During the very recent 2021 United Nations Climate Change Conference (COP26), it became clearer and clearer that equity is a central issue in the current climate action debate. Emerging economies consider currently suggested abatement policies as unjust due to historical CO_2 generation of high-income countries and the strongly increasing need for energy in low-income countries (Bordoff, 2021; Farand, 2021; Mercado, 2021). The literature speaks of a so-called *double inequality* due to an inverse relationship between the distribution of risks and responsibilities (Barrett, 2013). Figure 1.1a depicts the distribution of CO_2 emissions around the world. This distribution is strongly contrasted with the distribution of vulnerability to climate impact in Figure 1.1b. One could get the idea that the colors in one map are the inversion of the colors in the other map.



(a) Distribution of Emissions in the World, 2019

CO_2 emissions produced from fossil fuels and cement since 1750, and measured in Billion Tonnes. (Data Source: Ritchie et al., 2020; Picture Source)

(b) Distribution of Climate Vulnerability, 2019

Vulnerability measures a country's exposure, sensitivity and ability to adapt to the negative impact of climate change. (Data Source: Chen et al., 2015; Picture Source)

Figure 1.1: Double Inequality: Emissions and Vulnerability

Inequality exacerbates as the world's wealthiest 10% are responsible for 50% of global emissions (Gore, 2015). Furthermore, the physical effects of climate change are expected to be heterogeneously distributed in the world (Green, 2016). Some regions and sub-populations will additionally be more vulnerable to the climate impacts as for example in case of severely increasing events such as droughts, flooding, and heatwaves, poorer regions run risk to become even worse-off as they are economically ill-equipped to adjust to these circumstances (Füssel, 2010; Masson-Delmotte et al., 2018; Thomas et al., 2019). In fact, 20 out of the 36 highest-emission countries are least vulnerable while 11 out of the 17 lowest-emission countries are most vulnerable to the climate change impact (Althor et al., 2016). While mitigation and adaptation policies will likely exhibit heterogeneous effects on different sub-populations, the distribution of policy costs and their benefits are another potential source of injustice which might exacerbate the situation of the worst-off (Atteridge and Remling, 2018). In order to enable international cooperation and have a shot at meeting the Paris Agreement target, we require policies that promote more equitable mitigation pathways (Mi et al., 2019). And while this problem of equity is receiving some traction in the current debate, appropriate climate policies are currently still lacking (Okereke and Coventry, 2016).

1.1.3. Integrated Assessment Models

This chapter is dedicated to IAMs, their criticism, and why we focus on the Regional Integrated Climate-Economy model (RICE). The name *Integrated Assessment Model* reveals already a lot about itself. IAMs combine the science of climate change with the economics and policies addressing GHG emissions. We can use IAMs to investigate the interaction between the climate and the global economy within a single modeling framework. We use IAMs to explore how GHG emissions affect climate change, how climate change effects economic production, and how economic production affects GHG emissions. These models can be used to project trends in emissions and GDP while basing these trends on climate change and abatement policies. They can help to assess the costs and benefits of various climate policies and to estimate the carbon price that would be required to attain particular emission reduction targets.

So, what are the components of an IAM? Figure 1.2 depicts the basic components and their interactions. At the top, we can see that economic output is computed which can be split into savings and consumption. Thus, a part of the economic output is consumed by the world population and another part is saved and will fed back into the economic calculus of the next time step. In the meantime, the economic output brings along GHG emissions. Radiative forcing affects the atmospheric temperature which has direct impacts on crops, forests, ecosystems, and more. These impacts are eventually translated into economic damages which will feed back into the economic output calculation of the next time step. This process repeats for a particular number of time steps.

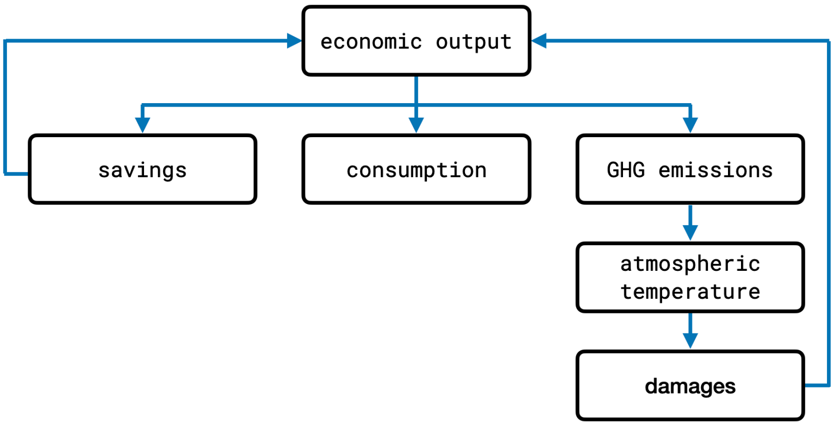


Figure 1.2: Integrated Assessment Model Structure

There are plenty of notable IAM frameworks, such as DICE (Nordhaus, 1992), RICE (Nordhaus and Yang, 1996), IMAGE (Stehfest et al., 2014), MESSAGEix (Huppmann et al., 2019), GCAM (Calvin et al., 2019), PAGE (Yumashev et al., 2019), or FUND (Tol, 1997). Within this research project, we will work with a version of the RICE model. There are multiple reasons for this. The Regional Integrated Climate-Economy model (RICE) is the immediate successor of the Dynamic Integrated Climate-Economy model (DICE) model for which its developer William Nordhaus has received the Nobel Memorial Prize in Economic Sciences. Over the last decades, the RICE model has become a common benchmark for IAMs. Although the DICE model is heavily criticized, it remains relevant in policy-making such as rather recently under the Obama administration to estimate the social costs of carbon (Aldy and Stavins, 2020). DICE, however, exhibits a very high degree of aggregation. It is geographically aggregated, so it does not provide any insights in how different regions are affected, i.e., how costs and benefits are distributed over space. The RICE model is an extension of the DICE model and is used to provide a higher geographical resolution[1]. Due to its simplicity, the RICE model is computationally relatively inexpensive. With comparably little compute, we can run plenty of optimizations and experiments to

---

[1] The appendix contains Table B.1 which provides an overview over the 12 regions of the PyRICE model.

explore its dynamics and gain insights. However, we are not only seeking to gain insights about the RICE model, but we are hopeful that the gained insights into the case study of the RICE model are transferable to other more complex and computationally expensive IAMs. That is where we expect the bigger policy impact eventually to be located.

The RICE model and IAMs in general have accumulated their fair share of criticism. Some researchers go as far as saying that climate policy analyses, which are based on IAMs, construct an illusionary perception of precision and certainty which misleads decision-makers to believe in a model's scientific legitimacy. IAM-based analyses of climate policy create a perception of knowledge and precision that is illusory, and can fool policy-makers into thinking that the forecasts the models generate have some kind of scientific legitimacy (Pindyck, 2020). IAMs have further been accused of being ill-equipped to adequately represent the complex economic reality of climate change crises (Stern, 2022). This is especially the case with the RICE model which exhibits an overly simplistic and outdated neo-classical money theory. Beyond that, others have also criticized its simple climate model with missing crucial tipping points and the general framework of having a central social planner which is utterly unrepresentative of the real-world situation. There are many more points of criticism, however, we would like to point to a last one that is especially dear to this research project. Many researchers have indicated that IAMs exhibit tacit but strong normative assumptions (Beck and Krueger, 2016; Jafino et al., 2021; Kowarsch, 2016). In fact, they hold the position that a great deal of equity-related problems have its origin in IAMs' inherent normative assumption embodied in a utilitarian social welfare function. Addressing this normative assumption is one of the key tenets of this research. The next section is closely related to tackling these normative assumptions by digging even deeper.

1.1.4. Premature Aggregation

Economist and former Chair of the Federal Reserve Ben Bernanke once stated ”Aggregate statistics can sometimes mask important information” (INET, 2012). Some might argue that aggregation does not only entail masking but the loss of information altogether. In 1950, Kenneth Arrow has investigated the question of how we can aggregate preferences of actors and what consequences such aggregations bring about (Arrow, 1950). Arrow has proven that there is no voting procedure for ordinal preferences which meets the following four requirements simultaneously: non-dictatorial, universal, independent of irrelevant alternatives, and Pareto-efficient. The corresponding theorem is known as Arrow's impossibility theorem or as Arrow's paradox. Amartya Sen has extended Arrow's proof to cardinal preferences (Sen, 2017, pp. 1882-186). Franssen (2005) has argued that the social-choice problem and the multi-criteria decision problem exhibit structural identity. One of the implications is that any aggregation of objectives will result in a situation where a subset of objectives will dominate the aggregate preference ordering. In other words, the aggregation over objectives will become dictatorial. One might argue that this dictatorial property of aggregation can be even further extended to the aggregation over world states, objectives, actors, time, and space. This is of course deeply problematic for any kind of decision analysis as it raises ethical concerns. In order to avoid Arrow's implications in which a subset of objectives will unintentionally dictate the optimized solutions in an unpredictable way, we attempt to avoid or postpone aggregation as much as possible (Kasprzyk et al., 2016). In this thesis, we aim to avoid such premature aggregation in the context of climate change decision support in various ways, such as by using many-objective planning with disaggregated objectives and multi-objective evolutionary algorithms to assist with discovering potential trade-offs between the objectives. Instead of providing a simple solution that is based on a highly aggregated objective, this approach allows for a posteriori decision support, which can enable a discussion and provide decision makers with a set of choices (Ciullo et al., 2019). Beyond the use of many-objective planning, we want to shed particular light on the disaggregation of utility and disutility which we consider to be incommensurable variables. A discussion that is based on Kulkarni's wok (2020) and is extended further, can be found in Section 2.2.1.

In this thesis, we argue that we can find climate justice behind the veil of aggregation. What we mean by this is that more equitable policy recommendations are obscured and lie hidden behind a bulwark of highly aggregated variables. If we look beyond this obstruction by the means of disaggregation, we are better equipped to find climate justice.

1.2. Related Work

This section provides a discussion of the relevant literature and the identified themes. Two systematic main literature scanning phases are conducted to arrive at the final collection of sources that are included in this review. Details can be found in Appendix A. The literature review yields the papers and dimensions that are presented in Table 1.1. Most presented papers are using one or two climate change IAMs in which one or more ethical principles have been part of the problem formulation[2]. The column *Research Angle* provides a short indication what the main focus of the corresponding paper is. The column *Model Name* shows what model has been used; it is mostly RICE or a variation thereof. The following column *Regions Count* indicates how many different regions have been considered by the model. The final three columns represent the identified core themes of this literature review which are *Ethical Principles*, *Focus on Dis/utility*, and *DMDU Use*[3]. These themes are discussed in the subsequent sections below.

---

[2] The exceptions are Ciullo et al. (2020) and Stanton (2011). Former do not use a climate change IAM but a model for flood risk management. The latter does not use any specific model but shows how a plethora of IAMs can be augmented to counter inequality.

[3] DMDU is a common shorthand for *decision-making under deep uncertainty*. More elaboration on this concept follows in Section 1.2.3.

Table 1.1: Selected Papers for the Literature Review

These papers are the result of a literature review. In the column *Objectives*, *welfare* refers to the weighted sum of discounted wellbeing which is based on consumption.

Paper	Research Angle	Model Name	Regions Count	Ethical Principles	Objectives	Focus on Dis/utility	DMDU Use
Nordhaus (1992)	introduction of DICE	DICE	1	utilitarian	welfare	no	no
Nordhaus & Yang (1996)	introduction of RICE	RICE	12	utilitarian	welfare	no	no
Tjallingii (2021)	effect of ethical principle on abatement pathways	RICE	12	utilitarian, prioritarian, sufficientarian, egalitarian	various, depending on problem formulation	no	yes
Kulkarni (2020)	effect of dissagregation of costs and benefits on climate action	DICE	1	utilitarian	atmospheric temperature, damages, total output, utility of consumption, disutility of damage, welfare	yes	yes
Lingeswaran (2019)	repercussions of fat-tailed distributions over uncertain parameters	DICE	1	utilitarian	atmospheric temperature, damages, total output, utility of consumption	no	yes
Budolfson et al. (2021)	maximizing social welfare instead of minimizing costs	RICE & FUND	12	utilitarian	welfare	partial	yes
Gazzotti et al. (2021)	comparing self-interested and cooperative behavior in many scenarios	RICE	50+	utilitarian + weights	welfare	no	no
Lamontagne et al. (2019)	evaluating the joint impact of HES uncertainties and abatement actions on climate-economic outcomes	CDICE + DOECLIM	1	utilitarian	discounted global utility	no	yes
Marangoni et al. (2021)	combining literature of 1) trade-offs between economic costs and environmental benefits 2) coupled human-Earth systems and their deep uncertainty	DICE + DOECLIM	1	utilitarian	discounted expected utility, warming above 2°C, damage costs, mitigation costs	partial	yes
Errickson et al. (2021)	importance of equity vs. climate uncertainty on social cost of methane	DICE & FUND	1	utilitarian + weights	welfare	no	no
Ciullo et al. (2020)	operationalizing ethical principles in flood risk management	-	-	utilitarian, constrained utilitarian, egalitarian, prioritarian	costs and variations on costs	no	yes
Dennig et al. (2015)	introduction of nested inequalities	RICE & NICE	12	utilitarian	welfare	no	no
Cantore & Padila (2010)	how do different climate policies affect emissions and income distribution over time	RICE	8	utilitarian	welfare	no	no
Ikefuji et al. (2020)	incorporates, possibly heavy-tailed, stochasticity	SDICE*	1	utilitarian	welfare	no	no
Mi et al. (2019)	introduces 4 equity principles	RICE	6	ability to pay, egalitarianism, grandfathering, historical responsibility	welfare	no	no
Anthoff & Emmerling (2019)	disentagle temporal inequality from spatial inequality	RICE & FUND	12	utilitarian + 2x weights	welfare	no	no
Dietz & Asheim (2012)	discounting utility	DICE	1	utilitarian + special discounting	welfare	no	yes
Shanton (2011)	introduction of Negishi welfare weights to counter inequality	-	-	utilitarian + weights	-	no	no

1.2.1. Theme: Ethical Principles

Generally speaking, a utilitarian perspective is the pre-dominant ethical assumption within IAMs. This is to the detriment of less well-off regions in the world. As Chancel (2020) points out, the current distribution of consumption per capita across nations is a sign for high degrees of inequality. And this inequality can be even exacerbated if we would follow the IAM optimal pathways as for example with DICE/RICE models (Dennig et al., 2015.). And although the search criteria contain equity-related terms, most of the selected papers exhibit either a utilitarian welfare function or an augmented version that incorporates population weights (see Table 1.1 under the corresponding column). Only a few papers (Ciullo et al., 2020; Mi et al., 2019; Tjallingii, 2021) introduce alternative distributive justice principles into the problem formulation of their particular model. Ciullo et al. formulate a *prioritarian*, an *egalitarian*, a standard *utilitarian*, and a *constrained utilitarian* principle. Within, the standard utilitarian approach, aggregate utility is being optimized. The constrained utilitarian approach reminds of a sufficientarian approach in which a threshold is defined that regions are not supposed to fall below – in this case, it is an individual region's status quo well-being. According to the prioritarian principle, the worst-off region's well-being has to be maximized, whereas in the egalitarian perspective, risk and benefits have to be as equally distributed as possible. Tjallingii exhibits virtually the same alternative principles. Mi et al. (2019), however, introduce four other equity principles. The first principle is *ability to pay* which states that a region's abatement costs are determined by how wealthy this region is. They understand *egalitarianism* as all people having equal rights to use atmospheric resources. *Grandfathering* entails that a region's current emission levels determine its future emission rights. *Historical responsibility* is understood as an equity principle in which past emissions are taken into consideration. The more a region has cumulatively emitted CO_2 equivalents, the more they are responsible to contribute to climate change mitigation now. In summary, most researchers use a utilitarian variant while only few go beyond this and introduce fully-fledged alternative principles.

1.2.2. Theme: Focus on Utility and Disutility

Most of the selected papers do not focus on a clear disaggregation of utility and disutility. Most papers have a clear focus on the maximization of welfare that is based on utility that is based on consumption. Budolfson et al. (2021), for example, focus on the maximization of social welfare because they consider the focus on costs to be too narrow as to ensure equity within their RICE model. Although, their problem formulation is a utilitarian one, they want to find more equitable pathways by arguing that utility maximization enables this. Marangoni et al. (2021) also use a utilitarian problem formulation with a simple DICE model. Although spatial and temporal variables are maximally aggregated, they are maximizing discounted expected utility and minimize costs (among other variables). So, in this sense, Marangoni et al. exhibit a meaningful disaggregation between utility and disutility while they are not explicitly formulating a disutility variable with e.g., its own discounting. Furthermore, Kulkarni's (2020) sole focus is the disaggregation of costs and benefits. Kulkarni uses a DICE model with a utilitarian problem formulation while investigating the impact of utility of consumption and disutility of damages on Pareto-optimal abatement pathways. Kulkarni makes a strong economic argument in favor of this kind of disaggregation by arguing how utility and disutility are incommensurable quantities. Applying such a disaggregation in other models and on other ethical principles might yield interesting results. In summary, most researchers do not consider a focus on the distinction between utility and disutility.

1.2.3. Theme: DMDU use

IAMs exhibit deep uncertainties in many of their key properties, such as for instance climate sensitivity and climate damages (Hwang et al., 2013; Knutti et al., 2017; Wagner and Zeckhauser, 2016; Weitzman, 2012). The term *deep uncertainties* is used in the literature in a way to refer to the lack of consensus on such external variables, the structure of the system, or which outcomes variables are of interest (Lempert, 2003). When facing problems that exhibit such deep uncertainties, the paradigm of *exploratory modeling* can become useful (Bankes, 1993). Marchau et al. (2019) summarizes in their book their approach of *decision-making under deep uncertainty* (DMDU). Often, Many Objective Evolutionary Algorithms (MOEAs) are used for optimization processes in which a Pareto-optimal set of candidate solutions are found that perform well (or well enough) for chosen objectives against a plethora of sampled scenarios (Coello et al., 2007). The last column of Table 1.1 indicates whether a paper makes use of such a DMDU approach. Seven out of 15 selected papers use such the DMDU approach. Dietz and Asheim (2012) are the earliest to use this approach by running a genetic algorithm (which they call *Riskoptimizer*). Lamontagne et al. (2019) make use of Sobol' sensitivity analysis to explore abatement pathways. Ciullo et al. (2020) use the ε-NSGAII (Kollat and Reed, 2005) search algorithm in combination with the Exploratory Modeling and Analysis Workbench (EMA-Workbench) (Kwakkel, 2017). The two Master theses (Kulkarni, 2020; Tjallingii, 2021) make also an extensive use of the EMA-Workbench and the ε-NSGAII search algorithm for their explorations. Finally, Marangoni et al. (2021) use Borg which is a state-of-the-art auto-adaptive evolutionary algorithm for multi-objective optimization (Hadka and Reed, 2013). In a nutshell, around half the papers use some form of DMDU while the application is mostly on a utilitarian problem formulation.

1.3. Research Objective

Kulkarni's disaggregation of costs and benefits show interesting results when applied to a utilitarian DICE model (2020). It would be interesting to see whether we would face similar results when applying such a disaggregation to non-utilitarian ethical principles. This would be another case of avoiding premature aggregation of potentially incommensurable variables. Given the previous discussions within the three themes, it appears that there is a research gap when it comes to the combination of using (a) a regionally disaggregated model (e.g., RICE), (b) a DMDU approach, (c) a focus on the disaggregation of utility and disutility which is applied to (d) alternative ethical principles. This research gap leads then to the research question below.

Main Research Question

How are Pareto-optimal climate abatement pathways affected by the disaggregation of utility and disutility in alternative ethical problem formulations when using an integrated assessment model under deep uncertainty?

The main research question asks how abatement pathways are affected by the application of specific ethical problem formulations. The phrasing of this question does not allow a simple conclusive answer. As the literature review shows, there is only very limited research on the effect of disaggregation of utility and disutility. This research project is an initial step to explore this uncharted space. Within this project, the PyRICE model is used and expanded which would render the research approach a modeling one. There is indeed a lack of understanding with respect to the socio-technical system of the the economy and the climate. The consequences of different problem formulations on the resulting pathways are unknown and can be explored. Given that the research design requires the context of decision-making under deep uncertainty, *exploratory modeling* is a natural fit (Bankes, 1993). Given all the various uncertainties that are related to the economy and the climate (see e.g., section 1.2.3), simple predictions cannot be made. Instead, policies and their associated pathways need to be discovered and explored.

In order to answer the main research question, smaller steps and sub-questions have to be distinguished. And for this purpose, we break down the main research question into its key components:

1. alternative ethical problem formulations
2. disaggregation of utility and disutility
3. Pareto-optimal climate abatement pathways
4. under deep uncertainty

Let us start with the alternative ethical problem formulations. One of the points of criticism against IAMs is the common use of a utilitarian framework (see Section 1.1.3). But this distributive justice principle[4] is not the only one in town. The field of ethics offers a plenitude of alternatives. Some have already been mentioned in Section 1.2, such as versions of egalitarianism, sufficientarianism, prioritarianism, and rawslianism. The first research question revolves around the question what alternative distributive justice principle should be considered for our problem formulations.

Sub-Question 1

What alternative distributive justice principles are suitable contenders to enable more equitable climate abatement pathways?

Next, we can direct our attention to the second component: the disaggregation of utility and disutility. As this component is at the core of the main research question, we build on Kulkarni's disaggregation work on the utilitarian DICE formulation (2020) and extend it to other ethical premises within the RICE model. This brings us to the following question:

Sub-Question 2

How can utility and disutility be disaggregated in each of the chosen distributive justice principles?

The third component refers to Pareto-optimal abatement pathways and how they are affected. This begs the question what pathways we exactly mean. Eventually, we want to compare the pathways across all problem formulations. Thus, we need to find sensible metrics across these formulations. With the help of a literature review and review of the PyRICE model, key performance indicators (KPIs) are identified which are used to compare the policies' performances on. Some of these KPIs are in the objective space of some problem formulations, but some are outside of that space. Consequently, the following research question shall be addressed:

Sub-Question 3

What are suitable key performance indicators to compare the performance of eventually found policies on?

The last component refers to deep uncertainty. Here, we want to know how the found policies perform under a set of uncertain parameters and whether the resulting Pareto-optimal policies exhibit robustness. The corresponding question is:

Sub-Question 4

How well do the found Pareto-optimal policies perform under uncertainty?

---

[4] In the context of this thesis, we equate the terms *distributive justice principle* and *ethical premise*.

For this purpose, we sample scenarios from the uncertainty space and acquire the performance, trade-offs, and robustness of the policies across this space in the phase of data analysis. With these pathways, we compare and contrast the performance of the identified policies and consequently answer how optimal pathways compare across the ethical problem formulations and how they compare between the aggregated and disaggregated problem formulations. To answer these questions, we need to compare the KPI pathways of the identified polices conditioned on the problem formulation. Having answered these questions will lead us to the point in which we can synthesize the insights and answer our main research question which is repeated here:

Research Question

How are Pareto-optimal climate abatement pathways affected by the disaggregation of utility and disutility in alternative ethical problem formulations when using an integrated assessment model under deep uncertainty?

The need for climate justice is voiced louder than ever. If alternative and equity-centered problem formulations do indeed yield substantially different abatement policies and pathways, there is an acute need for developing such models and exploring more distributively just paths into the future where the focus on utility versus disutility could play a great role.

1.4. Thesis Outline

After the introduction, we introduce the PyRICE model in the following chapter (Chapter 2), where we talk about its inputs, outputs, and its input-output-mapping. We answer sub-questions 1, 2, and 3 in this chapter. Chapter 3 offers information on the methods that we are using to investigate our research questions. In Chapter 4, we present our results and address sub-question 4. In Chapter 5, we continue with a discussion in of our key findings and the research limitations. Eventually, we conclude with Chapter 6, summarizing our research answers and inferring the scientific and policy implications.

2

The PyRICE Model

In this study, the RICE model is transformed into a simulation model and embedded in a many-objective simulation-optimization setup in order to find Pareto-optimal climate mitigation pathways. We call it the PyRICE model and a version has been initially developed by Tjallingii (2021) in the context of his master thesis. This model exhibits also Tjallingii's extension of alternative ethical principles with utilitarianism, egalitarianism, sufficientarianism, and prioritarianism[1]. The author of this thesis has refactored this model during a previous internship at the HIPPO Lab at TU Delft in 2021 in order enable modularity, adherence to object-oriented programming principles, re-usability, and compliance with PEP8[2]. In the context of this master thesis, the author extends the refactored model where necessary.

The intention behind this chapter consists in providing insight in how the PyRICE model works. In order to explain the key aspects of the PyRICE model, we use the framework of XLRM which consists of four components (Lempert, 2003, p. 70):

- **X: exogenous uncertainties**. These are variables that a decision-maker has no control over but will, however, still affect the system and therefore the performance of policies. All uncertainty variables span the uncertainty space. A point in that space is called a scenario.
- **L: levers**. These are variables that that a decision-maker has control over and with which they can affect the system. All lever variables span the policy space. A point in that space is called a policy.
- **R: relations**: This component refers the inner workings of the system, describing a mapping between X and L on the one hand, and the metrics (M) on other hand.
- **M: metrics**: These are variables that a system outputs and that a decision-maker cares about[3].

A visual overview is presented in Figure 2.1. The subsequent subsections each explain one of the major components. Merely, the order is altered to: R, M, L, and then X.

---

[1] Why these four ethical principles have been chosen by Tjallingii and why we decided to continue with them, is discussed in Section 2.2.1.

[2] PEP8 is a style guide for Python code and can be found here.

[3] In the remainder of this thesis, we equate the terms *metrics* and *objectives*.

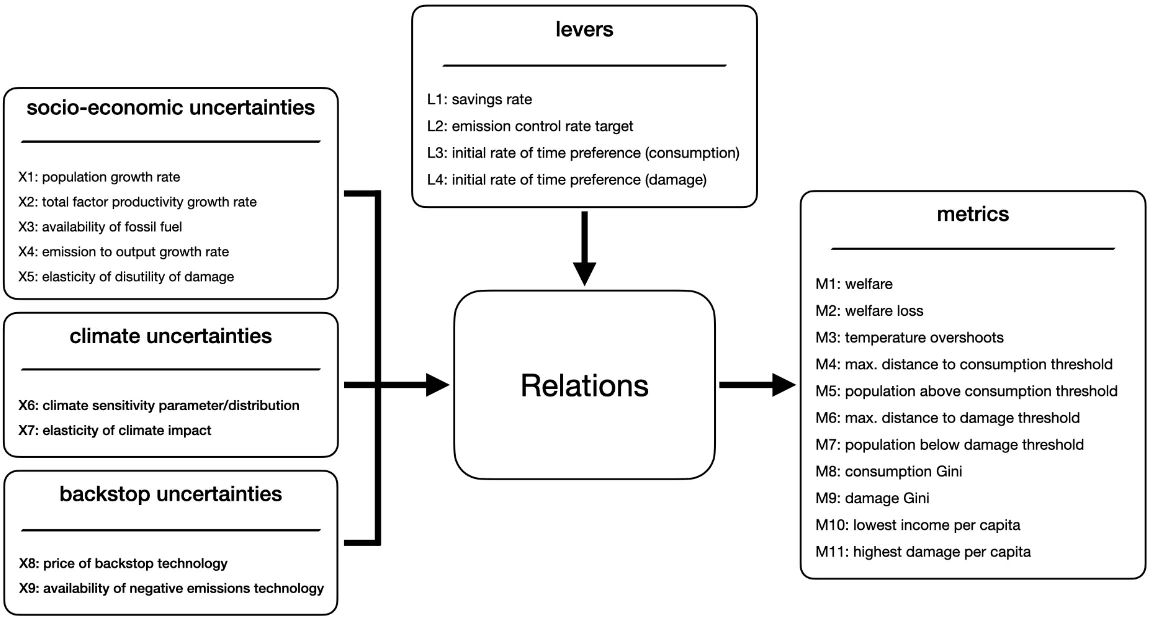


Figure 2.1: XLRM for PyRICE

This figure contains the XLRM components of the PyRICE model. On the left, all uncertainties are listed with their corresponding category. The details on the metrics are omitted because they depend on which problem formulation is chosen. Details on these problem formulations follow in Section 2.2.

2.1. R: Relations

The relations of the PyRICE model can be viewed through the lens of four separate sub-models which are depicted in Figure 2.2 below.

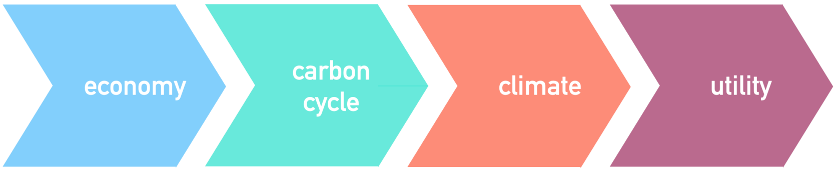


Figure 2.2: Sub-Models of PyRICE

The direction of the arrows represents the order in which information is generally processed.

2.1.1. Economic Sub-Model

We start with the *economy model* which is based on economic growth theory in the context of neoclassical money theory. According to the Ramsey-Cass-Koopman model, an economy starts out with its economic production which is partially consumed today and partially saved and therefore reinvested as capital stock for future consumption (Koopmans, 1966). The representation of agency is a bit tricky as we have to imagine each region to have only a single agent whose life spans the entire time horizon of 300 years. Consequently, postponing consumption by saving money in e.g. 2005 will conceptually not benefit potential descendants but the same agent in e.g. 300 years – let it be in a discounted fashion though. Generally, agents seek to maximize their overall discounted utility. For this, we need a region's GWP or total output which can be computed by the Cobb-Douglas production function (Equation B.1). As the PyRICE model adjusts its economic output by incorporating abatement and damage costs, we arrive at modified production function (Equation B.3). Abatement costs are based on the declining costs of backstop technology due to increasing technological efficiency (see Equation B.4) (Nordhaus, 2010). The damage function here is the default Nordhaus damage function which is based on atmospheric temperature and sea level rises (see Equation B.5). In combination with the modified production, the *savings rate* (which is one of the policy levers) determines consumption and investment levels (see Equation B.6 and Equation B.7). What role discounting plays in the calculation of `welfare` and `welfare loss` is described later in Section B.3.3.

2.1.2. Carbon-Cycle Sub-Model

The global economic activity eventually affects the carbon-cycle of the planet. This sub-model is first used to compute the global emissions which are based on each region's industrial emission and the global land usage emissions. The former is dependent on the production output of the economic sub-model, the emission output ratio, and the emission control rate (see Equation B.9). Using the global emissions, we can compute global carbon concentration, namely CO_2 concentrations in the atmosphere and CO_2 absorption and release in upper and lower oceans (see Equation B.10). Eventually, we can use these to compute the levels of radiative forcing resulting in higher energy influx than efflux given higher CO_2 concentrations (see Equation B.11).

2.1.3. Climate Sub-Model

Radiative forcing strongly affects the degree of atmospheric and ocean temperature increases. Other factors are climate equation coefficients for the upper and lower stratum, heat transfer coefficient between the two strata, and climate sensitivity parameters (see Equation B.12).

2.1.4. Utility Sub-Model

Eventually, we can use the outputs from the economic and the climate sub-model to compute various outcome variables. Originally, these variables are related to computing utility and welfare. That is why we call it the *utility sub-model*. In the context of this research project, we consider more than just one aggregate objective which is why we created a separate sub-model. What these objectives consist of, is described in the next section which focuses on the metrics within the XLRM framework.

2.2. M: Metrics

The set of metrics depends on the selection of the problem formulation. The set of problem formulations can be described as

$$\mathcal{PF} = \mathcal{E} \times \mathcal{A} \tag{2.1}$$

where

$\mathcal{PF}$	set of problem formulations
$\mathcal{E}$	set of ethical premises
$\mathcal{A}$	set of aggregation levels

In the following, we explain what ethical premises we choose (Section 2.2.1), and what we mean with aggregation levels (Section 2.2.2). This is followed by an overview over the different problem formulations and a deeper dive into their corresponding metrics (Section 2.2.3).

2.2.1. Alternative Distributive Justice Principles

Let us start with the the set of ethical premises $\mathcal{E}$ and tackle our first research sub-question which is repeated here:

Sub-Question 1

What alternative distributive justice principles are suitable contenders to enable more equitable climate abatement pathways?

These results are based on Tjallingii's literature review with respect to alternative welfarist principles within IAMs (2021). In a nutshell, the welfarist approach presupposes that we can gauge the value of policies on the basis of their impact on the *welfare* of the people. In the framework of welfarism, we can rank policies according to their induced welfare levels (Sen, 2017). Neglecting all procedural aspects, we only care about the distribution of welfare. The question is now what such welfare is based on – among many others, it could be pleasure, preference satisfaction, happiness, or consumption. For this purpose, welfarists are required to specify their so-called *social welfare functions* which are used to determine the ideal allocation of welfare across all parties (Lamont and Favor, 2017). The standard social welfare function in IAMs is the utilitarian social welfare function in which we maximize aggregate discounted utility (M. D. Adler, 2019). As we have described earlier, the use of this normative assumption in IAMs is a common point of criticism that we can address by considering alternative social welfare functions. Common alternatives are forms of sufficientarianism, egalitarianism, and prioritarianism which are summarized in Table 2.1 (M. Adler et al., 2017; M. D. Adler, 2019; Dietz and Asheim, 2012; Robiou du Pont et al., 2017).

Table 2.1: Ethical Premises

Ethical Premise	Definition
utilitarianism	maximize aggregate wellbeing
sufficientarianism	minimize the number of people below a critical consumption threshold
egalitarianism	minimize inequality of wellbeing
prioritarianism	maximize wellbeing of the worst-off

For the purpose of this research project, we use these four social welfare functions in the PyRICE model. Having covered the set of ethical premises $\mathcal{E}$, we can turn to the second component of our problem formulations – the set of aggregation levels $\mathcal{A}$.

2.2.2. Aggregation and Disaggregation

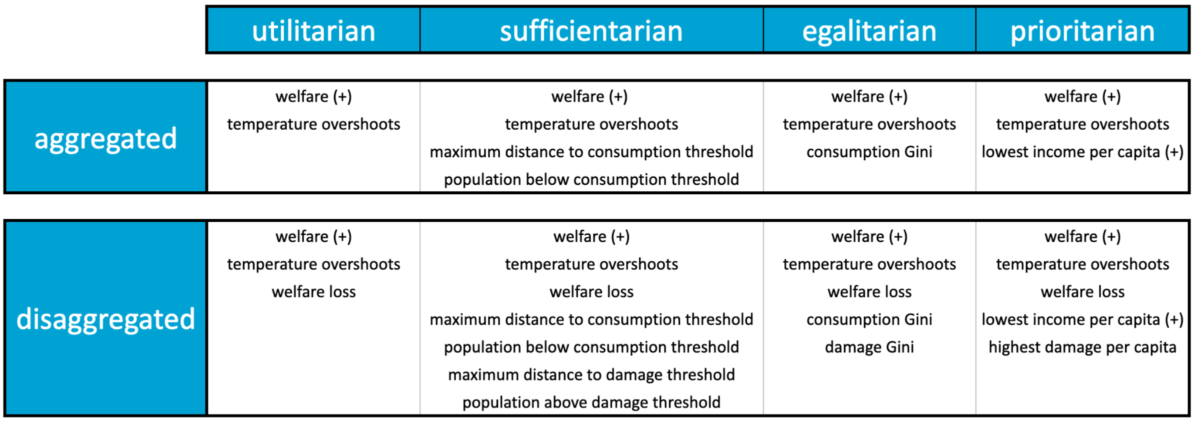
In our case, the set of aggregation levels $\mathcal{A}$ contains only two elements: *aggregated* and *disaggregated*. These levels of aggregation refer to the question whether utility and disutility are aggregated or disaggregated. To make this a bit clearer, let us first have a look at Kulkarni's research who argues that we should indeed disaggregate these two variables (2020). Kulkarni analyzed how the disaggregation of utility and disutility can affect climate action in the IAM of DICE. His arguments focused on a) the inadequacy of accounting for the threat of climate damages on wellbeing and b) the inadequacy of accounting for social preferences to damage risk aversion. Point a) includes reduced capabilities that are the repercussion of climate damages, the shortcoming of a Pigouvian tax (e.g. Social Cost of Carbon), and the non-existence of discounting climate damages. All in all, Kulkarni makes a strong point that concepts such as utility and disutility are two incommensurate quantities. If we consider only utility as an objective, its underlying consumption is the most crucial aspect. Remembering Figure 1.2, we know that climate damages are fed back into the model and therefore affect the economic output and by transition the consumption of the next time step. Focusing only on utility (and thus on consumption) accounts *indirectly* for damages. Given Kulkarni's arguments, however, it would be advisable to see the two as separate quantities and objectives. He applied this disaggregation of utility and disutility on the standard utilitarian formulation of the DICE model. We want to apply such a disaggregation in a similar manner to all four ethical premises in $\mathcal{E}$. The next section provides an overview over the resulting problem formulations $\mathcal{PF}$ and how we perform these disaggregations.

2.2.3. Problem Formulations

An overview of the problem formulations and their corresponding metrics can be found in Table 2.2. In the following, we discuss each of the metrics.

Table 2.2: Problem Formulations

The columns represent the ethical premises while the rows represent the aggregation level. A cell represents a single problem formulation. The corresponding metrics are listed in its cell. Each metric in this table corresponds to the year 2105. A (+) indicates that the corresponding objective needs to be maximized. If this addition is missing, the objective needs to be minimized.



	utilitarian	sufficientarian	egalitarian	prioritarian
aggregated	welfare (+) temperature overshoots	welfare (+) temperature overshoots maximum distance to consumption threshold population below consumption threshold	welfare (+) temperature overshoots consumption Gini	welfare (+) temperature overshoots lowest income per capita (+)
disaggregated	welfare (+) temperature overshoots welfare loss	welfare (+) temperature overshoots welfare loss maximum distance to consumption threshold population below consumption threshold maximum distance to damage threshold population above damage threshold	welfare (+) temperature overshoots welfare loss consumption Gini damage Gini	welfare (+) temperature overshoots welfare loss lowest income per capita (+) highest damage per capita

Each white cell represents a problem formulation's set of objectives. We can see that within an ethical premise, all objectives of the aggregated version are also represented in the disaggregated version. Thus, the objectives of aggregated problem formulations are real subsets of disaggregated versions.

For simplicity's sake, we introduce shorthands for the different problem formulations. Because the level of aggregation refers always to an ethical premise, we use a normal capital letter (U, S, E, and P) to indicate the ethical premise and a subscript capital letter (A and D) to indicate the level of aggregation. A complete list is shown below in Table 2.3. In the results section (Chapter 4), we make extensive use of these shorthands.

Table 2.3: Shorthands for Problem Formulations

Shorthand	Problem Formulation
U_A	Utilitarian Aggregated
U_D	Utilitarian Disaggregated
S_A	Sufficientarian Aggregated
S_D	Sufficientarian Disaggregated
E_A	Egalitarian Aggregated
E_D	Egalitarian Disaggregated
P_A	Prioritarian Aggregated
P_D	Prioritarian Disaggregated

In the next step, we want to dive deeper in the differences between these two types of problem formulations and eventually answer the second sub-research question which is repeated here:

Sub-Question 2

How can utility and disutility be disaggregated in each of the chosen distributive justice principles?

For this purpose, the following subsections explore the metrics and how we disaggregate them where possible. Except for Section B.3.2 (for `temperature overshoots`), all other subsections consider one or two utility-related metrics and their corresponding disutility-related metrics. Let us go through the objectives column by column.

The objectives `welfare` is a commonly used objective in IAMs (Equation B.14) and is part of each of our problem formulations. It is based on discounted utility of consumption which is defined by the Constant Relative Risk Aversion utility function. Furthermore, `welfare` relies on an endogenous *social discount factor* (Equation B.16) which is based on an *endogenous social rate of consumption* (Equation B.17). This rate depends on the *initial rate of time preference for consumption* (which is one of the policy levers) and an endogenous *rate of change of consumption* (Equation B.18). In the disaggregated version, we can see the objective `welfare loss` which can be considered the counter part of `welfare` and is represented in each disaggregated problem formulation. Its computation is very similar to the computation of `welfare` and is strongly based on Kulkarni (2020). Instead of utility, we compute disutility of damages with its own coefficient of relative risk aversion – but not for consumption as in utility but for climate damages. The other underlying endogenous functions are following the same trend. Details can be found in the appendix under Section B.3.3. The next objective is `temperature overshoots` which describes the number of 2°C atmospheric overshoots up to a given year[4]. This objective has recently been increasingly used in the IAM research community and is therefore also used in our research and shared by every problem formulation (Drouet et al., 2021). It is also the only objective that does not have a disaggregation equivalent. This covers the first column and therefore all utilitarian objectives.

The same three objectives are recurring in other problem formulations which is why we will not repeat them. In the second column, we can see the sufficientarian objectives. In the aggregated version, two objectives are added (compared to the utilitarian one) – `maximum distance to consumption threshold` and `population below consumption threshold`. According to the sufficientarian principle,

---

[4] E.g., `temperature overshoots 2105` would refer to the number 2°C atmospheric overshoots up to and including the year 2105. It shall be noted that it does not count every year but considers a step size of 10 years. I.e., up to the year 2205, there are maximally 20 temperature overshoots as $(2205 - 2005)/10 = 20$ where 2005 is the starting year and 10 is the step size.

we stipulate a consumption threshold which we borrow from the World's bank definition of a poverty line[5]. Then, we can examine how many people turn out to be below this consumption threshold[6]. Additionally, we also consider the set of all regions' quintiles, how short they fall from reaching the consumption threshold. The maximum distance to this threshold is represented by the variable `maximum distance to consumption threshold`. The disaggregated equivalents for these two sufficientarian objectives are `maximum distance to damage threshold` and `population above damage threshold`. They work in an analogous way with the only difference that we consider not pure damages but the ratio of damages over consumption. The reason for this is that absolute values of damages are lacking the context and do not express how severe the damages are for a specific region. Consumption levels provide the necessary context. The corresponding threshold is set to damages being maximally as high as 5% of the consumption level[7].

The additional objectives of the egalitarian problem formulations are rather simple. We consider a Gini index for consumption across the 12 PyRICE regions (`consumption Gini`) and a Gini index for damages over consumption (`damage Gini`). For the prioritarian objectives, we have `lowest income per capita` which considers the average income of all regions' income quintiles and picks out the worst-off. Similarly, for `highest damage per capita`, we take each regions' damage over consumption per capita and pick out the worst-off. More details can be found in the appendix under Section B.3.5 and Section B.3.5.1. This concludes the metrics of our problem formulations and brings us further to the component of the XLRM framework: L for the levers.

2.3. L: Levers

We have only four policy levers which are listed in Table 2.4 below.

Table 2.4: Levers

Levers	Range
savings rate	$[0.1, 0.5]$
emission control rate target	$[2065, 2305]$
initial rate of social time preference consumption	$[0.001, 0.015]$
initial rate of social time preference damage	$[0.001, 0.015]$

The savings rate describes the percentage of disposable income that a region decides to save rather than spends on consumption. The savings rate is exogenously set to converge to the steady-state savings rate for the world. The emission control rate target describes the year in which the world exhibits net-zero emission for the first time. The initial rates of social time preferences represent how much consumption or damage by future generations is valued. These are, however, only initial values. The resulting social discount rates are endogenous, adjusting with e.g., population, consumption, and damage growth. These two levers are used in Equation B.17 and Equation B.22, respectively. More information with respect to variable names, value ranges, and units can be found in the appendix in Table B.4.

---

[5] The chosen poverty line is \$3.20 per person per day. We have also considered the World Bank's lower poverty line of \$1.90 but our preliminary analyses have shown that this threshold was not strict enough to meaningfully affect the outcomes (World Bank, 2018, p. 68). Furthermore, this consumption threshold is only an initial one and adjusting with a region's growing consumption over time.

[6] For this purpose, we split a region's population in five income quintiles and consider for each quintile, whether the average consumption is above the given threshold.

[7] This value is chosen on the basis of the recent IPCC reports defining severe key risks to aggregate economic output as "economic losses due to climate change to match or exceed losses during the world's worst historical economic recession", which is roughly a 5% GDP drop during the 2008-2009 recession (Masson-Delmotte et al., 2021).

2.4. X: Uncertainties

In the framework of decision-making under deep uncertainty (DMDU), we naturally account for uncertainties. The final selection of parametric uncertainties can be found in Table 2.5 below.

Table 2.5: Uncertainties

Uncertainty	Type	Range
population growth rate	socio-economic	{short-term, long-term}
total factor productivity growth rate	socio-economic	{short-term, long-term}
availability of fossil fuel	socio-economic	$[4000, 13649]$
emission to output growth rate	socio-economic	{short-term, long-term}
elasticity of disutility of damage	socio-economic	$[0.001, 0.6]$
climate sensitivity parameter	climate	{normal, log normal, Cauchy}
elasticity of climate impact	climate	$\{-1, 0, 1\}$
price of backstop technology	backstop	$[1260, 1890]$
availability of negative emissions technology	backstop	{yes (20%), no}

We distinguish three types of uncertainties: socio-economic, climate, and backstop uncertainties which are based on Tjallingii (2021), Kulkarni (2020), and Lingeswaran (2019) who all worked on predecessor versions of the PyRICE model. In short, most socio-economic and backstop uncertainties relate to shared socio-economic pathways (SSPs) which have been developed by the IPCC community to improve the comparison and translation of research findings across studies on climate change (Lamontagne et al., 2018). There is one additional socio-economic uncertainty is the *elasticity of disutility of damage* which describes a coefficient of relative risk aversion for climate damage and is used to compute the endogenous social rate of damage. This uncertainty has been introduced by Kulkarni (2020) to disaggregate utility and disutility and is adopted for this research as well. The climate uncertainties are based on the work of Lingeswaran (2019). Again, more information with respect to variable names, value ranges, explanations, and units can be found in the appendix in Table B.3.

2.5. Key Performance Indicators

After having covered all components of the XLRM framework, we would like to show an addition that is strongly related to the metrics. We are talking about the key performance indicators (KPIs). We need to identify them to answer our third sub-question which is repeated in the following:

Sub-Question 3

What are suitable key performance indicators to compare the eventually found policies on?

An overview of the KPIs can be found in Table 2.6 below.

Table 2.6: Key Performance Indicators

This table contains a list of all used KPIs, whether the KPIs are in the objective space Ω which contains all objectives that are part of the eight problem formulations, and their corresponding equations. Only the last equation is not shown here as the associated equation is in a long matrix equation. All details on the equations can be found in Appendix B.

KPI	KPI $\in \Omega$	Equation
welfare	yes	$W_{c,t,r} = \sum_{i=1}^{t} U(C)_{i,r} \cdot R_{c,i,r} \cdot L_{i,r}$
welfare loss	yes	$W_{d,t,r} = \sum_{i=1}^{t} V(\Omega)_{i,r} \cdot R_{d,i,r} \cdot L_{i,r}$
temperature overshoots	yes	$O_t = \sum_{i=2005}^{t} f(T_i)$
distance to consumption threshold	yes	$d_{c,t} = max(\ \forall(q \in Q)\ f(q, \theta_{c,t})\ )$
distance to damage threshold	yes	$d_{d,t} = max(\ \forall(r \in R)\ f(r, \theta_{d,t})\ )$
population below consumption threshold	yes	$POP_{c,t} = \sum_{r=0}^{12} \sum_{q=0}^{5} f(p_{r,q,t})$
population above damage threshold	yes	$POP_{d,t} = \sum_{r=0}^{12} f(p_{r,t})$
consumption Gini	yes	$G_{c,t} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \lvert C_{t,i} - C_{t,j} \rvert}{2n^2}$
damage Gini	yes	$G_{d,t} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \lvert \Omega_{t,i} - \Omega_{t,j} \rvert}{2n^2}$
lowest income per capita	yes	$C_{c,min,t} = min[C_{t,r,q}$ for region r in quintile $q]$
highest damage per capita	yes	$D_{c,max,t} = max[\Omega_{t,r,q}$ for region r in quintile $q]$
GWP	no	$Y_{gross,t,r} = A_{t,r} \cdot L_{t,r}^{1-\gamma} \cdot K_{t,r}^{\gamma}$
damages	no	$\Omega_{t,r} = \phi_{1,r} \cdot T_{ATM,t} + \phi_{2,r} \cdot T_{ATM,t}^{\phi_{3,r}}$
industrial emissions	no	$E_t = \sum_{r=1}^{R} (E_{ind,t,r}) + E_{land,t}$
consumption per capita	no	$CPC_{r,t} = \frac{C_{r,t}}{L_{r,t}}$
damage per capita	no	$DPC_{r,t} = \frac{\Omega_{r,t}}{L_{r,t}}$
atmospheric temperature increase	no	see Equation B.12

As we can see, most KPIs are part of the objective space and therefore objectives that are optimized for some problem formulations. However, we add a few more KPIs that extend beyond the objective space. The reason for doing this is that most objectives are not shared across all problem formulations (only `welfare` and `temperature overshoots` are). Furthermore, we want to compare the pathways of easily understood and commonly used KPIs such as atmospheric temperature increases (Lamontagne et al., 2019; Marangoni et al., 2021); global CO_2 emissions, gross world product (GWP), and global damages (Gazzotti et al., 2021; Tjallingii, 2021). Furthermore, we use consumption and damage per capita to compare regional pathways across the problem formulations.

3

Decision-Making under Deep Uncertainty

3.1. Background

This section offers the theoretical underpinning for the subsequent methods. First, we introduce the main framework of this thesis: multi-scenario multi-objective robust decision-making (Section 3.1.1). This is followed up by an elaboration on Pareto-optimality (Section 3.1.2) and how directed search can be used for its purpose (Section 3.1.3).

3.1.1. Multi-Scenario Multi-Objective Robust Decision-Making

Within the DMDU framework, there is an approach that is called *multi-objective robust decision-making* (MORDM) (Kasprzyk et al., 2013; Watson and Kasprzyk, 2017). The steps of this framework are shown in Figure 3.1. Generally, policies are searched with the help of an evolutionary algorithm, which optimizes given metrics. In the first step, the model has to be specified with all its hyperparameters, including the problem formulation (i.e., which metrics we care about). In the second step, we identify vulnerable scenarios which are future states that especially grim. Usually, MORDM relies on a single reference scenario which can be problematic (Eker and Kwakkel, 2018). Relying on multiple scenarios can extend MORDM to MS-MORDM and render the policies overall more robust (Watson and Kasprzyk, 2017). With the help of these so-called reference scenarios, we can use directed policy search to discover optimal policies for a problem formulation. These identified policies are then tested against an ensemble of possible future scenarios which are sampled from an uncertainty space. Eventually, various robustness metrics are used to determine the policies' robustness (Kasprzyk et al., 2013). After having gone through all the above steps, we can go through another iteration by re-specifying the model and discover even more vulnerable scenarios. This can be done by conducting a vulnerability analysis, e.g., sub-pace partitioning via PRIM (Friedman and Fisher, 1999) or CART (Breiman et al., 2017); or a factor prioritization and fixing via a global sensitivity analysis such as SOBOL (Jaxa-Rozen and Kwakkel, 2018). Eventually, we conduct a trade-off analysis between the different problem formulations.

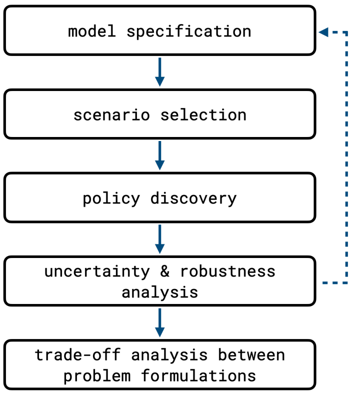


Figure 3.1: MS-MORDM Steps

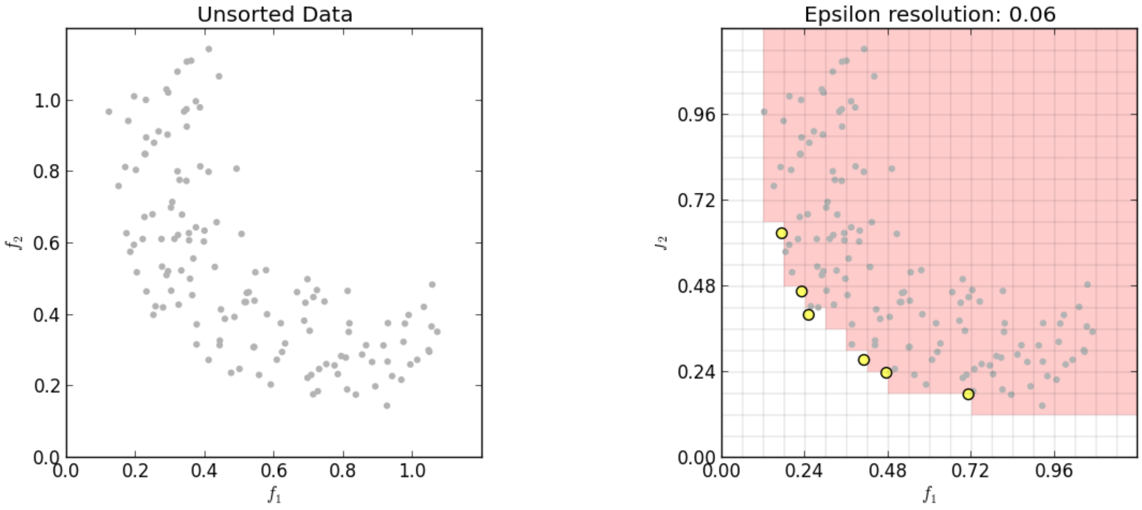
3.1.2. Pareto-Optimality

If there would be only a single (aggregate) objective, we could use any run-of-the-mill optimization algorithm. Optimality in this context means to find solutions that move the value of the objective in the desired direction. However, we are facing a problem that contains many objectives. Consequently, the intuitive understanding of optimality is not holding up anymore. A many-objective optimization problem can be seen in the following way (Osyczka, 1985): The policy vector (with its levers as dimensions) optimizes a vector function which has our metrics as its dimensions. These metrics have their own objective functions which form a formal description of criteria which can be in conflict with each other.

Optimality refers here to finding solutions which are Pareto-optimal. This means, that we cannot find any solution that improves the value of one metric without impairing the value of at least one other metric. As we cannot find a single global optimum in such a situation, we shift the aim to finding good trade-offs between the metrics. According to Coello et al. (2007, p. 10), we can formally express Pareto-optimality like this: A solution $x \in \Omega$ is considered to be Pareto-optimal with respect to the entire decision variable space Ω if and only if there is no $x' \in \Omega$ for which $v = F(x') = (f_1(x'), ..., f_n(x'))$ dominates $u = F(x) = (f_1(x), ..., f_n(x))$ F represents an objective vector function while f_i represents an objective function for metric i out of n metrics.

3.1.3. Directed Search

Directed search is used for both, the scenario and policy discovery step. The search refers to searching for Pareto-optimal solutions. Many-objective evolutionary algorithms (MOEA) are one way to find such solutions. In our case, we make use of the ϵ-NSGA-II algorithm (P. Reed and Devireddy, 2004). The NSGA-II algorithm is used to rank candidate solutions with Pareto-dominance. With each iteration, candidate solutions are sorted to identify the non-dominated set of candidate solutions. The ϵ in ϵ-NSGA-II introduces ϵ-dominance by specifying intervals for each objective dimension and can be considered the level of precision for our found solutions. For better illustration, let us consider Figure 3.2 below.



(a) Data before sorting

(b) Data after epsilon-non-dominated sorting

Figure 3.2: Example for ϵ-Dominated Sorting

Minimizing both objectives f_1 and f_2.
Source: (Woodruff and Herman, 2013)

For simplicity's sake, we have only two objectives f_1 and f_2. We aim to minimize both objectives. Thus, solutions in the bottom-left corner are more desirable. In the left figure, we can see data points in this objective space. In order to find the Pareto-front, we use an ϵ-value for each objective. In this example, it is the same for both objectives. The right figure shows the resulting grid. The red epsilon-boxes and all their contained candidate solutions are dominated by the yellow candidate solutions and considered ϵ-non-dominated. Given this current set of data points, there are no candidate solutions that are e.g. one cell further to the left without being one cell higher. This means that we cannot find any solution that is better for one objective without making it worse for the other objective. This is in a nutshell how directed search with ϵ-NSGA-II works in order to find our Pareto-optimal solutions.

3.2. Open Exploration

In order to gain insights into the behaviors dynamics of the model and feed this data for the scenario selection process, we use the method of open exploration. Consequently, we explore the input-output mapping of the system with the input being a vector containing levers and uncertainties, and the output being a vector function with its metrics. In the framework of an experimental setup, we can run the PyRICE simulation model over thousands of times while sampling scenarios from the uncertainty space. As a policy, we consider the Nordhaus policy as it represents a common benchmark. The Nordhaus policy is the policy that is the result of William Nordhaus' original RICE optimization model.

3.3. Scenario Selection

Figure 3.3 shows the basic steps of the scenario selection process. Through scenario selection, we identify a subspace of the uncertainty space which is vulnerable to a set of policies. We use two methods to find vulnerable scenarios – *time series clustering* and *directed scenario search*. Both methods have drawbacks. Time series clustering is dependent on performing random experiments (open exploration). Theoretically, we do not know how many scenarios are considered sufficient for this purpose. In order to avoid potential drawbacks of using a low number of scenarios, we complement this method with directed scenario search. Directed scenario search in turn does not guarantee evenly spaced solutions. The combination of both should give us a better combined set vulnerable scenarios. Eventually, we intend to find four reference scenarios that are maximally diverse and represent the vulnerable space. These reference scenarios can then be used to find policies that optimize our objectives by feeding them to the MOEA ϵ-NSGA-II as described in Section 3.1.3. Our scenario selection process consists of two main steps. First, we identify a large number of vulnerable scenarios with time series clustering (see Section 3.3.1) and directed scenario search (see Section 3.3.2). Eventually, we select a tiny subset of these consisting of only four scenarios (see Section 3.3.3).

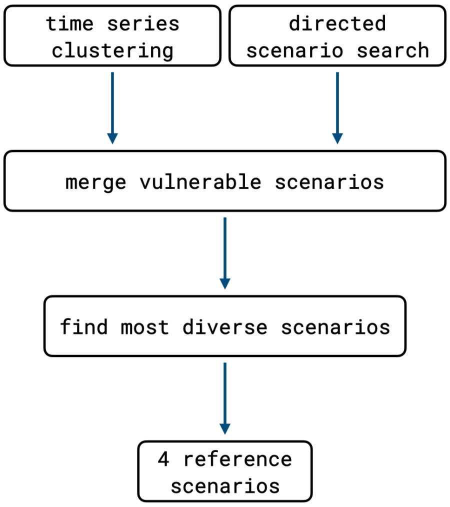


Figure 3.3: Scenario Selection Process

3.3.1. Time Series Clustering

As we want to find a vulnerable subspace of the uncertainty space, we could make use of an unsupervised learning algorithm – clustering. Such algorithms are used to automatically find grouping in data. For each grouping or cluster, there is one point which is supposed to be most similar to all its cluster members and most distinct to all of its non-cluster members. If we would have no time series data, this would fit our purpose very well. However, we are faced with pathways (e.g., atmospheric temperature increase over many years) which represents time series data. In a similar way, time series clustering algorithms are taking time series data as input and returning these time series data subset into clusters that maximize similarity within clusters and maximize diversity between clusters (Liao, 2005). We use a time series clustering algorithm and apply it to the data we have gathered during the process of open exploration as envisioned by Steinmann et al. (2020). Each of our eleven metrics (see Table B.2) offers a set of pathways with 30’000 scenarios. We find various clusters for the pathways of each metric. Naturally, we select those clusters that generate the most unfavorable pathways. Eventually, we combine the associated scenarios of these most unfavorable pathways into a set of scenarios that will be used later in the final scenario selection step (see Section 3.3.3).

Here, we explain how the time series clustering works in detail. As a first step, we need the data from the open exploration process. For each metric, we require to compute the pairwise similarity of the time series data. This pairwise similarity can be computed with the complexity-invariant distant (CID) method as it performs well when it comes to identifying tipping points in complex non-linear systems (Steinmann et al., 2020). The algorithm is presented and devised by Batista et al. (2014). After computing all pairwise distances between the pathways for a given metric, we apply agglomerative hierarchical clustering on the distances. The result of this clustering process is the average silhouette widths for different numbers of clusters (Rousseeuw, 1987). A comparison between the average silhouette widths helps to decide on how many clusters to consider for the selection. The average silhouette width represents the clustering quality for all observations of a cluster. This is complemented by a visual inspection of the resulting clusters.

3.3.2. Directed Scenario Search

Directed scenario search can be considered as an alternative method to time series clustering. Both methods can be used to identify vulnerable scenarios. We make use of the ϵ-NSGA-II algorithm as introduced in Section 3.1.3. For each problem formulation, we run an optimization process while sampling from the lever space, and trying to find those scenarios that optimize the corresponding metrics. As we are looking for the worst-case scenarios, we need to flip the optimization direction of the metrics. The worst-case scenarios for each problem formulation are eventually combined into one big set of worst-case scenarios.

3.3.3. Scenario Subsetting

Combining the results from time series clustering and directed scenario search, we have identified a set of worst-case scenarios. If we would theoretically use this merged set of worst-case scenarios and discover policies in Section 3.4 for each of them, the resulting policies would likely be more diverse and robust. However, computationally, this becomes infeasible very quickly. If the run of an optimization process takes hours to complete, running thousands of optimizations due to the high number of reference scenarios, could take decades. Naturally, we need to deal with this situation by reducing the number of scenarios. Eker and Kwakkel (2018) suggest a systematic approach to this selection process in which both, policy relevance and scenario diversity are considered. For this purpose, they find a small number of scenarios that maximizes their diversity. This selection is based on a diversity criterion as defined by Carlsen et al. (2016). In this context, we would need to find all combinations of four candidate scenarios and compute the distances between the four scenarios. To avoid a combinatorial explosion, we will sample the sets of combinations. In fact, out of 15 trillion possible combinations, we sample only 10 million. Eventually, our process yields four reference scenarios that represent a diverse set of worst-off scenarios. This set can then be used for the policy discovery process in Section 3.4.

3.4. Policy Discovery

Now, that we have our four reference scenarios, we can use the ϵ-NSGA-II algorithm to find optimal policies. We explore our eight problem formulations. As the algorithm uses stochasticity to sample candidate solutions, there is a possibility that the resulting Pareto-optimal solutions are sensitive to the used random seed. To counter this potential dependency, we use two different seeds and compare additionally whether the Pareto-optimal pathways differ indeed. If they would differ, we would need to use a higher number of seeds for further optimization runs. However, for now, we have to run optimizations with 8 problem formulations, 4 reference scenarios, and 2 seeds, i.e., $8 \cdot 4 \cdot 2 = 64$ optimization runs. Eventually, we merge all found policies per problem formulation and investigate the resulting pathways. In order to assess whether the optimizations have converged, we use the convergence metric ϵ-progress which is used to evaluate the accuracy level of the found policy in the Pareto-approximate set (P. M. Reed et al., 2013). It is a measure that tracks how many new candidate solutions have been added to the set of non-dominated solutions over the number of function evaluations. A flattening curve suggests convergence.

Eventually, we re-evaluate the performance and robustness of the found policies under uncertainty. For this purpose, we make use of diverse robustness metrics according to the decision criteria as described by McPhail et al. (2018). For simplicity's sake, we consider only three KPIs which are also the most shared objectives across all problem formulations. Following the categorization by McPhail et al., we identify a) what system performance the KPI belongs to, b) whether we the KPI has absolute or relative performance values, and c) whether we are relatively risk averse with respect to the KPI. The KPIs, their decision criteria values, and the eventually chosen robustness metrics are shown in Table 3.1 below.

Table 3.1: Robustness Metrics

The KPI names are in the first column. The middle three column represent the decision criteria to determine the final column which is the resulting robustness metric.

KPI	system performance	performance values	relative risk aversion	robustness metric
welfare	optimize	absolute	low	Hurwicz's optimism-pessimism rule
welfare loss	optimize	relative	high	90th percentile minimax regret
atmospheric temperature	satisfactory	absolute	high	Starr's domain criterion

While we want to maximize `welfare` and minimize `welfare loss`, we do not need to necessarily optimize atmospheric temperature. A *good enough* value or range is sufficient in order to avoid catastrophic repercussions. The *performance values* refer to whether it is better to think about the absolute performance of a policy on a metric or whether it makes more sense to consider its relative performance to the alternative policies. For `welfare` and `atmospheric temperature`, we care about the absolute performance values, while for `welfare loss`, we could consider either. Eventually, pick the *relative* performance value. And finally, with respect to the *relative risk aversion*, we reason that we have a relatively low risk aversion for `welfare`. Although, it would not desirable to perform low on `welfare`, it is still okay to accept such performances. For `welfare loss` and `atmospheric temperature`, we exhibit a relatively high risk aversion. High atmospheric temperature increases are not only undesirable but could be catastrophic and cause many deaths. We want to err on the side of being rather risk averse. The same logic applies to `welfare loss`.

4

Results

4.1. Scenario Selection

According to the process that we describe in Section 3.3, we identify 4’350 worst-case scenarios through time series clustering and 59 worst-case scenarios through directed scenario search. Eventually, we merge both sets and select four scenarios that are approximately maximally diverse and representative of all worst-case scenarios. These scenarios are then our four reference scenarios. The intermediate steps and results of time series clustering and directed scenario search can be found in the appendix under Section C.3.1 and Section C.3.2. Here we focus on presenting the final four reference scenarios.

The merged scenarios from time series clustering and directed scenario search result in 4’351 distinct scenarios. This is interesting as 58 out of 59 vulnerable scenarios from directed scenario search are also in the set of the 4’350 vulnerable scenarios from time series clustering. After maximizing the diversity of a set of four scenarios, we yield the set of our four reference scenarios[1]. Let us put the reference scenarios into context. Figure 4.1 depicts a pair plot of a sampled set of objectives[2]. Gray dots represent scenarios of all 30’000 scenarios that have been explored during open exploration (see Figure C.2 for the corresponding pathways in the appendix). The selected reference scenarios are colored diamonds. We observe that the reference scenarios are not always well distributed. For example, in the panels for `welfare` versus `welfare loss`, we can see that reference scenarios 1, 2, and 3 are very close together while reference scenario 0 is at a great distance. In case of the bottom-left four panels, we can see that the reference scenarios are more equally distributed across the spanned subspace. This sampled set of objectives is representative of how well the reference scenarios are distributed for all other objectives. This is likely a limitation of choosing only four reference scenarios and using eleven dimensions in the objective space. More considerations follow in the limitations (Section 5.2). Using these four reference scenarios, we proceed with the subsequent policy discovery process.

4.2. Policy Discovery

Now that we have identified our reference scenarios, we can use directed policy search to find optimal outcomes and their corresponding policies. In this section, we elaborate on the convergence behavior of directed policy search (Section 4.2.1), its sensitivity to seeds (Section 4.2.2), the resulting and se-

---

[1] See in the appendix, Figure C.7 to view the corresponding parallel axes plot.

[2] In order to view a pair plot that contains all objectives, follow this link to the image on GitHub. It is not included in the appendix because the resolution of one A4 page is not sufficient to spot all the details.

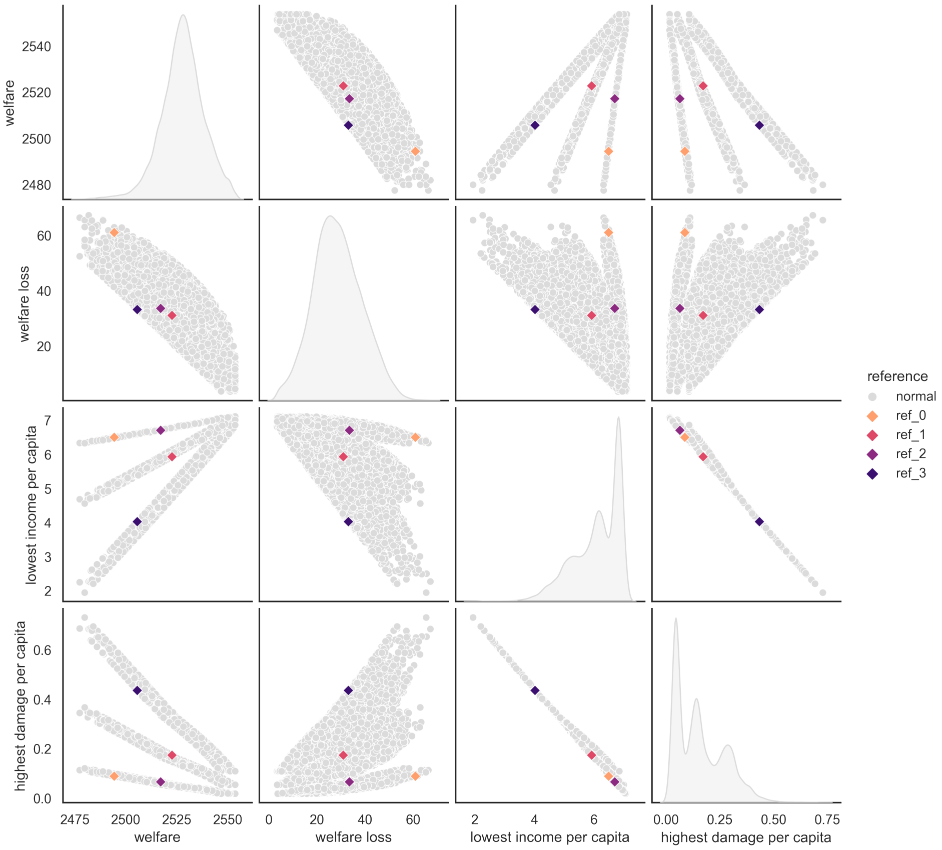


Figure 4.1: Reference Scenarios in Context: A Pair Plot

This figure shows a subset of pair plots for four objectives and the location of how the four chosen reference scenarios perform in this objective subspace.

lected policies (Section 4.2.3), a high-level comparison between the problem formulations' outcomes (Section 4.2.4), the resulting pathways (Section 4.2.5), the trade-offs (Section 4.2.6), and the policies' robustness (Section 4.2.7).

4.2.1. Convergence

We run 32 optimizations with 8 problem formulations, 4 reference scenarios, and 2 seeds. The convergence behavior for a selected subset of the problem formulations can been seen in Figure 4.2. An overview of all problem formulations can be found in Figure C.18. The results for the remaining problem formulations resemble the here shown ones. As we can see, the curves are generally saturating, we can conclude that the optimizations are converging. We can see though that the utilitarian aggregated formulation for reference scenario 1 and 3 do show a slight difference in converging behavior for the seeds. The optimization with seed 1 has in both cases a substantial amount of ϵ-progress. Consequently, we might get marginally better solutions if we would increase the `nfe` (number of function evaluations) further for the optimizations with utilitarian aggregated problem formulation.

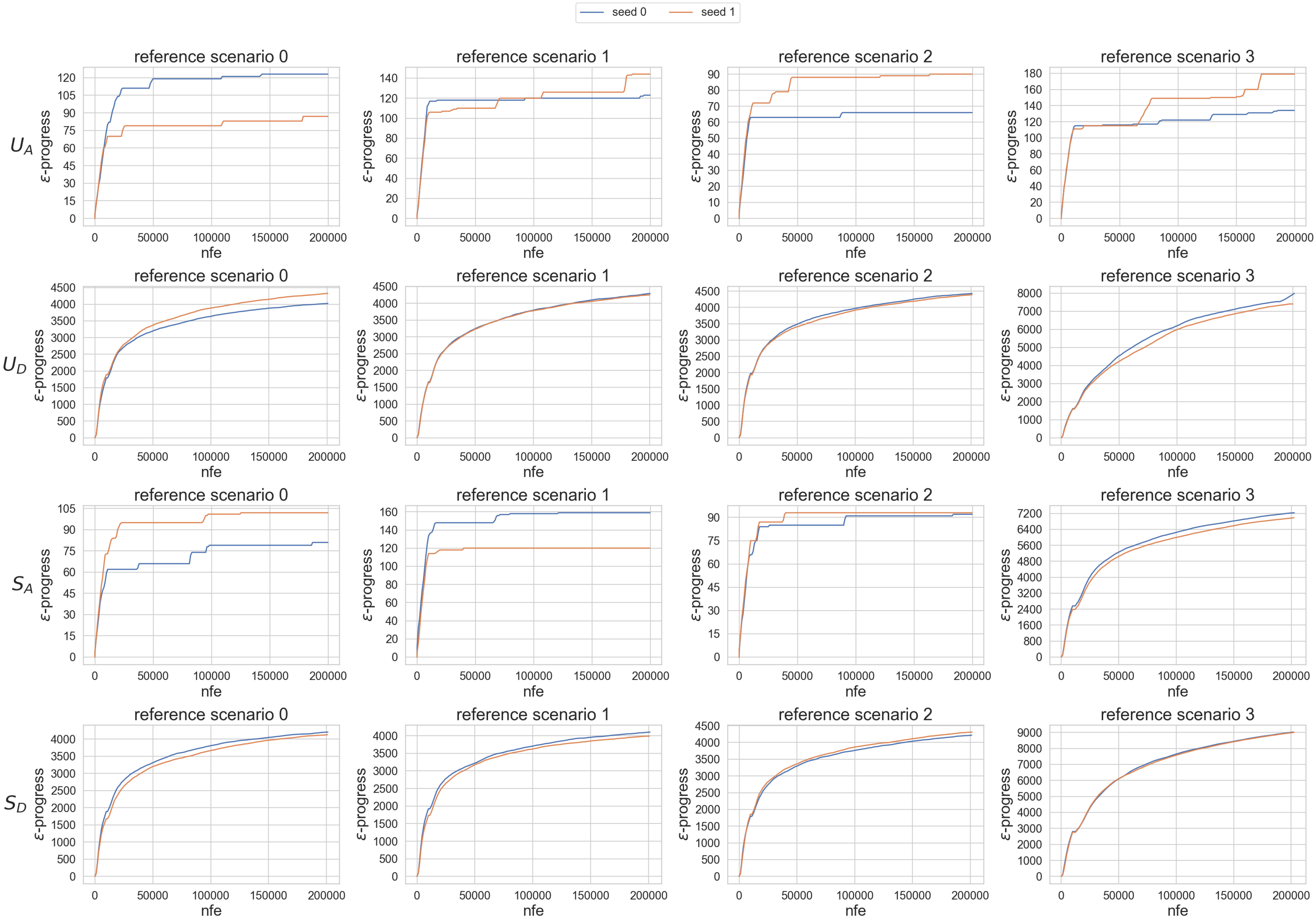


Figure 4.2: Convergence of Directed Policy Search via ϵ-Progress (Subset)

Each row represents a problem formulation. Each column represents the corresponding reference scenario that has been used during the optimization. Each subplot contains two lines for which a different seed has been used.

Summary of Convergence

All 32 optimization processes exhibit an acceptable degree of convergence with respect to ϵ-progress.

4.2.2. Sensitivity to Seeds

In order to check whether the optimization setup is sensitive to a chosen seed, we select two different seeds. Given a comparison of the Pareto-optimal pathways, we conclude that the optimization setup is not sensitive to stochasticity. As this analysis does not contribute to the main story line of our research, we refer to a detailed analysis in the appendix (see Section C.4.1).

4.2.3. Policy Selection

The optimization processes yield mostly big numbers of solutions. The third column in Table 4.1 shows these numbers. The utilitarian aggregated problem formulation stands out as we find only 12 solutions for 8 optimization runs. With 4 reference scenarios and 2 seeds, that is usually barely more than one solution per optimization run. The other problem formulations yield clearly more. This is likely connected to the number of objectives and their data types. The utilitarian problem formulation has only two objectives which are `welfare` and `temperature overshoots`. The first objective is of type float while the second one is of type integer with a small range. Usually, temperature overshoots in the year 2105 yield values such as `5, 6, 7`. The leading objective is probably `welfare` here. Even very low ϵ-values do not yield more solutions. Thus, the nature of this problem formulation's objective space is a fairly simple one. Furthermore, the egalitarian disaggregated problem formulation yields a very high number of found policies with $53'124$. This is especially due to the very small ϵ-value for `damage Gini`.

Table 4.1: Directed Policy Search: Overview of Found and Selected Policies

Problem formulations are all listed in the first column. The remaining columns display the number of associated objectives, the number of found policies (through directed policy search), and the number of selected policies (through non-dominated sorting in post).

problem formulation	# of objectives	# of found policies	# of selected policies
U_A	2	12	12
U_D	3	4528	15
S_A	4	1810	12
S_D	7	4913	14
E_A	3	474	18
E_D	5	53124	12
P_A	4	32	14
P_D	7	569	16

Eventually, we apply nondominated sorting by increasing the ϵ-values successively. In order to be able to investigate at a manageable number of policies, we adjust the ϵ-values in a way, such that we arrive at a set size of 12 to 18 policies per problem formulation. What exact ϵ-values are used for this, can be found in this README.md. For future research, we recommend to run optimization processes with these ϵ-values to reduce compute time. Figure 4.3 offers a visualization of all Pareto-optimal policies and the selected Pareto-optimal policies.

The limits of all axes are adjusted such that they are shared between all parallel axes plots. This facilitates better comparability. We can observe that the policies exhibit a very similar pattern for the utilitarian and sufficientarian problem formulations (blue and green). The similarity is especially bigger between U_A and S_A on the one hand and U_D and S_D on the other. All of these four problem formulations have a minimal `initial rate of social time preference for consumption` of around `0.001` which is the minimum that has been explored for this variable in the optimization setup. This is due to the fact that that a decrease in this variable increases the computed `welfare`. Because `welfare` is a dominant objective in the utilitarian and sufficientarian problem formulations, it makes sense that the lever `initial rate of social time preference for consumption` would turn out as small as possible. For U_A and S_A, we can also see that the savings rate is concentrated between 23% and 25% while there is a bigger spread for all other problem formulations (10% to 25%). Moreover, P_A shares a similar savings rate to the utilitarian and sufficientarian problem formulations. E_A exhibits a `net-zero`

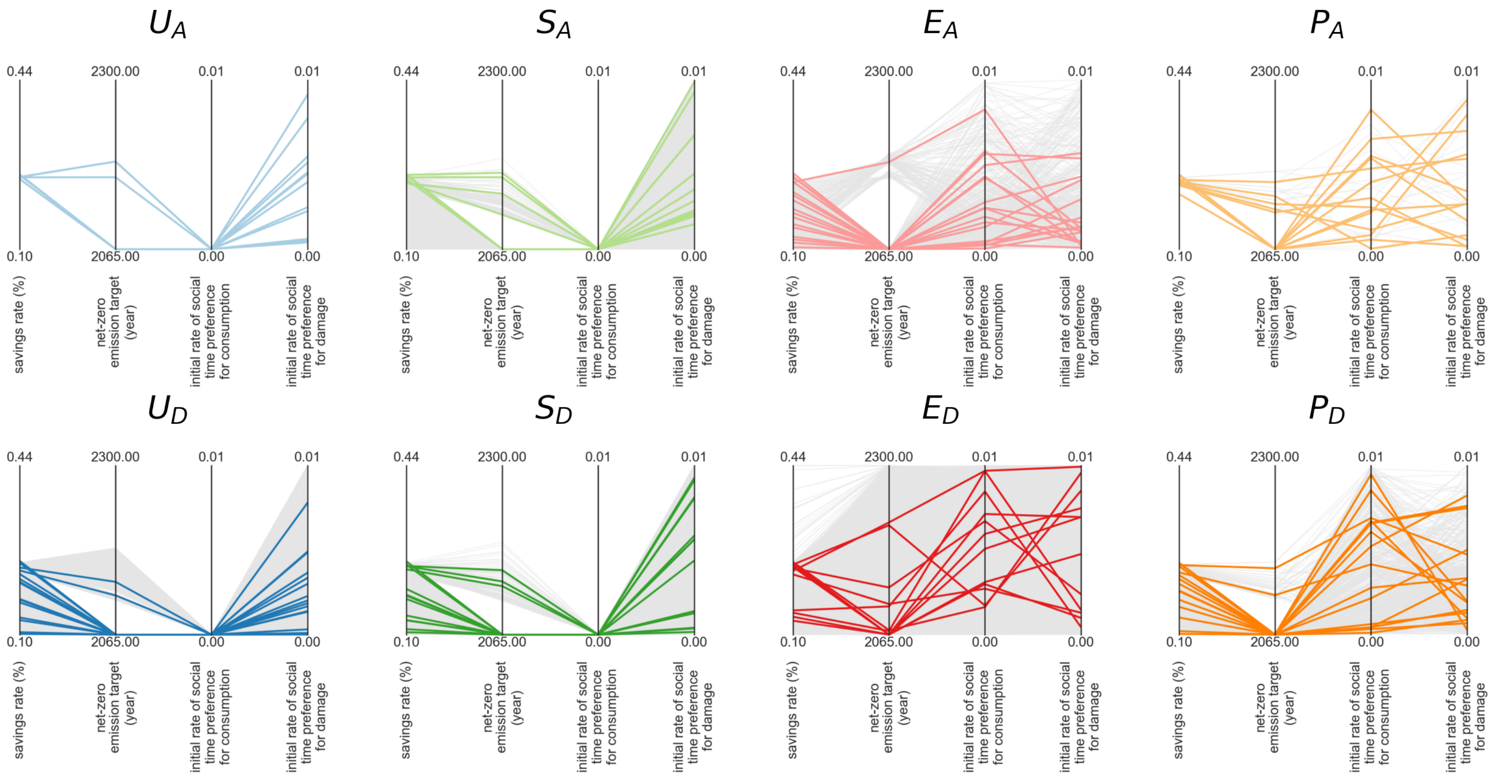


Figure 4.3: Selected Pareto-Optimal Policies

Parallel axes plots for each problem formulation in lever space. Gray lines represent all optimal policies. The colored lines represent the selected optimal policies. There is no desired direction on the axes.

`emission target` of mostly 2065 which is the considered minimum for this model (only once is the year 2186) part of the Pareto-optimal solutions. This makes E_A the problem formulation that yields policies with the consistently earliest `net-zero emission target`. Interestingly, E_D yields some of the latest `net-zero emission targets` with years going up to 2221.

Summary of Policy Selection

The utilitarian and sufficientarian policies resemble each other very strongly while they resemble each other even stronger within the same aggregation level. The biggest difference between utilitarian and sufficientarian problem formulations on the one hand and egalitarian and prioritarian problem formulations on the other hand, is the `initial rate of social time preference for consumption` which is minimal for the former and widely spread for the latter.

4.2.4. High-Level Comparison

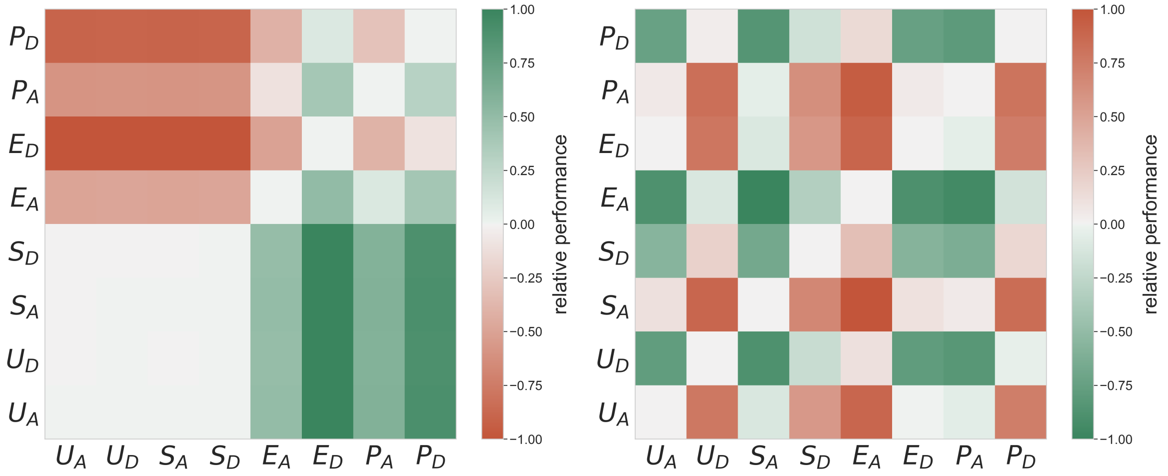
Let us face a subset of the outcomes of the found policies and compare them. What we want to explore, is how the median outcomes compare pairwise between the optimal policies of two problem formulations. For instance, how does the performance of the optimal policies of U_A compare against the one of U_D. For this purpose, we sample 50 scenarios from the set of all vulnerable scenarios (see scenario selection). Then, we run all selected Pareto-optimal policies on the same 50 sampled vulnerable scenarios. Using the same scenarios improves comparability. Let us consider the results for `welfare` and `welfare loss`. For each problem formulation, we have `number of its Pareto-optimal policies` times `50 scenarios`. Next, we compute each problem formulation's median for these two variables. An outcome score $s_{o,i,j}$ between two problem formulations i and j is calculated in the following way:

$$s_{o,i,j} = \frac{\tilde{o}_i - \tilde{o}_j}{max(\{(\tilde{o}_x - \tilde{o}_y)\ \forall x, y \in PF\})} \tag{4.1}$$

where

$s_{o,i,j}$	outcome score between two problem formulations i and j for outcome o
$\tilde{o}_a$	median of outcome o for problem formulation a
PF	set of all problem formulations

The denominator is used to normalize the number such that the range for $s_{o,i,j}$ is $[-1, 1]$. Figure 4.4a and Figure 4.4b show the relations between the problem formulations' median outcomes for `welfare` and `welfare loss`, respectively.



(a) Welfare

(b) Welfare Loss

Figure 4.4: Relation between Problem Formulations for Median Welfare and Welfare Loss

A grid cell represents the relation between a row's and a column's problem formulation. Green represents that the row's problem formulation is more desirable than the column's one.

For example, in Figure 4.4a, the top-left cell is $s_{welfare,P_D,U_A}$ and represents how well P_D's optimal policies perform on the variable `welfare` against U_A's optimal policies under the same scenarios. In this case, it is intensely red which means that P_D is performing very badly compared to U_A with respect to `welfare`. We can also follow from this that U_A is performing very well compared to P_D. In fact, the scores are mirrored and inverted on the diagonal from bottom-left to top-right being the mirror axis.

Let us look first at `welfare` in Figure 4.4a. We can observe that there is virtually no difference between utilitarian and sufficientarian problem formulations. Moreover, because the bottom-right block is completely green, we can conclude that utilitarian and sufficientarian problem formulations dominate egalitarian and prioritarian problem formulations with E_D and P_D performing particularly badly. Consequently, in the case of `welfare`, the selection of the ethical premise has clearly a higher impact than the aggregation level. However, the effect of the aggregation level has still a clear pattern as we can see that the disaggregated problem formulations for E and P perform clearly worse than the aggregated problem formulations. All utilitarian and sufficientarian problem formulations perform virtually the same on `welfare`. Additionally, the egalitarian and prioritarian problem formulations are associated with Pareto-optimal policies that cause more redistribution of resources. In fact, if we look at the particular egalitarian objectives, we can see that for the `consumption Gini`, P_D and E_A perform very well which means that the optimal policies of these two problem formulations yield higher equality of consumption across the globe (see Figure C.10a). And the optimal policies of P_A and E_D yield a higher equality of climate damages than any other problem formulations (see Figure C.10b). Hence, we can find more equality for egalitarian and prioritarian problem formulations but higher `welfare` for utilitarian and sufficientarian problem formulations.

While `welfare` is dominated by the choice of the ethical premise, `welfare loss` is dominated by the aggregation level. Let us analyze `welfare loss` in Figure 4.4b. We can observe that the disaggregated problem formulations usually outperform the aggregated ones as the rows for P_D, S_D, and U_D are mostly green. E_A and E_D are the exception. In fact, according to the saturated green in E_A's row, E_A performs the best with respect to `welfare loss`. E_A has the objectives `welfare`, `temperature overshoot`, and `consumption Gini`. Thus, there is no objective that is directly related to damages or `welfare loss`. And although, E_D's objectives extend E_A's objectives by `welfare loss` and the `damage Gini`. Nevertheless, E_D performs among the worst problem formulations with respect to `welfare loss`. When we look at the `damage Gini` for E_D in Figure C.10b, we can see that E_D is actually performing very well although `welfare loss` performs relatively bad. Overall, most KPIs follow the pattern of `welfare loss` in terms of aggregation level being the more impactful factor. Among these KPIs are the following and can be found in the appendix:

- `consumption Gini` (Figure C.10a)
- `damage Gini` (Figure C.10b)
- `maximum distance to damage threshold` (Figure C.12a)
- `population above damage threshold` (Figure C.12b)
- `lowest income per capita` (Figure C.11a)
- `highest damage per capita` (Figure C.11b)
- `economic damages` (Figure C.13a)
- `GWP` (Figure C.13b)
- `increase in atmospheric temperature` (Figure C.14)

Summary of High-Level Comparison

For `welfare`, the level of aggregation has an impact, however, the ethical premise has clearly a higher impact. For most other KPIs, the level of aggregation has clearly a greater impact with the notable exception of E_A and E_D. Overall, we can conclude that the level of aggregation is more dominant than the ethical premise when it comes to the median performance of most KPIs.

4.2.5. Pathways

After acquiring a high-level overview by comparing relative performances of KPI medians, let us have a closer look at global pathways[3]. Here, we want to discuss the pathways for three central objectives: `welfare`, `welfare loss`, and `atmospheric temperature increase`. In Figure 4.5, we can see that the `welfare` pathways are consistent with the results that we obtained from comparing the medians. Here, we can observe that for utilitarian and sufficientarian problem formulations, we get virtually no variation across different scenarios. Pathways for E_A, P_A, and P_D resemble each other strongly with only small variations. Again, E_D sticks out with being more different. In fact, optimal policies of E_D are never

---

[3] We have also analyzed regional pathways which can be found in the appendix under Section C.4.4.

reaching `welfare` levels as high as for any of the other problem formulations' optimal policies. We can confirm here that the choice of ethical premise has the greater impact on the pathways.

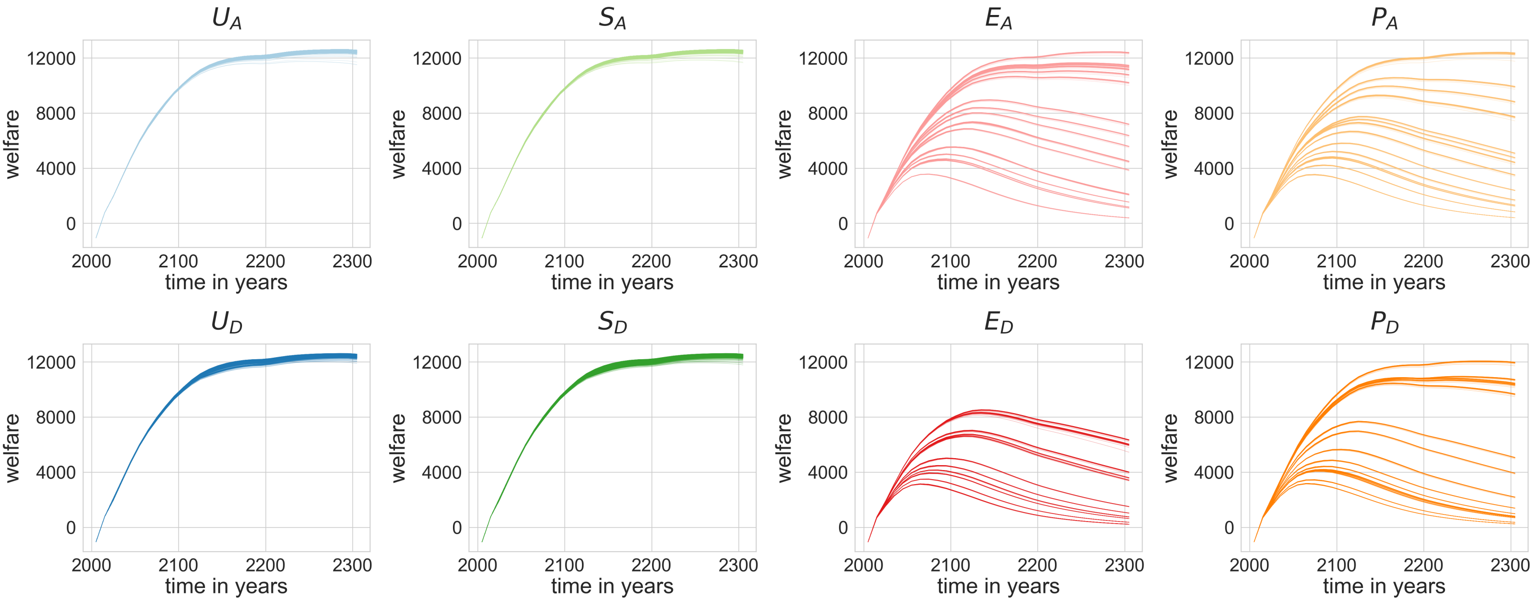


Figure 4.5: Global Pathways for Welfare

Each panel represents the results given the Pareto-optimal policies of a problem formulation in combination with 50 vulnerable scenarios. Each combination (policy x scenario) is one line in a panel.

The pathways for `welfare loss` are consistent with the findings from comparing medians, too. We can observe in Figure 4.6 that the welfare loss is steadily increasing for all problem formulations. However, aggregated problem formulations exhibit `welfare loss` more frequent higher increases. We can see that most of the disaggregated problem formulation panels (bottom row) do not show high and thus undesirable `welfare loss` as the highest values do not reach 600s while the panels for aggregated problem formulations (top row) reach the 600s and even 700s. A notable exception is again E_D which exhibits some pathways that even approximate a value of 800 which represents the largest visible `welfare loss`. In the case of `welfare loss`, we can conclude again that the level of disaggregation has usually a bigger impact on the pathways than the ethical premise.

Finally, let us inspect the pathways for `atmospheric temperature increase` in Figure 4.7. We can observe a general trend of pathways having lower `atmospheric temperature increases` for disaggregated problem formulations (lower row). The disaggregated problem formulations show lower maximal values and their pathways are stronger concentrated in the sub-2.5°C-area. Again, exceptions are E_A and E_D. This is a common thread throughout the analysis. Having mentioned this, we can conclude again that the level of aggregation has a stronger impact on the pathways than the ethical premise[4].

Summary of Pathways

Global pathways for `welfare` are strongly determined by the ethical premise. For `welfare loss` and `atmospheric temperature increase`, and `industrial emissions`, aggregation is the predominant factor on the global pathways.

---

[4] For additional pathways on global industrial emissions, see Figure C.15.

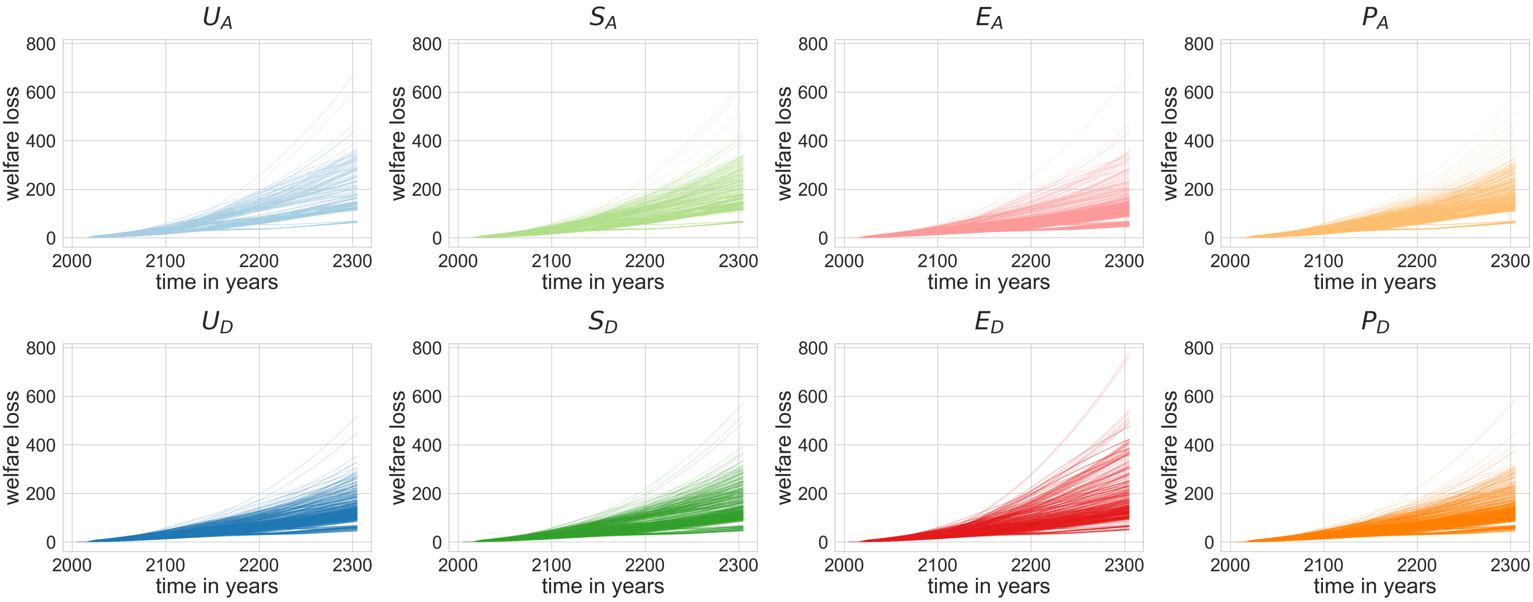


Figure 4.6: Global Pathways for Welfare Loss

Each panel represents the results given the optimal-policies of a problem formulation in combination with 50 vulnerable scenarios. Each combination (policy x scenario) is one line in a panel.

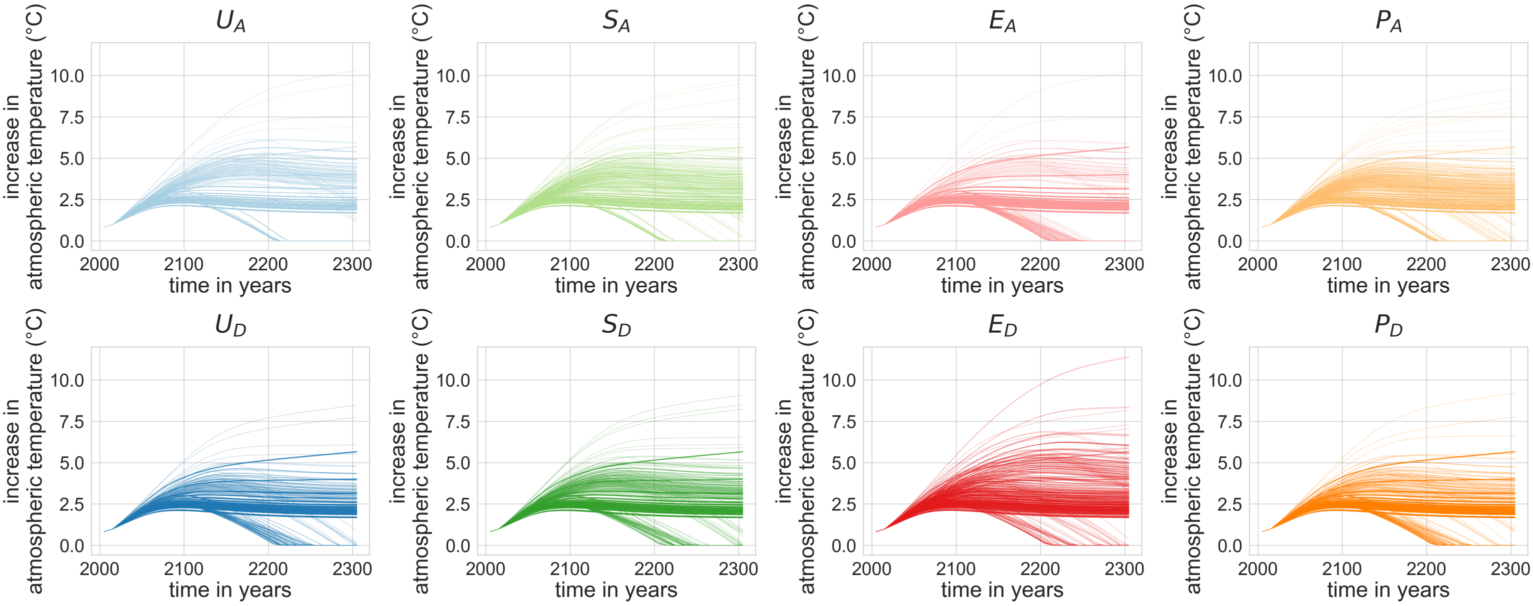


Figure 4.7: Global Pathways for Atmospheric Temperature Increase

Each panel represents the results given the optimal-policies of a problem formulation in combination with 50 vulnerable scenarios. Each combination (policy x scenario) is one line in a panel.

4.2.6. Trade-Offs

Figure 4.8 depicts the trade-offs between four KPIs for all eight problem formulations. The KPIs are `welfare`, `welfare loss`, `atmospheric temperature`, and `industrial emission`. The limits for each KPI (top and bottom) are shared by all problem formulations to enable better comparability.

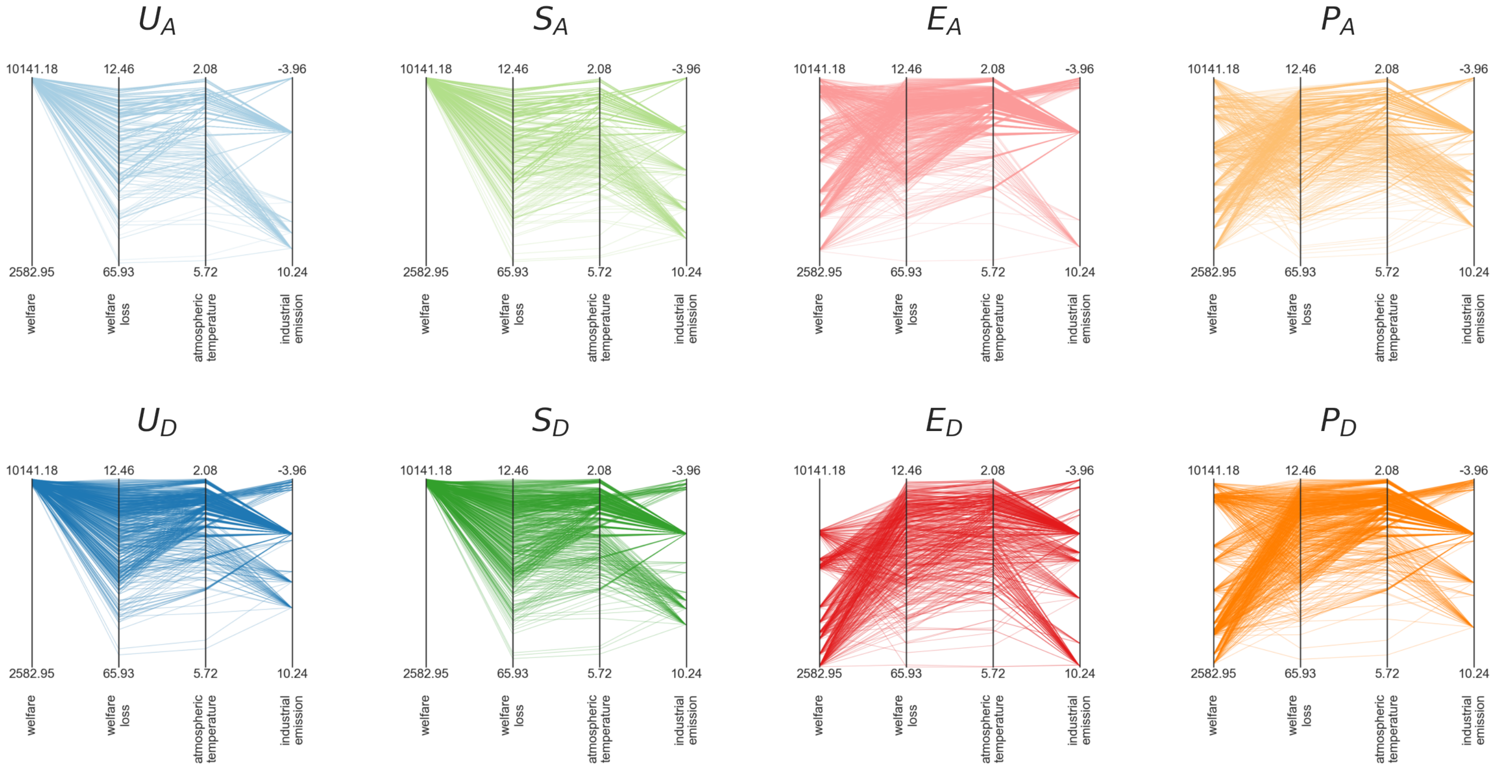


Figure 4.8: Trade-Offs: Welfare, Welfare Loss, Atmospheric Temperature, and Industrial Emissions

Each panel represents the results given the optimal-policies of a problem formulation in combination with 50 vulnerable scenarios. Each combination (policy x scenario) is one line in a panel. On each axis, top values are more desired than bottom values. Each objective refers to the year 2105. Desired values can be found at the top of the axes.

At first glance, we can directly make out that the utilitarian and sufficientarian trade-off patterns look very much alike. And the egalitarian and prioritarian trade-offs share a strong resemblance as well. Therefore, simply based on these observations, the ethical premise has a very strong effect on the trade-offs. If we look closer, we can also perceive that trade-offs patterns look even more similar between U_A and S_A, and between U_D and S_D. Consequently, aggregation plays also an important role. Interestingly, while for utilitarian and sufficientarian problem formulations, there is always a maximum `welfare` level, we can see that there is a maximum spread across `welfare loss` which entails that a high degree of `welfare` can also have high and low degrees of `welfare loss`. For the egalitarian and prioritarian problem formulations, we can also observe high `welfare` values being connected to a wide range of `welfare loss`. However, if we look at the bottom of the axes for the egalitarian and prioritarian problem formulations, we can see that the most undesirable `welfare` values are connected to the more desired `welfare loss` values. For egalitarian and prioritarian problem formulations, we can thus see an inverse proportional relationship between the desirability of `welfare` and `welfare loss`.

Summary of Trade-Offs

The choice of the ethical premise has the strongest impact on the patterns of the trade-offs as they split the problem formulations in two bigger groups – group 1: U_A, U_D, S_A, and S_D, and group 2: E_A, E_D, P_A, and P_D. However, the aggregation level has still a perceivable impact within these two groups.

4.2.7. Robustness

We have seen how pathways and trade-offs are affected by the ethical premise and the aggregation level of the problem formulations. In this last section, we want to investigate their influence in different kinds of scenarios[5]. Let us look at the performance of the optimal policies on different sets of scenarios. Figure 4.9.

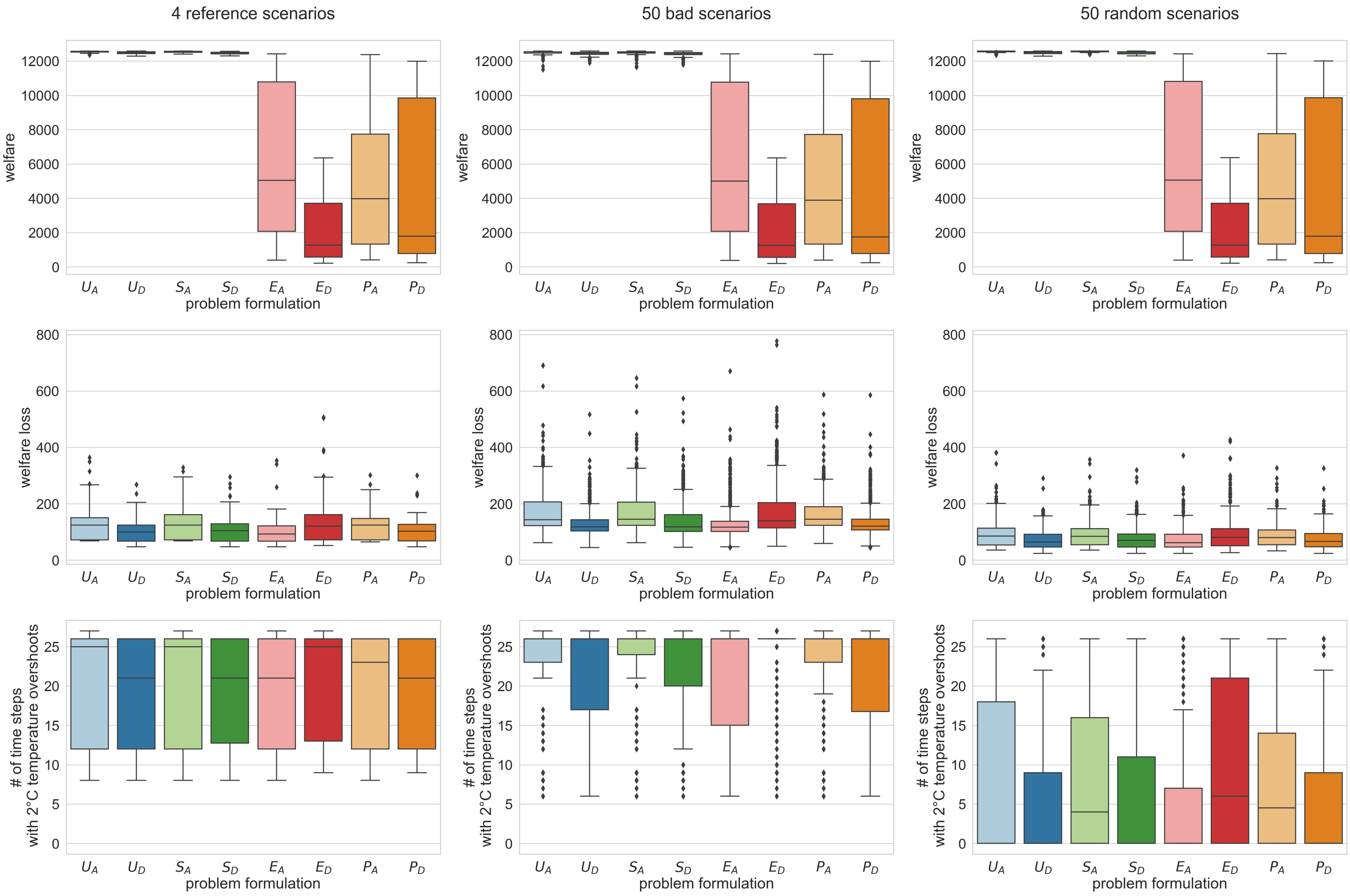


Figure 4.9: Performance under Various Scenario Sets: Reference, Vulnerable, and Random

The rows represent KPIs at the year 2305. The columns represent with what scenarios the experiments have been run in order to obtain the KPI values. The box plots refer to the 8 problem formulations. For `welfare`, higher values are considered to be better. For the other two KPIs, lower values are considered more desirable.

Let us have a look at the results for the objectives that most problem formulations share: `welfare`, `welfare loss`[6], and `temperature overshoots`. Figure 4.9 shows the results for the year 2305. This year are chosen because all three variables are considering the information of their entire path and can be considered cumulative functions. For example, `temperature overshoots` is literally just counting

---

[5] Additionally we also conduct a separate robustness analysis including various robustness metrics. In order to keep the main story line straight forward, we put this particular analysis into the appendix under Section C.4.5. The relevant results are that the robustness trade-offs are stronger determined by the choice of ethical premise, however, the aggregation level has a noticeable effect within subgroups of problem formulations.

[6] (The metric `welfare loss` is only shared by the disaggregated formulations.

how many 2°C overshoot time steps there have been up to the year 2305. The other two variables consider a discounted sum of `welfare` or `welfare loss` from 2005 to 2305, respectively.

Previous plots have only considered our selected Pareto-optimal policies combined with *50 bad scenarios*. To gain more insights, we combine the policies once with the *4 reference scenarios*, and once with *50 random scenarios*. We can see that for example for `welfare`, we get very similar results when looking across all scenario sets (see first row). The policies perform slightly worse under the 50 bad scenarios when we compare against the reference and the random scenarios. E.g., the lower whiskers and outliers are a bit lower for utilitarian and sufficientarian problem formulations. But the differences are minimal. The differences between the scenario sets are slightly more pronounced for `welfare loss` (second row). Interquartile ranges, whiskers, and outliers are higher in the 50 bad scenarios than for the other two scenario sets. Finally, we look at the last row of Figure 4.9 which depicts the performance on `temperature overshoots`. Here, we can see striking differences between the scenario sets. In the first column, we observe that the median values are high and thus undesirable, ranging from 21 to 25. The interquartile ranges are very wide ranging from 12 to 26 for almost all problem formulations. In the second column, we can see that that medians are very high and undesirable as well. The interquartile ranges are significantly shorter and are more distinct between the problem formulations. The last column shows the performance of Pareto-optimal policies on the set of completely random scenarios. We can discern that the medians are now very low and desirable, ranging from 0 to 6. Interquartile ranges share this pattern.

The found Pareto-optimal solutions perform well on the scenarios that we optimized them for (4 reference scenarios). They perform similar or very well for random scenarios as they provide more favorable circumstances. However, it looks more bleak for the 50 bad scenarios as they are more unfavorable and potentially far away from the 4 reference scenarios in uncertainty space. Except for the KPI of `welfare`, the choice of ethical premise has little effect compared to the choice of whether to aggregate or disaggregate utility and disutility. Thus, this inference does not only hold for the year 2105 is presented in our former analysis but also throughout the entire time horizon from 2005 up to 2305. E_A and E_D are again the notable exceptions. The level of aggregation shows are strong effect on the outcomes, even across different kinds of scenario sets.

Summary of Robustness

The previously identified patterns hold true for different sets of scenarios (reference, bad, and random). The level of aggregation has mostly the biggest effect on the outcome distributions with `welfare` being a special case in which the ethical premise is pre-dominant.

5

Discussion

On the basis of the presented results in the previous chapter, we discuss our key findings (Section 5.1) and showcase the limitations of our research (Section 5.2).

5.1. Key Findings

We identify four key findings which we present in the subsequent sections. In order to put our findings into context, let us reiterate our research angle by restating the main research question:

Research Question

How are Pareto-optimal climate abatement pathways affected by the disaggregation of utility and disutility in alternative ethical problem formulations when using an integrated assessment model under deep uncertainty?

So, what do we find when we explore this question?

5.1.1. Dominance of Aggregation Levels over Ethical Premises

Our disaggregation of utility and disutility in alternative ethical problem formulations has a significant impact on the Pareto-optimal climate abatement pathways. This is in line with Kulkarni's findings on a utilitarian aggregated problem formulation of the DICE model (2020). Beyond this, our most paramount finding is that the aggregation levels clearly dominate the ethical premises on almost all KPIs. In other words, disaggregating an ethical premise (e.g., U_A to U_D) has a stronger affect on the Pareto-optimal abatement pathways than changing the ethical premise (e.g., U_A to P_A). This applies to most KPIs such as `welfare loss`, `atmospheric temperature increase`, `industrial emissions`, and many more. This should not be a huge surprise as for example, `welfare loss` is based on damages and disaggregated problem formulations contain damage-related objectives. Consequently, disaggregation positively affects the performance on e.g `welfare loss`. Disaggregated problem formulations yield Pareto-optimal policies that outperform their aggregated counterparts on almost all KPIs but on `welfare`[1]. For `welfare`, the level of aggregation has an impact on the pathways and trade-offs, however, the ethical premise has clearly a higher impact. In fact, for this KPI, we can infer that that the utilitarian and sufficientarian Pareto-optimal policies dominate the egalitarian and prioritarian ones.

---

[1] A notable exception are the egalitarian problem formulations which are discussed in the ensuing key findings.

Why is this the case? Every single problem formulation includes `welfare` as one of their objectives. U_A and U_D have the lowest number of objectives (2 and 3, respectively) which could be interpreted as `welfare` being more prioritized when trying to find the Pareto-front during the optimization process. Furthermore, in our analysis, the sufficientarian objectives turn out do not add much to the shared utilitarian objectives. This entails that there are no trade-offs between purely utilitarian and the additional sufficientarian objectives. All utilitarian and sufficientarian problem formulations perform virtually the same on `welfare`. In contrast, the egalitarian and prioritarian problem formulations are associated with Pareto-optimal policies that cause more redistribution of resources.

5.1.2. Correlation between Low Welfare and High Welfare Loss

Our results indicate that the negative correlation between `welfare` and `welfare loss` is not as evident as the findings of Marangoni et al. (2021) suggest. In our case, we have to make two distinctions: one for utilitarian and sufficientarian problem formulations and one for egalitarian and prioritarian problem formulations. For the former, high `welfare` is connected to a wide range of `welfare loss` (low and high values). For the egalitarian and prioritarian problem formulations, it is the case that lower `welfare` values are mostly correlated with lower `welfare loss` values. Higher `welfare` values are again pairable with a wide range of `welfare loss` values. This is not the case for Marangoni et al. (2021). This is strange as there are many parallels otherwise. For example, our correlation between `welfare loss` and `atmospheric temperature increase` mirrors their correlation between damage costs and 2°C temperature overshoots. And in both research designs, similar objectives are shared as can be seen in Table 5.1 below.

Table 5.1: Comparison between Our Objectives and Marangoni et al.'s Objectives

The left column contains the objectives used in most of our problem formulations. Welfare loss is only used in disaggregated problem formulations though. The right column represents the corresponding objectives in Marangoni et al. (2021). In contrast to Marangoni et al., we do not have an objective that corresponds to the mitigation costs.

Our Objectives	Marangoni et al.'s Objectives
welfare	discounted expected utility
2°C temperature overshoots	2°C temperature overshoots
welfare loss	damage costs
–	mitigation costs

Marangoni et al.'s objectives strongly resemble our utilitarian disaggregated problem formulation. The main differences consist in a) our lack of a mitigation costs objective and b) our more elaborated objective `welfare loss` which is based on damage costs but extended to include *relative risk aversion* to damages[2] and a separate *discount rate for damages*. It could be the case that the latter two have a crucial influence in the outcome differences. We treat the relative risk aversion as an uncertainty variable. Depending on the scenario, this variable might be e.g. higher which would in turn reduce the disutility of the damages. Let us imagine two scenarios. In both scenarios, the resulting damages are on the same high level. In the first scenario, the relative risk aversion to damages is low. This would entail a high level of disutility of damages which would imply a high degree of welfare loss. In the second scenario, the relative risk aversion to damages is high. Disutility of damages would be lower now and so would be welfare loss. Although the damages are the same in both scenarios, welfare

---

[2] The relative risk aversion variable is also known as the *elasticity of disutility of damages* or `emdd` for short. Details can be found in the appendix under Equation B.20.

loss can differ widely[3] Furthermore, given the the equation of the endogenous social rate of damage (Equation B.22), strongly discounting future damages might have similar affects. These points could explain the difference between our and Marangoni et al.'s trade-offs of `welfare` and `welfare loss`. The same pattern of dispersion occurs for other problem formulations where the same explanation holds.

5.1.3. Dominance of Egalitarian Aggregated Pareto-Optimal Policies

Throughout our analysis, we see again and again how egalitarian Pareto-optimal policies do not follow the same pattern as the Pareto-optimal policies of the other ethical premises. A second key finding of ours consists in the observation that the egalitarian aggregated Pareto-optimal policies perform generally better than the Pareto-optimal policies of any other problem formulation. This applies to many KPIs such as `welfare loss`, `atmospheric temperature increase`, `industrial emissions`, etc. The only exception is `welfare`. Pareto-optimal policies of the utilitarian and sufficientarian problem formulations perform clearly better than the Pareto-optimal policies of any other problem formulations – including the egalitarian aggregated one. However, among the remaining four problem formulations (E_A, E_D, P_A, and P_D), the egalitarian aggregated one performs best. This insight seems to be at odds with Tjallingii's findings (2021) who states that the prioritarian and sufficientarian Pareto-optimal policies indicate sooner emission reductions when compared to the utilitarian and particularly the egalitarian ethical premise (Tjallingii, 2021, p. 90). If we look at our net-zero emission targets and the global emissions pathways, we can clearly conclude that the egalitarian aggregated problem formulation yields Pareto-optimal policies that result in distinctly sooner targets and actual reductions than any of its problem formulation counterparts. What does cause these conflicting findings? The answer lies probably in the exact choice of the the problem formulations. While Tjallingii focuses purely on Gini indices for various years (2055, 2105, 2155, 2205, and 2305), we consider only one Gini index for the year 2105 and supplement this objective with `welfare`, and `temperature overshoots`. Consequently, we are not taking Tjallingii's pure egalitarian interpretation but a modified one that includes other desirable objectives that go beyond equality. This suggests that a *relaxed* egalitarian problem formulation can turn around the performance from virtually worst to virtually best. This gives us pause for thought. Even within the same ethical premise, some potentially minor change can have a huge impact on the performance of the resulting Pareto-optimal policies. This renders careful specification of the problem formulations a key component in the analysis of equity in IAMs.

5.1.4. Shared Misery via Egalitarian Disaggregated Pareto-Optimal Policies

The last key finding is also related to the egalitarian ethical premise. The Pareto-optimal policies of the egalitarian disaggregated problem formulation perform the worst on virtually all KPIs. These includes `welfare loss`, `atmospheric temperature increase`, and in this case even `welfare`. Usually, `welfare` is an exceptional case for our other findings but in this case, the egalitarian Pareto-optimal policies are consistent throughout most KPIs. The only exception, which the egalitarian disaggregated Pareto-optimal policies perform better on, is `damage Gini`. For this KPI, egalitarian disaggregated Pareto-optimal policies actually perform the best. How can this be explained? The title of this subsection gives away the answer. In order to enable relatively evenly distributed ratios of damages over consumption across the 12 regions of our PyRICE model, higher damages are generally accepted which levels the playing field through shared misery. The performance of this problem formulation's Pareto-optimal policies is very interesting as it stands out like a sore thumb compared to the other disaggregated problem formulations. A key difference between their objectives sheds light on the reason for this observation. Let us have a look at what objectives the disaggregation adds to the problem formulation of each of the ethical premises.

---

[3] For more details on the relationship between these variables, see Equation B.19 and Equation B.20.

Table 5.2: Additional Objectives through Disaggregation

The left column shows the ethical premise while the right column shows those objectives that are added to the aggregated version of the corresponding ethical premise.

Ethical Premise	Additional Disaggregated Objectives
utilitarianism	welfare loss
sufficientarianism	welfare loss, distance to and population above a damage threshold
egalitarianism	welfare loss, damage Gini
prioritarianism	welfare loss, highest damage per capita

Let us put aside `welfare loss` for a moment which is added to each ethical premise. Beyond this, we can see that the sufficientarian ethical premise gets two more objectives. Both of them relate clearly to minimizing some absolute damage-related value – in this case minimizing the maximum distance to the damage threshold and minimizing the number of people that are beyond a damage threshold. For the prioritarian premise, it is similar as we aim to minimize the region's damage per capita which is worst-off. The egalitarian premise adds an objective that does not indicate an inherent direction of the damages. In contrast to the other objectives, equality of damages does not necessarily entail a reduction of damages. We can imagine a loop hole that allows the optimizer to increase the damages to enable their equality across the regions. And this is exactly what is happening. It might be interesting to see what Pareto-optimal policies would be the consequence of leaving out the objective `damage Gini` and add only `welfare loss` instead to the egalitarian disaggregated problem formulation. We would expect the results to be generally more desirable. Furthermore, we find it worth mentioning that the objective `damage Gini` is not based on *pure* damages but on the ratio of damages over consumption. While trying to minimize this objective, the trade-off between between consumption and damages is taken into account. In other words: Even if pure damages turn out to be higher, this is acceptable as long as consumption is also higher in a given region. What would this mean in practise? In order to allow more equality of damages over consumption, regions are allowed to suffer substantially higher damages if their consumption is accordingly high. This circumstance might facilitate further higher overall damages.

5.2. Research Limitations

One researcher's limitations are another researcher's food for thought. In this sense, we want to disclose the inherent limitations of our research and provide hints for future avenues. We present limitations of our model (Section 5.2.1), of our problem formulations (Section 5.2.2), of the used methods (Section 5.2.3), and finally of the narrow focus of distributive justice (Section 5.2.4).

5.2.1. Limitations of the Model

First, let us have a look at the limitations of the PyRICE model. In Section 1.1.3, we have already discussed general problems with modeling per se, with IAMs in general, and with the RICE model. But with the PyRICE model, we have introduced further elements that exhibit limitations themselves. By transforming the original RICE model into a simulation model and extending it further and further, we run the risk of introducing more complexity. For instance, introducing a new variable `welfare loss` with its own equations and rather unexplored uncertainty variable `emdd` (elasticity of disutility of damage), we bring in even more assumptions which implies that we have to be even more careful when it comes to interpreting our results and inferring policy recommendations. The same would hold true for more complex IAMs such IMAGE and MESSAGE.

5.2.2. Limitations of Problem Formulations

Furthermore, we choose very specific assumptions with respect to how we disaggregate the problem formulations for diverse ethical premises. In this context, we select specific versions of the four ethical premises. Different kinds of ethical premises would have been also possible to include in the analysis. And even with the currently chosen ethical premises, we settle on *our* particular understanding of them. Variations might yield very different results in terms of Pareto-optimal pathways as the comparison between our findings and Tjallingii shows (2021). Moreover, for the disaggregated problem formulations, we sometimes use a measure that is based on climate damages only (such as in the utilitarian) and sometimes we use a measure that is based on the ratio of climate damages over consumption (such as for all the remaining ethical premises). A deeper exploration in the differences between these two approaches (either only damages or damages over consumption) is an interesting avenue. This could also shed light on the surprising results for both egalitarian problem formulations. In addition, we do not use *pure* ethical premises. For example, for the egalitarian aggregated problem formulation, we could have focused only on minimizing the `consumption Gini` to reduce inequality. However, we expected that the resulting Pareto-optimal pathways would be less interesting as `welfare` would be completely ignored. The resulting pathways might have yielded very low and utterly unacceptable levels of `welfare` which is again coherent with Tjallingii's results. Due to time and compute constraints, we are unable to try these *purer* problem formulations in our optimization setup. We would strongly expect that the outcome differences between Pareto-optimal policies for ethical premises would be more salient than in our current setup. A similar concern can be raised about the objective `temperature overshoots`. Nowadays, this objective is a more commonly used one and its inclusion in every problem formulation strikes us as reasonable. However, this variable has already a strong effect on potentially decreasing `damages` or increasing`welfare loss`. It would be interesting to run the optimizations without this objective. We would expect that the outcome differences would be even more pronounced between the aggregated and disaggregated problem formulations.

5.2.3. Limitations of the Methods

Our limitations are not limited to the model itself but extend to the used methods. For example, in the context of scenario discovery, we conduct an open exploration with 30'000 random scenarios and the Nordhaus policy. However, other researchers have developed other policy recommendations that are based on their extended version of RICE. In order to enable wider exploration, it would be advisable to conduct such an exploration. The vulnerable subspace, that we would explore and cluster with our method of time series clustering, could look different, yield eventually other reference scenarios, and eventually affect the process of directed policy search. But even with this currently explored space, we have other limitations to list. E.g., in the process of scenario selection, we calculate the distance and diversity of a set of four candidate scenarios. The entire set of possible combinations with four candidate scenarios is above 15 trillion. We are unable to conduct an exhaustive analysis and compute all potential distances. We sample merely 10 million such sets. We have discussed how the reference scenarios are not ideally spread in the uncertainty space (see Figure 4.1). A bigger sample size or other optimization methods might improve the selection of the reference scenarios further. And of course, we do not conduct a scenario discovery because we only went through one iteration of the MS-MORDM process. Further iterations, including further vulnerability analyses, could further solidify the results but we do not consider the lack of further iterations as a serious threat to the validity of our findings. The step of policy discovery follows the step of scenario selection. Here, it might be intriguing to use more MOEA performance metrics in order to assess the quality of the Pareto-approximate set that we get out of the optimization process. Currently, we have only discussed ϵ-progress but one could also consider P. M. Reed et al.'s suggestions (2013):

- absolute failure: the average distance between the reference set and the approximation set
- sensitivity to gaps in the Pareto front: the maximum distance to be translated by the approximation set in order to dominate the reference set
- convergence & diversity: hypervolume

Looking into these, might give us more insights about the differences between the Pareto-optimal policies of the various problem formulations. But again, their use would likely firm up our findings and not pose a threat to their validity.

5.2.4. Flavors of Justice

We have considered limitations of the model and limitations of our methods. But let us zoom out a bit further and consider a potentially deeper issue with this research. Justice is a concept with a very broad range of definitions. For example, we could distinguish the following four flavors:

Table 5.3: Flavors of Justice

Name	Focus
distributive justice	distribution of benefits and costs
procedural justice	decision-making process
recognitional justice	recognizing all the actors and diversity
restorative justice	recovering from past unjust situations

What we tackle in our research is the first flavor of justice – the distributive one. Given our research framework, this choice makes sense. Simulation models offer states over time. We can look at a time point and it investigate its state which represents how particular quantities are distributed. The distribution of benefits and costs, or of `welfare` and `welfare loss` are easily accessible with such simulation models. And although we have not integrated restorative justice into our PyRICE model, there is a way. For example, Mi et al. (2019) have done so to include the notion of *historical responsibility* into the RICE model. However, we would have problems integrating or exploring the other flavors of justice. The PyRICE model does not offer any opportunity to integrate procedural justice. The decision-making process is utterly out of the model's scope. The same holds for recognitional justice. The *actors* are 12 regions but they do not act independently within the model. The levers are set to a global scale and not in the hands of each individual region. Even if the levers would be fine-grained in this way, we have only 12 regions and their quintiles categorized by their income classes. This does not constitute the recognition of all actors or their diversity. Generally, we can see that using models such as PyRICE can be limited to account for justice in the broader sense. We have to conclude that such simulation models can merely be a part of the solution. We also need other complementary tools to advance into the direction of enabling a transition of more climate justice.

6

Conclusion

6.1. Answers

We use this research to look behind the veil of aggregation and figure out whether we find climate justice. For this purpose, we tackle our defined six sub-research-questions. Here, we want to provide a summary of our answers. In Section 2.2.1, we discuss the following question:

Sub-Question 1

What alternative distributive justice principles are suitable contenders to enable more equitable climate abatement pathways?

For this purpose, we identify four prevailing distributive justice principles which are summarized in Table 6.1 below.

Table 6.1: Summary: Ethical Premises

Ethical Premise	Definition
utilitarianism	maximize aggregate wellbeing
sufficientarianism	minimize the number of people below a critical threshold
egalitarianism	minimize inequality of wellbeing
prioritarianism	maximize wellbeing of the worst-off

After delineating these four ethical premises, we stand before the challenge of disaggregating utility and disutility within each ethical premise. For this purpose, we have to answer the second sub-question:

Sub-Question 2

How can utility and disutility be disaggregated in each of the chosen distributive justice principles?

Table 6.2: Summary: Disaggregation

Ethical Premise	Disaggregation
utilitarianism	distinguish welfare and welfare loss with their own and independent risk aversions & discounting where welfare is based on consumption and welfare loss is based on climate damages
sufficientarianism	distinguish a consumption threshold based on a defined poverty line and a threshold that refers to the ratio of consumption to climate damages
egalitarianism	distinguish equality of consumption and equality of climate damages
prioritarianism	distinguish the worst-off income quintile of a region and the worst-off ratio of climate damages to consumption of a region

Table 6.2 contains not all objectives of the corresponding problem formulations. This table has merely the function to show what the key differences are between an aggregated and a disaggregated problem formulation within the same ethical premise. In order to compare the effect of disaggregation and the selection of the ethical premises on climate abatement pathways, we needed to chose the KPIs whose pathways we want to investigate. Some of these objectives are also part of the KPIs but the KPIs are not limited to these objectives only. The next question is:

Sub-Question 3

What are suitable key performance indicators to compare the performance of eventually found policies on?

Table 6.3: Summary: Key Performance Indicators

This table contains a list of all used KPIs and whether the KPIs are in the objective space Ω which contains all objectives that are part of the eight problem formulations.

KPI	KPI $\in \Omega$
welfare	yes
welfare loss	yes
temperature overshoots	yes
distance to consumption threshold	yes
distance to damage threshold	yes
population below consumption threshold	yes
population above damage threshold	yes
consumption Gini	yes
damage Gini	yes
lowest income per capita	yes
highest damage per capita	yes
GWP	no
damages	no
atmospheric temperature increase	no
industrial emissions	no
consumption per capita	no
damage per capita	no

Having identified the relevant ethical premises, designed both – aggregated and disaggregated versions of them, and chosen our KPIs to investigate their pathways, we resume to run the optimizations, perform various experiments with the found Pareto-optimal policies, and analyze the resulting pathways, trade-offs, and robustness comparisons such that we can answer the last sub-question:

Sub-Question 4

How well do the found Pareto-optimal policies perform under uncertainty?

On a high-level, we can summarize our main findings as

1. dominance of aggregation levels over ethical premises
2. correlation between low welfare and high welfare loss
3. general dominance of egalitarian aggregated Pareto-optimal policies
4. shared misery via egalitarian disaggregated Pareto-optimal policies

And with the obtained answers to the sub-questions, we can address the main research question:

Main Research Question

How are Pareto-optimal climate abatement pathways affected by the disaggregation of utility and disutility in alternative ethical problem formulations when using an integrated assessment model under deep uncertainty?

The effect of disaggregating utility and disutility is stronger than originally expected. Using disaggregated problem formulations yields substantially different pathways even within the same ethical premise – at least for most KPIs. When zooming out, the optimal policies originating from aggregated problem formulations yield pathways that look very similar. The same holds for the outcomes given optimal policies of disaggregated problem formulations.

6.2. Implications

As we have found answers to our research questions and are approaching the closing of this thesis, we eventually face the inevitable question: What does all of this entail? We distinguish two kinds of implications: scientific (Section 6.2.1) and policy implications (Section 6.2.2).

6.2.1. Scientific Implications

Let us put this thesis first back into the context of this field's current research landscape. We have made use of exploratory modeling and decision-making under deep uncertainty (DMDU) as the decision-makers and stakeholders do not know or are unable to agree on the the inner mechanics of the climate-economy system, what future scenarios are most likely, or even what objectives are most relevant. Given this nature of general deep uncertainty, we consider it prudent to not simply reduce the uncertainty to mere point estimates or consider only a handful of scenarios. On the contrary, we urge to embrace the uncertainty and consider wide ranges of scenarios, identifying optimal solutions for not only one but many objectives, and exploring the model behavior instead of providing dubious point predictions. With so much inherent uncertainty in IAMs, the DMDU approach strikes us an almost necessary tool to navigate the precariousness of climate change policy. Furthermore, by using the DMDU methodology and our focus on disaggregating incommensurate objectives such as utility and disutility, we have attempted to avoid or at least ameliorate the implications that originate from Arrow's paradox.

To recap one crucial implication: Any aggregation of objectives will result in a situation where a subset of objectives will dominate the aggregate preference ordering. In other words, the aggregation over objectives will become dictatorial as a subset of objectives will unintentionally dictate the optimized solutions in an unpredictable way. Using many-objective planning with multi-objective evolutionary algorithms (MOEAs) and disaggregated objectives can likely support climate change decision support in a better way than the default approach does in the current scientific IAM-community. With our research, we hope to have contributed a bit more to the spread of this methodology among climate change modeling and policy analysis.

Beyond the spread of the DMDU methodology, our main objective was to improve IAMs, such that they can generate more effective, robust, and equitable climate abatement strategies. Given that we have shown successfully how climate abatement pathways differ substantially when using aggregated versus disaggregated problem formulations with alternative ethical premises, we should consider shifting our attention away from other issues of the debate spotlight such as what damage function is best or what discount rate is most appropriate. The choice of the ethical premise and the choice of whether and how to disaggregate objectives could be of greater concern for model-based climate change policy than the two currently heatedly debated issues. In fact, in a previous preliminary analysis, the author has tested whether ethical premise or damage function has a bigger impact on Pareto-optimal climate abatement pathways. The results pointed clearly into the direction of ethical premise being more paramount. And as this current research has shown that the level of disaggregation has a bigger impact on the same pathways, we could reluctantly conclude by transition that the level of disaggregation is indeed more important than the choice of the damage function. But also without shifting the debate away from the damage function and the discount rate, we have shown that it is key to be explicit about normative assumptions. Maybe this research can contribute to induce a state in the modeling community in which the days of the tacit use of utilitarian normative assumptions are counted. We need to be honest, transparent, and explicit about our assumptions, let them be descriptive or normative. And of course, this does recommendation does not stop at the RICE model but can likely be applied to other more complex IAMs. We have simply used the RICE model as a computationally inexpensive case study to gain insights about whether disaggregation of utility and disutility are worthy of being explored any further. As the results are positive, it would be interesting to apply similar ethical premises and disaggregations to IAMs such as the Integrated Model to Assess the Global Environment (IMAGE) or the Model for Energy Supply Strategy Alternatives and their General Environmental Impact (MESSAGE). We would expect to see similar patterns there. An adoption of explicit normative assumptions and carefully argued and implemented (disaggregated) objectives within the problem formulations is a potential and desirable development. But the contributions are not only found in the yard of the scientific IAM-community but also in the realm of policy-making.

6.2.2. Policy Implications

International cooperation is crucial if we want to comply with the Paris Agreement and avoid catastrophic outcomes. However, we are in a situation in which the global south is more vulnerable to future climate damages and which feels unfairly treated in the deliberation process of climate action. Climate justice is necessary to enable the international cooperation that we are seeking. Logistic and technological challenges are already huge, but finding a *fair* consensus on how to proceed is also a non-trivial pursuit. Especially, the wealthy regions that are least vulnerable to climate damages and often responsible for most industrial emissions, appeal to the authority of the longstanding tool of IAMs which provide them with a flair of evidence-based decision-making when the IAMs evidently proclaim that aggregate welfare maximizing policies are the way to go. These policies are good on an aggregate level but not from the perspective of individual regions. The default way might actually exacerbate global inequality by up to 117% (Gazzotti et al., 2021). What we need to do, is looking under the figuratively normative hood of IAMs and make explicit assumptions about ethical premises and incommensurate objectives. This leads to another complication. In a morally anti-realistic world where there is no objective morality, fairness lies in the eye of the beholder. Consequently, we have to ask ourselves: Who decides what normative assumptions to use in our models? The modelers, the decision-makers, or the stakeholders? As we have discussed above, we are facing deep uncertainties which also imply *normative* or *moral uncertainty* as we do not know which theory of rational choice is the correct one (MacAskill et al., 2020). And this normative uncertainty might prompt even more contestation and potential policy deadlocks. This would pose a danger of failing to comply with the Paris Agreement. But what is the alternative? Pretending that the objective of aggregate welfare is acceptable and that we go on with such utilitarian assumptions? No, we can and maybe even should use improved models, equipped with alternative ethical premises and different levels of disaggregation, to better identify sources of injustice and gain a better understanding of policy implications using different problem formulations. What we likely need, is a stronger dialogue between the modelers and policy analysts on the one side and the stakeholders and decision-makers on the other side. The latter ones cannot just blindly trust in the *magical* outputs of the model but need to be involved to decide what problem formulations we need to use as there is no correct way to frame a complex real-world problem. As unmitigated climate damages exhibit an independent impact on a region's well-being, we could render IAMs more useful for climate policy if we a) acknowledge that the classical notion of welfare is obsolete, b) use a multi-objective approach, and c) let the decision-makers decide how they want to trade-off the various objectives in post.

References

Adler, M., Anthoff, D., Bosetti, V., Garner, G., Keller, K., & Treich, N. (2017). Priority for the worse-off and the social cost of carbon. *Nature Climate Change*, *7*(6), 443–449. https://doi.org/10.1038/nclimate3298

Adler, M. D. (2019). *Measuring social welfare: An introduction*. Oxford University Press, USA. https://doi.org/10.1093/oso/9780190643027.001.0001

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/S2010007820400023

Althor, G., Watson, J. E., & Fuller, R. A. (2016). Global mismatch between greenhouse gas emissions and the burden of climate change. *Scientific Reports*, *6*(1), 1–6. https://doi.org/10.1038/srep20281

Anthoff, D., & Emmerling, J. (2019). Inequality and the social cost of carbon. *Journal of the Association of Environmental and Resource Economists*, *6*(2), 243–273. https://doi.org/10.1086/701900

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of political economy*, *58*(4), 328–346. https://doi.org/10.1086/256963

Atteridge, A., & Remling, E. (2018). Is adaptation reducing vulnerability or redistributing it? *Wiley Interdisciplinary Reviews: Climate Change*, *9*(1), e500. https://doi.org/10.1002/wcc.500

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations Research*, *41*(3), 435–449. https://doi.org/10.1287/opre.41.3.435

Barrett, S. (2013). Local level climate justice? adaptation finance and vulnerability reduction. *Global Environmental Change*, *23*(6), 1819–1829. https://doi.org/10.1016/j.gloenvcha.2013.07.015

Batista, G. E., Keogh, E. J., Tataw, O. M., & de Souza, V. (2014). Cid: An efficient complexity-invariant distance for time series. *Data Mining and Knowledge Discovery*, *28*(3), 634–669. https://doi.org/10.1007/s10618-013-0312-3

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

Bordoff, J. (2021). The developing world needs energy — and lots of it. https://foreignpolicy.com/2021/10/29/cop26-climate-summit-developing-countries-energy-glasgow/

Bostrom, N., & Cirkovic, M. M. (2011). *Global catastrophic risks*. Oxford University Press. https://doi.org/10.1093/oso/9780198570509.001.0001

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge. https://doi.org/10.1201/9781315139470

Budolfson, M. B., Anthoff, D., Dennig, F., Errickson, F., Kuruc, K., Spears, D., & Dubash, N. K. (2021). Utilitarian benchmarks for emissions and pledges promote equity, climate and development. *Nature Climate Change*, *11*(10), 827–833. https://doi.org/10.1038/s41558-021-01130-6

Calvin, K., Patel, P., Clarke, L., Asrar, G., Bond-Lamberty, B., Cui, R. Y., Di Vittorio, A., Dorheim, K., Edmonds, J., Hartin, C., et al. (2019). GCAM v5. 1: Representing the linkages between energy, water, land, climate, and economic systems. *Geoscientific Model Development*, *12*(2), 677–698. https://doi.org/10.5194/gmd-12-677-2019,2019

Cantore, N., & Padilla, E. (2010). Equality and CO2 emissions distribution in climate change integrated assessment modelling. *Energy*, *35*(1), 298–313. https://doi.org/10.1016/j.energy.2009.09.022

Carlsen, H., Lempert, R., Wikman-Svahn, P., & Schweizer, V. (2016). Choosing small sets of policy-relevant scenarios by combining vulnerability and diversity approaches. *Environmental Modelling & Software*, *84*, 155–164. https://doi.org/10.1016/j.envsoft.2016.06.011

Chancel, L. (2020). *Unsustainable inequalities*. Harvard University Press. https://doi.org/10.4159/9780674250673

Chen, C., Noble, I., Hellmann, J., Coffee, J., Murillo, M., & Chawla, N. (2015). University of Notre Dame global adaptation index country index technical report. *ND-GAIN: South Bend, IN, USA*. https://gain.nd.edu/our-work/country-index/

Ciullo, A., de Bruijn, K. M., Kwakkel, J. H., & Klijn, F. (2019). Accounting for the uncertain effects of hydraulic interactions in optimising embankments heights: Proof of principle for the IJssel river. *Journal of Flood Risk Management*, *12*(S2), e12532. https://doi.org/10.1111/jfr3.12532

Ciullo, A., Kwakkel, J. H., De Bruijn, K. M., Doorn, N., & Klijn, F. (2020). Efficient or fair? Operationalizing ethical principles in flood risk management: A case study on the Dutch-German Rhine. *Risk Analysis*, *40*(9), 1844–1862. https://doi.org/10.1111/risa.13527

Coello, C. A. C., Lamont, G. B., Van Veldhuizen, D. A., et al. (2007). *Evolutionary algorithms for solving multi-objective problems* (Vol. 5). Springer. https://doi.org/10.1007/978-0-387-36797-2

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

Dietz, S., & Asheim, G. B. (2012). Climate policy under sustainable discounted utilitarianism. *Journal of Environmental Economics and Management*, *63*(3), 321–335. https://doi.org/10.1016/j.jeem.2012.01.003

Drouet, L., Bosetti, V., Padoan, S. A., Aleluia Reis, L., Bertram, C., Dalla Longa, F., Després, J., Emmerling, J., Fosse, F., Fragkiadakis, K., et al. (2021). Net zero-emission pathways reduce the physical and economic risks of climate change. *Nature Climate Change*, *11*(12), 1070–1076. https://doi.org/10.1038/s41558-021-01218-z

Eker, S., & Kwakkel, J. H. (2018). Including robustness considerations in the search phase of many-objective robust decision making. *Environmental Modelling & Software*, *105*, 201–216. https://doi.org/10.1016/j.envsoft.2018.03.029

Errickson, F. C., Keller, K., Collins, W. D., Srikrishnan, V., & Anthoff, D. (2021). Equity is more important for the social cost of methane than climate uncertainty. *Nature*, *592*(7855), 564–570. https://doi.org/10.1038/s41586-021-03386-6

European Commission. (2022). *Climate change consequences*. https://ec.europa.eu/clima/climate-change/climate-change-consequences_en

Farand, C. (2021). Emerging economies slam cop26 net zero push as 'anti-equity'. https://www.climatechangenews.com/2021/10/20/emerging-economies-slam-cop26-net-zero-push-anti-equity/

Franssen, M. (2005). Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. *Research in Engineering Design*, *16*(1), 42–56. https://doi.org/10.1007/s00163-004-0057-5

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, *9*(2), 123–143. https://doi.org/10.1023/A:1008894516817

Füssel, H.-M. (2010). How inequitable is the global distribution of responsibility, capability, and vulnerability to climate change: A comprehensive indicator-based assessment. *Global Environmental Change*, *20*(4), 597–611. https://doi.org/10.1016/j.gloenvcha.2010.07.009

Gazzotti, P., Emmerling, J., Marangoni, G., Castelletti, A., van der Wijst, K.-I., Hof, A., & Tavoni, M. (2021). Persistent inequality in economically optimal climate policies. *Nature Communications*, *12*(1), 1–10. https://doi.org/10.1038/s41467-021-23613-y

Gini, C. (1997). Concentration and dependency ratios. *Rivista di Politica Economica*, *87*, 769–792.

Gore, T. (2015). Extreme carbon inequality: Why the Paris climate deal must put the poorest, lowest emitting and most vulnerable people first. https://www-cdn.oxfam.org/s3fs-public/file_attachments/mb-extreme-carbon-inequality-021215-en.pdf

Green, D. (2016). The spatial distribution of extreme climate events, another climate inequity for the world's most vulnerable people. *Environmental Research Letters*, *11*(9), 091002. https://doi.org/10.1088/1748-9326/11/9/091002

Hadka, D., & Reed, P. (2013). Borg: An auto-adaptive many-objective evolutionary computing framework. *Evolutionary Computation*, *21*(2), 231–259. https://doi.org/10.1162/EVCO_a_00075

Huppmann, D., Gidden, M., Fricko, O., Kolp, P., Orthofer, C., Pimmer, M., Kushin, N., Vinca, A., Mastrucci, A., Riahi, K., et al. (2019). The MESSAGEix integrated assessment model and the ix modeling platform (ixmp): An open framework for integrated and cross-cutting analysis of energy, climate, the environment, and sustainable development. *Environmental Modelling & Software*, *112*, 143–156. https://doi.org/10.1016/j.envsoft.2018.11.012

Hwang, I. C., Reynès, F., & Tol, R. S. (2013). Climate policy under fat-tailed risk: An application of DICE. *Environmental and Resource Economics*, *56*(3), 415–436. https://doi.org/10.1007/s10640-013-9654-y

Ikefuji, M., Laeven, R. J., Magnus, J. R., & Muris, C. (2020). Expected utility and catastrophic risk in a stochastic economy-climate model. *Journal of Econometrics*, *214*(1), 110–129. https://doi.org/10.1016/j.jeconom.2019.05.007

INET. (2012). *Ben Bernanke to economists: More philosophy, please*. https://www.ineteconomics.org/about/news/2012/ben-bernanke-to-economists-more-philosophy-please

IPCC. (2022). *Climate change 2022. impacts, adaptation and vulnerability. summary for policymakers*. https://www.ipcc.ch/report/ar6/wg2/downloads/report/IPCC_AR6_WGII_SummaryForPolicymakers.pdf

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. *Wiley Interdisciplinary Reviews: Climate Change*, *12*(4), e721. https://doi.org/10.1002/wcc.721

Jaxa-Rozen, M., & Kwakkel, J. (2018). Tree-based ensemble methods for sensitivity analysis of environmental models: A performance comparison with Sobol and Morris techniques. *Environmental Modelling & Software*, *107*, 245–266. https://doi.org/10.1016/j.envsoft.2018.06.011

Kasprzyk, J. R., Nataraj, S., Reed, P. M., & Lempert, R. J. (2013). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, *42*, 55–71. https://doi.org/10.1016/j.envsoft.2012.12.007

Kasprzyk, J. R., Reed, P. M., & Hadka, D. M. (2016). Battling Arrow's paradox to discover robust water management alternatives. *Journal of Water Resources Planning and Management*, *142*(2), 04015053. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000572

Knutti, R., Rugenstein, M. A., & Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, *10*(10), 727–736. https://doi.org/10.1038/ngeo3017

Kollat, J. B., & Reed, P. M. (2005). The value of online adaptive search: A performance comparison of NSGAII, ε-NSGAII and ε-MOEA. *International Conference on Evolutionary Multi-Criterion Optimization*, 386–398. https://doi.org/10.1007/978-3-540-31880-4_27

Koopmans, T. C. (1966). On the concept of optimal economic growth. *Econometric approach to development planning, 1st edn. North Holland, Amsterdam*, 225–287. https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=1391&context=cowles-discussion-paper-series

Kowarsch, M. (2016). A pragmatist orientation for the social sciences in climate policy. *Boston Studies in the Philosophy and History of Science*. https://doi.org/10.1007/978-3-319-43281-6

Kulkarni, S. (2020). *Disaggregating the costs and benefits of climate action: An assessment using DICE* (Master's thesis). Delft University of Technology. http://resolver.tudelft.nl/uuid:8d81c2db-766d-4460-aaf4-2a91961f9f27

Kwakkel, J. H. (2017). The exploratory modeling workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling & Software*, *96*, 239–250. https://doi.org/10.1016/j.envsoft.2017.06.054

Lamont, J., & Favor, C. (2017). Distributive justice. https://plato.stanford.edu/entries/justice-distributive/

Lamontagne, J. R., Reed, P., Marangoni, G., Keller, K., & Garner, G. G. (2019). Robust abatement pathways to tolerable climate futures require immediate global action. *Nature Climate Change*, *9*(4), 290–294. https://doi.org/10.1038/s41558-019-0426-8

Lamontagne, J. R., Reed, P. M., Link, R., Calvin, K. V., Clarke, L. E., & Edmonds, J. A. (2018). Large ensemble analytic framework for consequence-driven discovery of climate change scenarios. *Earth's Future*, *6*(3), 488–504. https://doi.org/10.1002/2017EF000701

Lempert, R. J. (2003). Shaping the next one hundred years: New methods for quantitative, long-term policy analysis. https://doi.org/10.7249/MR1626

Levin, K., Cashore, B., Bernstein, S., & Auld, G. (2012). Overcoming the tragedy of super wicked problems: Constraining our future selves to ameliorate global climate change. *Policy Sciences*, *45*(2), 123–152. https://doi.org/10.1007/s11077-012-9151-0

Liao, T. W. (2005). Clustering of time series data—a survey. *Pattern Recognition*, *38*(11), 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025

Lingeswaran, S. (2019). *Redefining integrated assessment models: An exploratory approach towards robust climate-economic policies* (Master's thesis). Delft University of Technology. https://repository.tudelft.nl/islandora/object/uuid%3A812be430-79fb-4900-a0e2-b0bc9c3d804b

MacAskill, W., Bykvist, K., & Ord, T. (2020). *Moral uncertainty*. Oxford University Press. https://www.moraluncertainty.com/s/Moral-Uncertainty.pdf

Marangoni, G., Lamontagne, J. R., Quinn, J. D., Reed, P. M., & Keller, K. (2021). Adaptive mitigation strategies hedge against extreme climate futures. *Climatic Change*, *166*(3), 1–17. https://doi.org/10.1007/s10584-021-03132-x

Marchau, V. A., Walker, W. E., Bloemen, P. J., & Popper, S. W. (2019). *Decision making under deep uncertainty: From theory to practice*. Springer Nature. https://doi.org/10.1007/978-3-030-05252-2

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., & Zhou, B. (2021). Summary for policymakers. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, *1*, 1–9. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM_final.pdf

Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., et al. (2018). Global warming of 1.5 c. *An IPCC Special Report on the Impacts of Global Warming*, *1*, 1–9. https://doi.org/10.1002/wcc.500

McPhail, C., Maier, H., Kwakkel, J., Giuliani, M., Castelletti, A., & Westra, S. (2018). Robustness metrics: How are they calculated, when should they be used and why do they give different results? *Earth's Future*, *6*(2), 169–191. https://doi.org/10.1002/2017EF000649

Mercado, A. (2021). The people hit worst by climate change get the least airtime at COP26. https://www.popsci.com/environment/climate-change-equity-cop26/

Mi, Z., Liao, H., Coffman, D., & Wei, Y.-M. (2019). Assessment of equity principles for international climate policy based on an integrated assessment model. *Natural Hazards*, *95*(1), 309–323. https://doi.org/10.1007/s11069-018-3408-7

Nordhaus, W. D., & Yang, Z. (1996). A regional dynamic general-equilibrium model of alternative climate-change strategies. *American Economic Review*, *86*(4), 741–765. http://www.jstor.org/stable/2118303

Nordhaus, W. D. (1992). *The 'DICE' model: Background and structure of a dynamic integrated climate-economy model of the economics of global warming* (Cowles Foundation Discussion Papers No. 1009). Cowles Foundation for Research in Economics, Yale University. https://ideas.repec.org/p/cwl/cwldpp/1009.html

Nordhaus, W. D. (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

Okereke, C., & Coventry, P. (2016). Climate justice and the international regime: Before, during, and after Paris. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(6), 834–851. https://doi.org/10.1002/wcc.419

Osyczka, A. (1985). Multicriteria optimization for engineering design. *Design Optimization* (pp. 193–227). Elsevier. https://doi.org/10.1016/B978-0-12-280910-1.50012-X

Pindyck, R. S. (2020). The use and misuse of models for climate policy. *Review of Environmental Economics and Policy*. https://doi.org/10.1093/reep/rew012

Reed, P., & Devireddy, V. (2004). Groundwater monitoring design: A case study combining epsilon dominance archiving and automatic parameterization for the NSGA-II. *Applications of Multi-Objective Evolutionary Algorithms* (pp. 79–100). World Scientific. https://doi.org/10.1142/9789812567796_0004

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., & Kollat, J. B. (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in Water Resources*, *51*, 438–456. https://doi.org/10.1016/j.advwatres.2012.01.005

Ritchie, H., Roser, M., & Rosado, P. (2020). CO2 and greenhouse gas emissions. *Our World in Data*. https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions

Robiou du Pont, Y., Jeffery, M. L., Gütschow, J., Rogelj, J., Christoff, P., & Meinshausen, M. (2017). Equitable mitigation to achieve the Paris agreement goals. *Nature Climate Change*, *7*(1), 38–43. https://doi.org/10.1038/nclimate3186

Roser, M., & Ortiz-Ospina, E. (2013). Global extreme poverty. *Our World in Data*. https://ourworldindata.org/extreme-poverty

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sen, A. (2017). Collective choice and social welfare. *Collective choice and social welfare*. Harvard University Press. https://doi.org/10.4159/9780674974616

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

Stehfest, E., van Vuuren, D., Bouwman, L., Kram, T., et al. (2014). *Integrated assessment of global environmental change with IMAGE 3.0: Model description and policy applications*. Netherlands Environmental Assessment Agency (PBL). https://dspace.library.uu.nl/bitstream/handle/1874/308545/PBL_2014_Integrated_Assessment_of_Global_Environmental_Change_with_IMAGE_30_735.pdf

Steinmann, P., Auping, W. L., & Kwakkel, J. H. (2020). Behavior-based scenario discovery using time series clustering. *Technological Forecasting and Social Change*, *156*, 120052. https://doi.org/10.1016/j.techfore.2020.120052

Stern, N. (2022). A time for action on climate change and a time for change in economics. *The Economic Journal*, *132*(644), 1259–1289. https://doi.org/10.1093/ej/ueac005

Thomas, K., Hardy, R. D., Lazrus, H., Mendez, M., Orlove, B., Rivera-Collazo, I., Roberts, J. T., Rockman, M., Warner, B. P., & Winthrop, R. (2019). Explaining differential vulnerability to climate change: A social science review. *Wiley Interdisciplinary Reviews: Climate Change*, *10*(2), e565. https://doi.org/10.1002/wcc.565

Tjallingii, I. (2021). *Accounting for distributive justice in integrated assessment models: Towards a more equitable climate policy agenda* (Master's thesis). Delft University of Technology. http://resolver.tudelft.nl/uuid:4d3a43b6-25d0-452d-9906-5da597fb0ea1

Tol, R. S. (1997). On the optimal control of carbon dioxide emissions: An application of FUND. *Environmental Modeling & Assessment*, *2*(3), 151–163. https://doi.org/10.1023/A:1019017529030

Wagner, G., & Zeckhauser, R. J. (2016). Confronting deep and persistent climate uncertainty. https://doi.org/10.2139/ssrn.2818035

Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. *Health Economics*, *17*(12), 1329–1344. https://doi.org/10.1002/hec.1331

Watson, A. A., & Kasprzyk, J. R. (2017). Incorporating deeply uncertain factors into the many objective search process. *Environmental Modelling & Software*, *89*, 159–171. https://doi.org/10.1016/j.envsoft.2016.12.001

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

Winsberg, E. (2018). *Philosophy and climate science*. Cambridge University Press. https://doi.org/10.1017/9781108164290

WMO. (2022). *2021 one of the seven warmest years on record, WMO consolidated data shows*. https://public.wmo.int/en/media/press-release/2021-one-of-seven-warmest-years-record-wmo-consolidated-data-shows#:~:text=The%20warmest%20seven%20years%20have,to%20record%20global%20average%20warming.

Woodruff, M., & Herman, J. (2013). *pareto.py: An ϵ-nondomination sorting routine*. https://github.com/matthewjwoodruff/pareto.py

World Bank. (2018). Poverty and shared prosperity 2018: Piecing together the poverty puzzle. https://doi.org/10.1596/978-1-4648-1330-6

Yumashev, D., Hope, C., Schaefer, K., Riemann-Campe, K., Iglesias-Suarez, F., Jafarov, E., Burke, E. J., Young, P. J., Elshorbany, Y., & Whiteman, G. (2019). Climate policy implications of nonlinear decline of Arctic land permafrost and other cryosphere elements. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-09863-x

A

Appendix: Literature Review

Two main literature scanning phases have been conducted to arrive at the final collection of sources that were included in this review.

The first phase yielded papers through a systematic use of the scientific database *Scopus*. The search terms were limited to the title, the abstract, and keywords. The search terms were a conjunction of three components. The first component of the conjunction is a disjunction of different IAM related terms, such as ”integrated assessment model*”, ”IAM*”, ”RICE*”, ”DICE*”, ”Dynamic Integrated model of Climate and the Economy*”, and ”climate-economy model*”. The second component is a disjunction of terms that are related to distributive justice, such as ”*equality*”, ”equity*”, ”distributive justice*”, ”fair*”, ”ethical principle*”, ”distribution of risk*”, ”utilitarian*”, ”probabilistic abatement pathways*”, and ”expected utility”. The final conjunctive component was the term ”climate*” in order to reduce the results climate-related papers. Further search restrictions that affected the source type (journals), document type (article and letter), language (English) and recent publication years (2018-2021) have been applied as well. The focus on recent years directs attention to the current state of art in this field and provides a reasonable starting point to identify a research gap. The aforementioned constraints have been used in a conjunctive form.

This search yielded 59 papers[1]. After carefully reading through the titles and abstracts, further papers have been filtered out and resulting at eight papers that have been included in Table 1.1. Furthermore, in consultation with senior researchers, two very recent and relevant Master's theses from the TPM faculty at the TU Delft have been brought to the attention of the author, namely Kulkarni (2020) and Tjallingii (2021). The second phase consisted in a backward reference search from the remaining eight papers and the two Master's theses.

---

[1] The search results can be viewed with this link. The number of current search results might slightly differ as time more articles might have been added to the Scopus database.

B

Appendix: Model

B.1. Regions

Table B.1: RICE Regions

RICE Region	Description
US	The United States of America
OECD-Europe	European countries that are members of the OECD
Japan	Japan
Russia	Russia
Non-Russia Eurasia	Countries that are in Eurasia excluding Russia
China	China
India	India
Middle East	Middle-Eastern countries
Africa	Countries of the African continent
Latin America	Countries of Latin America
OHI	Other higher-income countries
Other non-OECD-Asia	Countries in Asia that are not members of the OECD

B.2. Relations

B.2.1. Economic Sub-Model

B.2.1.1. Production Function

$$Y_{gross,t,r} = A_{t,r} \cdot L_{t,r}^{1-\gamma} \cdot K_{t,r}^{\gamma} \tag{B.1}$$

where

$Y_{gross,t,r}$	gross output per region per period
$A_{t,r}$	total factor productivity per region per period
$L_{t,r}$	labour force per region per period
$K_{t,r}$	capital stock per region per period
γ	output elasticity of capital

B.2.1.2. Capital Function

$$K_{t+1,r} = I_{t,r} + (1-\delta) \cdot K_{t,r} \tag{B.2}$$

where

$K_{t,r}$	capital stock per region per period
$I_{t,r}$	investment per region per period
δ	depreciation rate of capital

B.2.1.3. Modified Production Function

$$Y_{net,t,r} = Y_{gross,t,r} \cdot (1-\Lambda_{t,r}) \cdot (1-\Omega_{t,r}) \tag{B.3}$$

where

$Y_{net,t,r}$	modified gross output per region per period
$\Lambda_{t,r}$	abatement costs per region per period
$\Omega_{t,r}$	damage costs per region per period

B.2.1.4. Abatement Function

$$\Lambda_{t,r} = \Theta_1 \cdot \mu_{t,r}^{\Theta_2} \tag{B.4}$$

where

$\Lambda_{t,r}$	abatement costs per region per period
Θ_1	calibrated parameter 1
Θ_2	calibrated parameter 2
$\mu_{t,r}$	emissions control rate per region per period

B.2.1.5. Damage Function

$$\Omega_{t,r} = \phi_{1,r} \cdot T_{ATM,t} + \phi_{2,r} \cdot T_{ATM,t}^{\phi_{3,r}} \tag{B.5}$$

where

$\Omega_{t,r}$	damage costs per region per period
$T_{ATM,t}$	atmospheric temperature per period

$\phi_{1,r}$	calibrated parameter 1 per region
$\phi_{2,r}$	calibrated parameter 2 per region
$\phi_{3,r}$	calibrated parameter 3 per region

B.2.1.6. Consumption and Investment Functions

$$C_{t,r} = (1 - \sigma) \cdot Y_{net,t,r} \tag{B.6}$$

$$I_{t,r} = \sigma \cdot Y_{net,t,r} \tag{B.7}$$

where

$C_{t,r}$	Consumption per region per period
$I_{t,r}$	Investment per region per period
$Y_{net,t,r}$	modified gross output per region per period
σ	savings rate

B.2.2. Carbon-Cycle Sub-Model

B.2.2.1. Global Emissions Function

$$E_t = \sum_{r=1}^{R} (E_{ind,t,r}) + E_{land,t} \tag{B.8}$$

where

E_t	global emissions
R	number of regions
$E_{ind,t,r}$	industrial emissions per region per period
$E_{land,t}$	textland usage emissions per period

B.2.2.2. Industrial Emissions Function

$$E_{ind,t,r} = \sigma_{t,r} \cdot Y_{gross,t,r} \cdot (1 - \mu_{t,r}) \tag{B.9}$$

where

$E_{ind,t,r}$	industrial emissions per region per period
$\sigma_{t,r}$	emission output ratio per region per period
$Y_{gross,t,r}$	gross output per region per period
$\mu_{t,r}$	emission control rate per region per period

B.2.2.3. Carbon Concentration Function

$$\begin{pmatrix} MAT_{t+1} \\ MU_{t+1} \\ ML_{t+1} \end{pmatrix} = \begin{pmatrix} 1.36 \\ 0 \\ 0 \end{pmatrix} \cdot E_t + \begin{pmatrix} \phi_{11} & \phi_{11} & 0 \\ \phi_{21} & \phi_{22} & \phi_{32} \\ 0 & \phi_{23} & \phi_{33} \end{pmatrix} \cdot \begin{pmatrix} MAT_t \\ MU_t \\ ML_t \end{pmatrix} \tag{B.10}$$

where

MAT_t	CO2 concentration in atmosphere per period
MU_t	CO2 absorption and release in upper oceans per period
ML_t	CO2 absorption and release in lower oceans per period

E_t	global emissions per period
ϕ_{xy}	calibrated parameters

B.2.2.4. Radiative Forcing Function

$$FORC_t = \eta \cdot \left[log_2 \left(\frac{MAT_t}{MAT_{eq}} \right) \right] + FORC_{ex,t} \tag{B.11}$$

where

$FORC_t$	radiative forcing per period
$FORC_{ex,t}$	external radiative forcing through other GHG per period
MAT_t	CO2 concentration in atmosphere per period
MAT_{eq}	pre-industrial CO2 equilibrium concentration

B.2.3. Climate Sub-Model

B.2.3.1. Atmospheric Temperature Function

$$\begin{pmatrix} T_{ATM,t+1} \\ T_{OCEAN,t+1} \end{pmatrix} = \begin{pmatrix} \chi_1 \\ 0 \end{pmatrix} \cdot FORC_t + \begin{pmatrix} 1 - \chi_1\chi_2 - \chi_1\chi_3 & \chi_1\chi_3 \\ \chi_4 & 1 - \chi_4 \end{pmatrix} \cdot \begin{pmatrix} T_{ATM,t} \\ T_{OCEAN,t} \end{pmatrix} \tag{B.12}$$

where

$T_{ATM,t}$	atmospheric temperature per period
$O_{OCEAN,t}$	temperature in oceans per period
$FORC_t$	radiative forcing per period
χ_1	climate equation coefficient for upper stratum
χ_2	climate sensitivity parameter
χ_3	heat transfer coefficient between upper and lower stratum
χ_4	climate equation coefficient for lower stratum

B.3. Metrics

B.3.1. Overview

B.3.2. Temperature Overshoots

For a given year, this variable represents how many time steps have seen atmospheric temperature increases of more than 2°C.

$$O_t = \sum_{i=2005}^{t} f(T_i) \tag{B.13}$$

where

O	number of time steps with 2°C atmospheric temperature overshoots
t	time step with $t \in [2005, 2305]$ and a step size of 10
T_t	Atmospheric temperature increase in °C at period t
$f(x)$	function that returns 1 if its argument x is higher than 2°C, else return 0

B.3.3. Welfare and Welfare Loss

Welfare is a commonly defined variable in the context of models such as DICE and RICE. The equation is given below.

Table B.2: M: Metrics

This overview table depicts all metrics. The column *variable name* represents the exact name of the variable within the code.

metrics	variable name	explanation	units
welfare	`global_per_util_ww`	Welfare of discounted utility based on consumption.	Dmnl
welfare loss	`global_per_disutility_ww`	Welfare of discounted disutility based on climate damages.	Dmnl
consumption Gini	`CPC_intra_gini`	Inequality of consumption across regions adjusted by population size.	Dmnl
damage Gini	`climate_impact_per_dollar_gini`	Inequality of consumption over damages across regions adjusted by population size.	Dmnl
lowest income class per capita	`worst_off_income_class`	Worst-off income class with the lowest income quintile across all regions.	$1000 per person
highest damage per capita	`worst_off_damage`	Worst-off quintile of climate impact over consumption across all regions.	Dmnl
maximum distance to consumption threshold	`max_utility_distance_threshold`	Maximum utility distance to the consumption threshold of all income quintiles across all regions.	trillion $
population below consumption threshold	`population_below_consumption_threshold`	Size of population that is above the $3.20 per person per day threshold.	million people
maximum distance to damage threshold	`max_disutility_distance_threshold`	Maximum disutility distance to the damage threshold of all income quintiles across all regions.	trillion $
population above damage threshold	`population_above_damage_threshold`	Size of population that is above a 5% threshold for consumption over damages.	million people
# of years with 2°C temperature overshoots	`temperature_overhoots`	Number of time steps that have an atmospheric temperature increase of 2°C or more.	Dmnl

$$W_{c,t,r} = \sum_{i=1}^{t} U(C)_{i,r} \cdot R_{c,i,r} \cdot L_{i,r} \tag{B.14}$$

where

$W_{c,t,r}$	welfare of region r based on consumption at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
$U(C)_{t,r}$	utility of consumption of region r at period t
$R_{c,t}$	social discount factor for consumption at period t
$L_{t,r}$	population of region r at period t

Utility of consumption can be defined by the Constant Relative Risk Aversion (CRRA) utility function (Wakker, 2008).

$$U(C)_{t,r} = \begin{cases} \frac{C_{t,r}^{1-\eta}}{1-\eta} & \text{if } \eta \neq 0 \\ ln(C_{t,r}) & \text{if } \eta = 0 \end{cases} \tag{B.15}$$

where

$U(C)_{t,r}$	utility of consumption of region r at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
$C_{t,r}$	consumption of region r at period t
η	coefficient of relative risk aversion for consumption

The social discount factor for consumption is endogenously defined as:

$$R_{c,t} = \frac{1}{1 + \rho_{c,t}} \tag{B.16}$$

where

$R_{c,t}$	social discount factor for consumption at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
$\rho_{c,t}$	endogenous social rate of consumption

The endogenous social rate of consumption is defined as:

$$\rho_{c,t} = \delta_c + \eta \cdot g_{c,t} \tag{B.17}$$

where

$\rho_{c,t}$	endogenous social rate of consumption
δ_c	initial rate of time preference for consumption
η	coefficient of relative risk aversion for consumption
$g_{c,t}$	rate of change of consumption at period t

And eventually, the rate of change of consumption is defined as:

$$g_{c,t} = \frac{cpc_{r,t} - cpc_{r,t-1}}{cpc_{r,t-1}} \tag{B.18}$$

where

$g_{c,t}$	rate of change of consumption at period t
$cpc_{r,t}$	consumption per capita of region r at period t

Welfare loss is a disaggregated variable that complements welfare. While `welfare` is based on consumption, `welfare loss` is based on suffered climate damages. The structure of `welfare loss` is very similar to the structure of `welfare`. We base this on the work of Kulkarni (2020). Instead of focusing on pure damages, it consider it relevant to also apply discounting and other adjustments in the same way as for consumption. The result can be seen below.

$$W_{d,t,r} = \sum_{i=1}^{t} V(\Omega)_{i,r} \cdot R_{d,i,r} \cdot L_{i,r} \tag{B.19}$$

where

$W_{d,t,r}$	welfare loss of region r based on damages at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
$V(\Omega)_{t,r}$	disutility of damages of region r at period t
$R_{d,t}$	social discount factor for damages at period t
$L_{t,r}$	population of region r at period t

$$V(\Omega)_{t,r} = \begin{cases} \frac{\Omega_{t,r}^{1-\tau}}{1-\tau} & \text{if } \tau \neq 0 \\ ln(\Omega_{t,r}) & \text{if } \tau = 0 \end{cases} \tag{B.20}$$

where

$V(\Omega)_{t,r}$	disutility of damages of region r at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
$\Omega_{t,r}$	damages of region r at period t
τ	coefficient of relative risk aversion for climate damage

The social discount factor for damages is also endogenously defined in a similar way:

$$R_{d,t} = \frac{1}{1 + \rho_{d,t}} \tag{B.21}$$

where

$R_{d,t}$	social discount factor for damages at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
$\rho_{d,t}$	endogenous social rate of damage

The endogenous social rate of damage is defined as:

$$\rho_{d,t} = \delta_d + \tau \cdot g_{d,t} \tag{B.22}$$

where

$\rho_{d,t}$	endogenous social rate of damage
δ_d	initial rate of time preference for damage
τ	coefficient of relative risk aversion for damages
t	time step with $t \in [2005, 2305]$ and a step size of 10
$g_{d,t}$	rate of change of damage at period t

And eventually, the rate of change of damage is defined as:

$$g_{d,t} = \frac{dpc_{r,t} - dpc_{r,t-1}}{dpc_{r,t-1}} \tag{B.23}$$

where

$g_{d,t}$	rate of change of damage at period t
$dpc_{r,t}$	damages per capita of region r at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10

B.3.4. Population and Maximum Distance to Thresholds

There are four objectives that are occur only in the sufficientarian problem formulations:

1. population below consumption threshold
2. maximum distance to consumption threshold
3. population above damage threshold
4. maximum distance to damage threshold

All of these objectives are desired to be minimized. The first two objectives are based on consumption and are part of the sufficientarian aggregated problem formulation. The third and fourth objectives are based on damages and are additionally added to the sufficientarian disaggregated problem formulation.

First, we discuss the aggregated version.

$$POP_{c,t} = \sum_{r=0}^{12} \sum_{q=0}^{5} f(p_{r,q,t}) \tag{B.24}$$

where

$POP_{c,t}$	population below consumption threshold at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
r	region r
q	quintile of a region r
f_x	function returning the population of argument x if x is above threshold, else 0
$p_{r,q,t}$	consumption of quintile q of region r at period t

The other sufficientarian aggregated objective refers to the maximum distance to the consumption threshold.

$$d_{c,t} = max(\ \forall(q \in Q)\ f(q, \theta_{c,t})\ ) \tag{B.25}$$

where

$d_{c,t}$	maximum distance to consumption threshold at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
Q	set of all quintiles of all regions
$f(x, \theta_{c,t})$	function returning a region's x distance to consumption threshold at period t
$\theta_{c,t}$	consumption threshold at period t

The other two sufficientarian disaggregated variables, are designed in a similar manner. The only difference is that we do not consider the separate quintiles but only regions.

$$POP_{d,t} = \sum_{r=0}^{12} f(p_{r,t}) \tag{B.26}$$

where

$POP_{c,t}$	population above damage threshold at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
r	region r
f_x	function returning the population of argument x if x is below threshold, else 0
$p_{r,t}$	damage of region r at period t

The other sufficientarian aggregated objective refers to the maximum distance to the consumption threshold.

$$d_{d,t} = max(\ \forall(r \in R)\ f(r, \theta_{d,t})\ ) \tag{B.27}$$

where

$d_{d,t}$	maximum distance to damage threshold at period t
t	time step with $t \in [2005, 2305]$ and a step size of 10
R	set of all regions
$f(x, \theta_{d,t})$	function returning a region's x distance to damage threshold at period t
$\theta_{d,t}$	damage threshold at period t

B.3.5. Consumption and Damage Ginis

Both egalitarian objectives are based on the Gini index developed by Corrado Gini (Gini, 1997).

$$G_{c,t} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |C_{t,i} - C_{t,j}|}{2n^2} \tag{B.28}$$

$G_{c,t}$	Gini index for consumption per period
$C_{t,r}$	consumption per region per period
t	time step with $t \in [2005, 2305]$ and a step size of 10
r	region r
n	number of regions

$$G_{d,t} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |\Omega_{t,i} - \Omega_{t,j}|}{2n^2} \tag{B.29}$$

$G_{d,t}$	Gini index for climate damages per period
$\Omega_{t,r}$	damages over consumption per region per period
t	time step with $t \in [2005, 2305]$ and a step size of 10
r	region r
n	number of regions

B.3.5.1. Worst-Off Consumption and Damage

$$C_{c,min,t} = min[C_{t,r,q} \text{ for region } r \text{ in quintile } q] \tag{B.30}$$

$C_{c,min,t}$	worst-off consumption per period
$C_{t,r,q}$	consumption per quintile per region per period
t	time step with $t \in [2005, 2305]$ and a step size of 10
r	region r
q	quintile q in a region r

$$D_{c,max,t} = max[\Omega_{t,r,q} \text{ for region } r \text{ in quintile } q] \tag{B.31}$$

$D_{c,max,t}$	worst-off damage over consumption per period
$\Omega_{t,r,q}$	damage over consumption per quintile per region per period
t	time step with $t \in [2005, 2305]$ and a step size of 10
r	region r
q	quintile q in a region r

B.4. Uncertainties

Table B.3: X: Uncertainties

This overview table depicts all uncertainties. The column *variable name* represents the exact name of the variable within the code.

uncertainty	variable name	explanation	range	units
population growth rate	`scenario_pop_gdp`	Scenario of population growth rate based on shared socioeconomic pathways.	`{short-term, long-term}`	1/year
total factor productivity growth rate	`t2xco2_dist`	Growth rate of the ratio of aggregate output to aggregate inputs.	`{short-term, long-term}`	1/year
availability of fossil fuel	`fosslim`	The availability of fossil fuel represents the limit on the maximum volume of industrial emissions.	`[4000, 13649]`	GtC
emission to output growth rate	`scenario_sigma`	Scenario that determines the CO2 efficiency development.	`{short-term, long-term}`	1/year
climate sensitivity parameter	`t2xco2_index`	This parameter represents the equilibrium temperature impact.	`{normal, log normal, Cauchy`	Dmnl
elasticity of climate impact	`scenario_elasticity_of_damages`	Scenario that determines the damage relation for lower income groups. Value of 1 entails a uniform distribution of climate impact across economic consumption. A value of -1 entails a damage distribution which predominantly affects the lower income shares.	`{-1, 0, 1}`	Dmnl
price of backstop technology	`scenario_cback`	Scenario that determines the price of backstop technology.	`[1260, 1890]`	2005$/tCO2
availability of negative emissions	`scenario_limmiu`	Scenario that determines whether negative emissions are possible.	`{yes (20%), no}`	Dmnl
elasticity of disutility of damage	`emdd`	Coefficient of relative risk aversion for climate damage. It's used as a parameter in the damage function and to compute the endogenous social rate of damage.	`[0.001, 0.6]`	Dmnl

B.5. Levers

Table B.4: L: Levers

This overview table depicts all levers. The column *variable name* represents the exact name of the variable within the code.

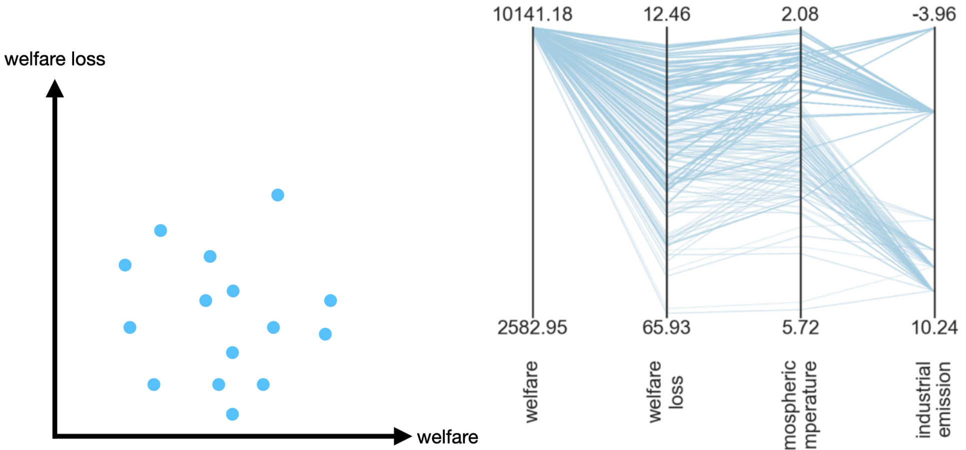
lever	variable name	explanation	range	units
savings rate	`sr`	The savings rate describes the percentage of disposable income that a region decides to save rather than spends on consumptions. The savings rate is exogenously set to converge to the steady-state savings rate for the world.	`[0.1, 0.5]`	dmnl
emission control rate target	`miu`	This variable describes the date (year) in which the world exhibits zero emission for the first time.	`[2065, 2305]`	date (year)
initial rate of social time preference for consumption	`irstp_consumption`	This variable represents how much consumption by future generations is valued. This is, however, only an initial value. The resulting social discount rate is endogenous, adjusting with e.g., population and consumption growth.	`[0.001, 0.015]`	dmnl
initial rate of social time preference for damage	`irstp_damage`	This variable represents how much damage by future generations is valued. This is, however, only an initial value. The resulting social discount rate is endogenous, adjusting with e.g., population and consumption growth.	`[0.001, 0.015]`	dmnl

C

Appendix: Results

C.1. Parallel Axes Plots: How to Read Them

This section has the intention to explain how parallel axes plots have to be read. We do this because we a) use them a lot in this project and b) we are aware that not all readers are familiar with them. So, why do we use parallel axes plots in the first place? Usually, we want to show the location of a point in some space. For example, let us consider the two variables `welfare` and `welfare loss`. A simulation run yields a result with values for these two objectives. In a two-dimensional space, this outcome would be represented by a point (see Figure C.1a below). Each point represents the results of a separate simulation run. If we increase the number of dimensions of the objective space by adding one more objective (e.g., `atmospheric temperature`), we could represent the result of a simulation run as a point in three-dimensional space. What happens if we increase the number of dimensions even further? One way of how to deal with this consists in using a parallel axes plot as shown in Figure C.1b. Let us add another objective (`industrial emission`) for illustration.



(a) Representation in 2D Space

Higher `welfare` value and lower `welfare loss` are desired.

(b) Representation in Higher-Dimensional Space

Values at the top of the axes are desired.

Figure C.1: Representation in Various Spaces

For each objective, we introduce a vertical axis. E.g., the left-most axis represents the dimension `welfare`. The axis spans a given range from top to bottom. More desirable values are on top and vice versa. E.g., higher `welfare` values are on top and reach a maximum of 10141.18 as we want to maximize this metric. As `welfare loss` should be minimized, we find the lower and therefore more desirable values on top and vice versa. But how do we represent the result of a simulation run? Such a result needs a value for each dimension. Figure C.1b shows lines that connect each axis once. One *point* in 2D or 3D space is now a line across all axes of the parallel axes plot. And the most desirable line would be a line that intersects all axes at the top.

A side note: Featuring the desirability of a variable value applies only to the objective space as depicted in Figure C.1. If we look at the individual dimensions of policies or scenarios, there is no preferred direction. Beyond this, everything works the same way.

C.2. Open Exploration

Here, we ran experiments with the Nordhaus policy and 30’000 different scenarios to conduct the time series clustering for scenario discovery. We can see the pathways for the metrics of all problem formulations in Figure C.2.

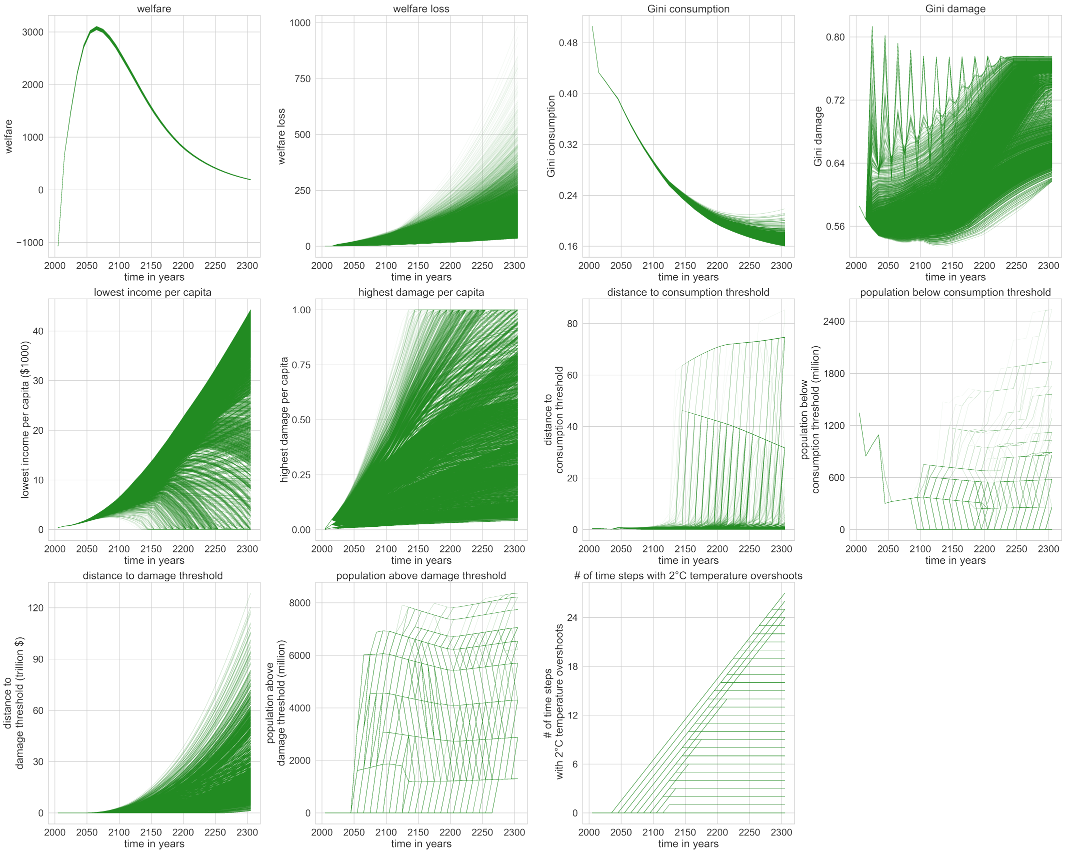


Figure C.2: Open Exploration with Nordhaus Policy and 30’000 Scenarios

C.3. Scenario Selection

C.3.1. Time Series Clustering

Building upon the open exploration data with 30’000 scenarios and the Nordhaus policy, we can use it to conduct time series clustering[1]. The process of time series clustering yields the average silhouette

---

[1] To have a look at the pathways for all metrics, see Figure C.2 in the appendix.

widths for all metrics and a number of clusters as depicted in Figure C.3.

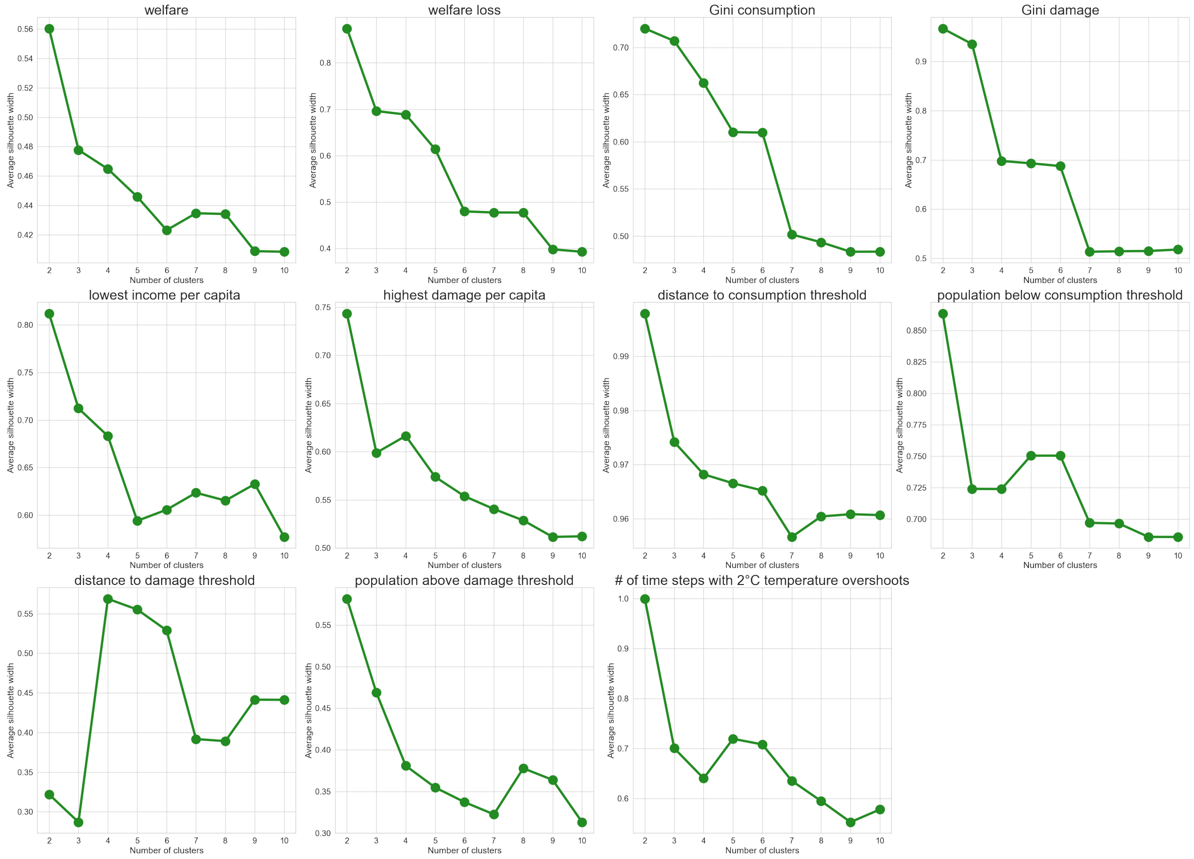


Figure C.3: Time Series Clustering: Average Silhouette Widths

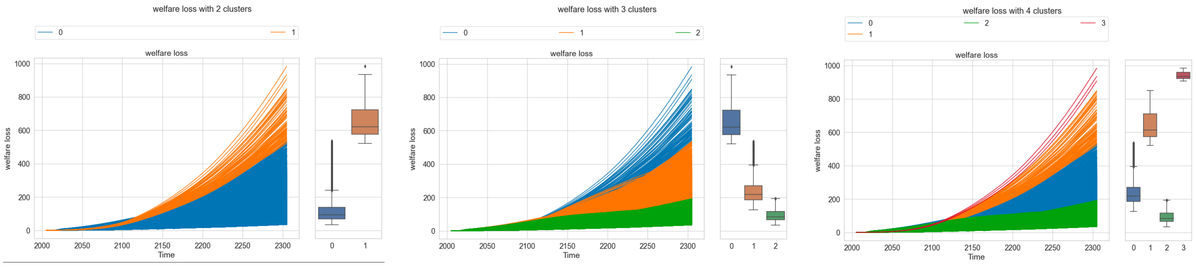
Average silhouette widths are calculated for a range of clusters $[2, 12]$. For all panels, the x-axis shows the number of used clusters while the y-axis show the average silhouette width. A lower average silhouette width is more desirable.

In order to determine how many clusters are most suitable for each metric, we need to identify the cluster number with the highest average silhouette width per metric. Unfortunately, this is usually not sufficient. An additional visual inspection of the pathways and their different clusters has to be inspected. Details can be found in the appendix but we provide here an example with the metric `welfare loss`. In Figure C.3 (first row, second column), we can see that there are three relative high average silhouette widths for clusters 2, 3, and 4. Thus, we plot the pathways for `welfare loss` with these cluster numbers in Figure C.4a, Figure C.4b, and Figure C.4c. Our observations are that the 3-cluster version captures a different separation than the 2-cluster version as not only exponential curves are visible. The 4-cluster version does not add much but splitting the highest cluster into two which generates a cluster for the highest disutility which are a very small set of scenarios as there are not many lines. As we want to cover a bigger space, we decide that the 3-cluster version (Figure C.4b) is the most appropriate one. And in Figure C.4b, we pick cluster `0` (the blue one) as the relevant worst-case scenario set.

We apply the same visual inspection to all pathways of all other metrics with their corresponding potential cluster numbers that we can read from Figure C.3[2]. Eventually, we merge all resulting scenarios into one set of worst-case scenarios. Like this, we cover all worst-case scenarios for each objective which might omit scenarios that share bad (but not worst) outcomes across various metrics. In total, we arrive at 4’350 worst-case scenarios. The resulting scenarios can be found in the appendix

---

[2] The clustered pathways for the other metrics can be found in this Jupyter notebook on GitHub.



(a) 2 Clusters **(b)** 3 Clusters **(c)** 4 Clusters

Figure C.4: Clustered Pathways for Welfare Loss

Pathways for the metric `welfare loss`. Time horizon spans from 2005 to 2305. Each line represents a model run. The same data is clustered in several ways.

under Figure C.5[3]. Given the float-variables, we can see that the scenarios span a wide space and are rather well distributed in uncertainty space. The variables `emdd` (elasticity of disutility of damage), `fosslim` (availability of fossil fuel), and `t2xco2_index` (climate sensitivity parameter) span virtually their entire ranges. Figure C.5 shows the resulting vulnerable scenarios that we identify via the time series clustering method.

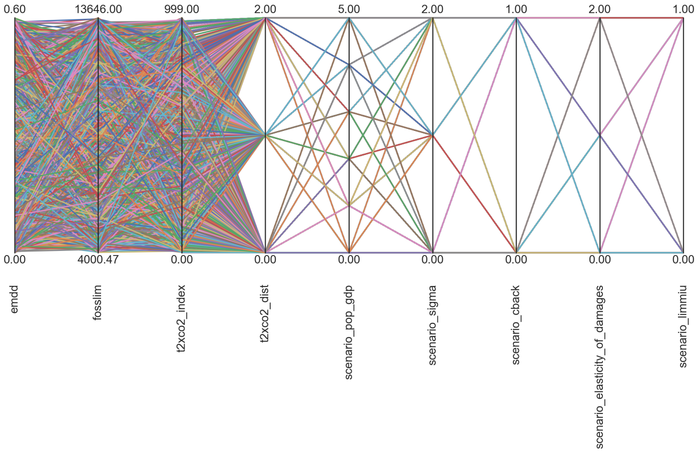


Figure C.5: 4’350 Worst-Case Scenarios from Time Series Clustering

Each axis represents an uncertainty dimension. Only the first three variables are of type float. The remaining variables are of type integer. There is no desired direction on the axes.

C.3.2. Directed Scenario Search

Directed scenario search yields 59 worst-case scenarios and can be viewed in the appendix under Figure C.6. We can see that these results are sparser than in the results of time series clustering. Interestingly, we can also see that the `emdd` values are strongly concentrated around the highest values. This makes sense as `emdd` represents the elasticity of disutility of damage. When this value is high, we have a high relative risk aversion for climate change. The metric `welfare loss` becomes correspondingly also very high.

Figure C.6 shows the resulting vulnerable scenarios that we identify via directed scenario search. Eventually, Figure C.7 shows the resulting reference scenarios.

---

[3] How to exactly read a parallel axes plot is explained in the appendix under Section C.1.

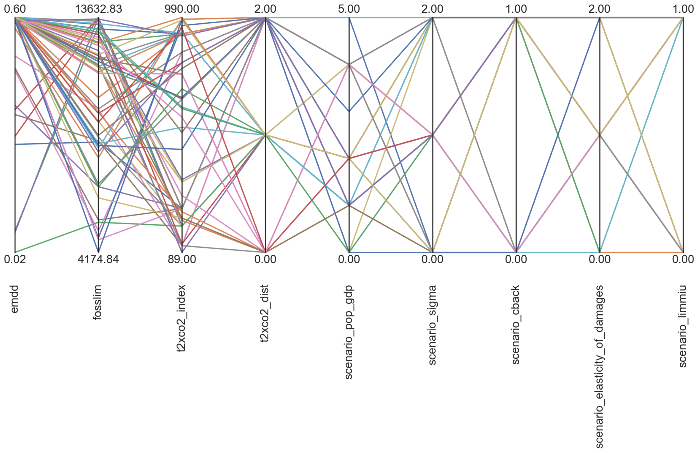


Figure C.6: 59 Worst-Case Scenarios from Directed Scenario Search

Each axis represents an uncertainty dimension. Only the first three variables are of type float. The remaining variables are of type integer. There is no desired direction on the axes.

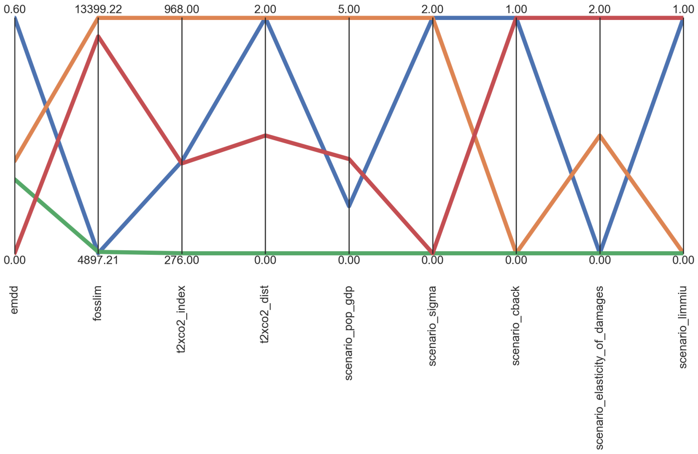


Figure C.7: 4 Reference Scenarios

Each axis represents an uncertainty dimension. Only the first three variables are of type float. The remaining variables are of type integer. There is no desired direction on the axes.

C.4. Policy Discovery

C.4.1. Sensitivity to Seeds

Figure C.8 and Figure C.9 show two example problem formulations which exhibit no discernible difference between the patterns of the KPI pathways. The order of plotting by seeds does not offer any interesting divergence between the two optimization results. We can conclude that the optimization process is not sensitive to the chosen seed. Of course, a more thorough investigation would be desirable. It could be the case, that the exact two seeds that we have used, are actually an exception. For now, we consider these results as evidence that the seed does not affect the optimization in any substantial way. The outcomes for other problem formulations show the same pattern.

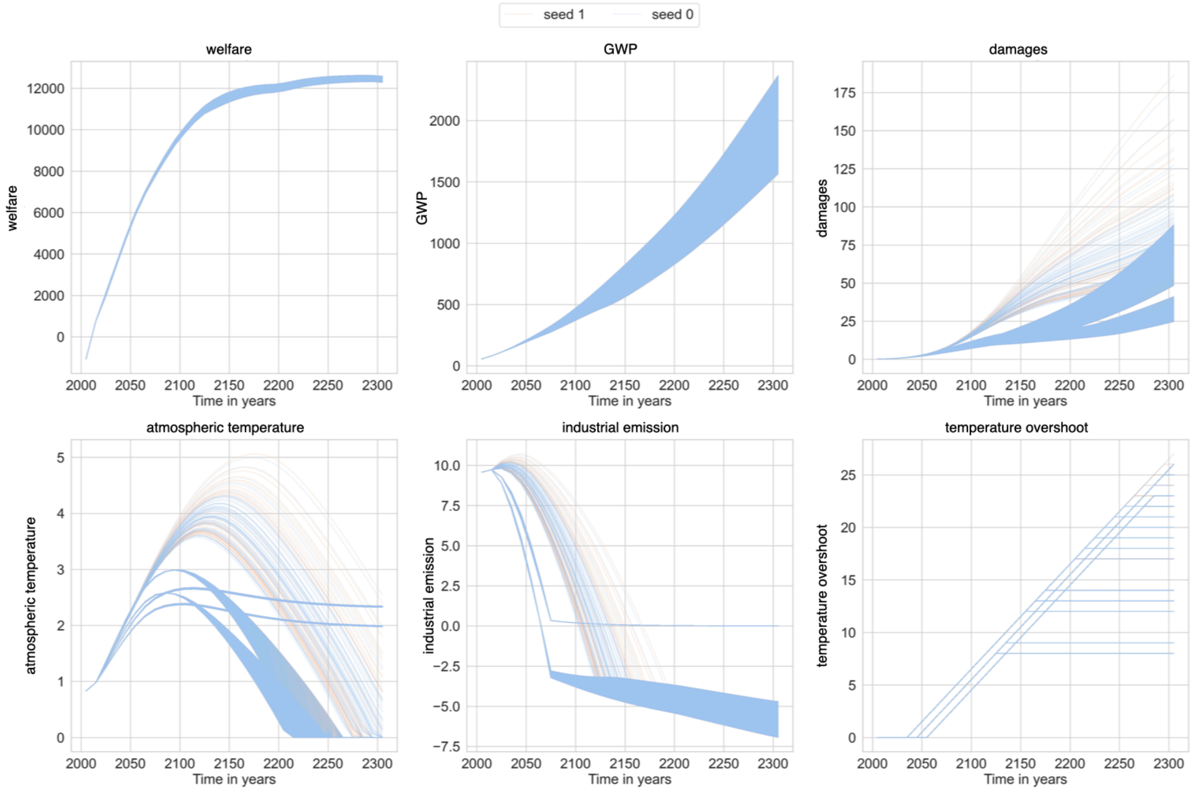


Figure C.8: Sensitivity to Seeds: Order of Plotting – Orange, Blue

Some Pathways are shown here. Time horizon spans from 2005 to 2305. Each line represents a policy run with its corresponding reference scenario. In both plots is exactly the same data. Only the order of plotting is different.

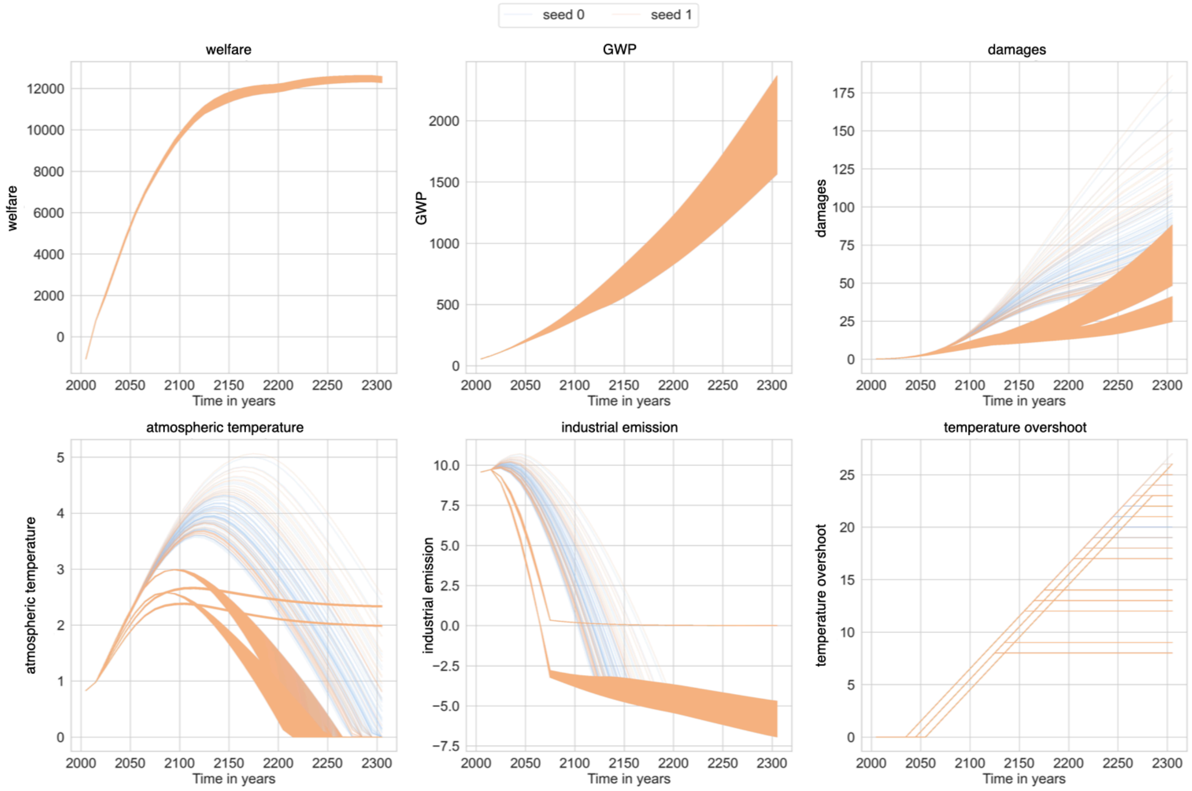
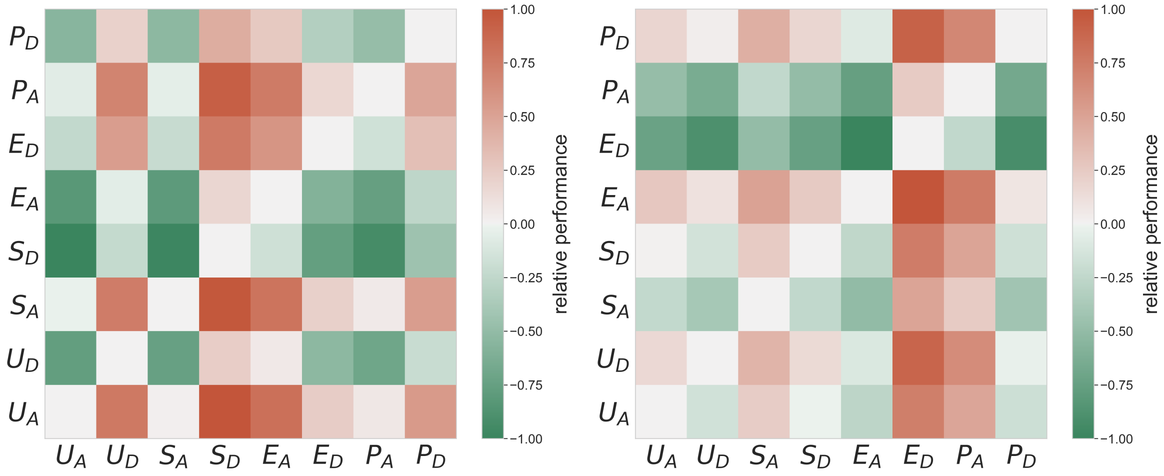


Figure C.9: Sensitivity to Seeds: Order of Plotting – Blue, Orange

Some Pathways are shown here. Time horizon spans from 2005 to 2305. Each line represents a policy run with its corresponding reference scenario. In both plots is exactly the same data. Only the order of plotting is different.

C.4.2. High-Level Comparison

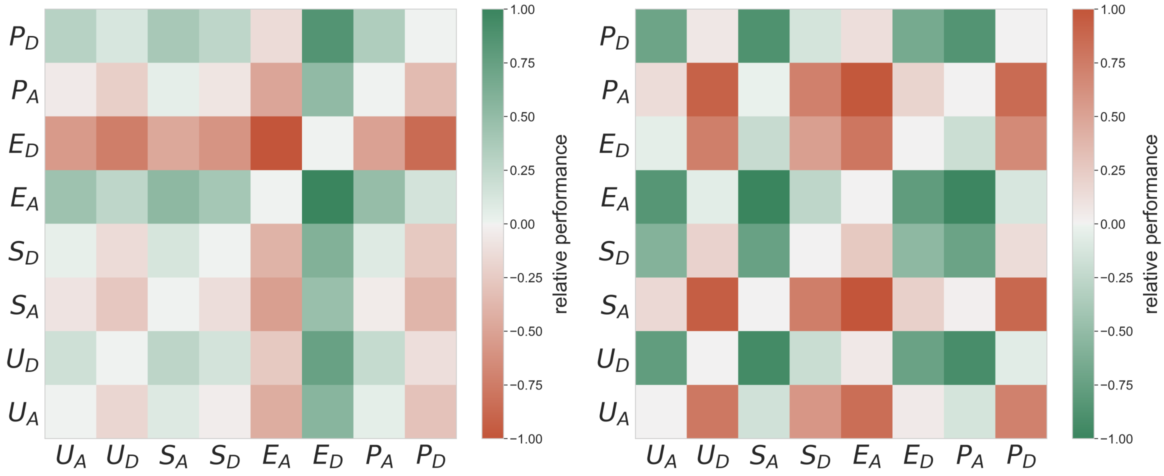


(a) Consumption Gini

(b) Damage Gini

Figure C.10: Relation between Problem Formulations for Median Egalitarian Objectives

A grid cell represents the relation between a row's and a column's problem formulation. Green represents that the row's problem formulation is more desirable than the column's one.



(a) Lowest Income per Capita

(b) Highest damage per capita

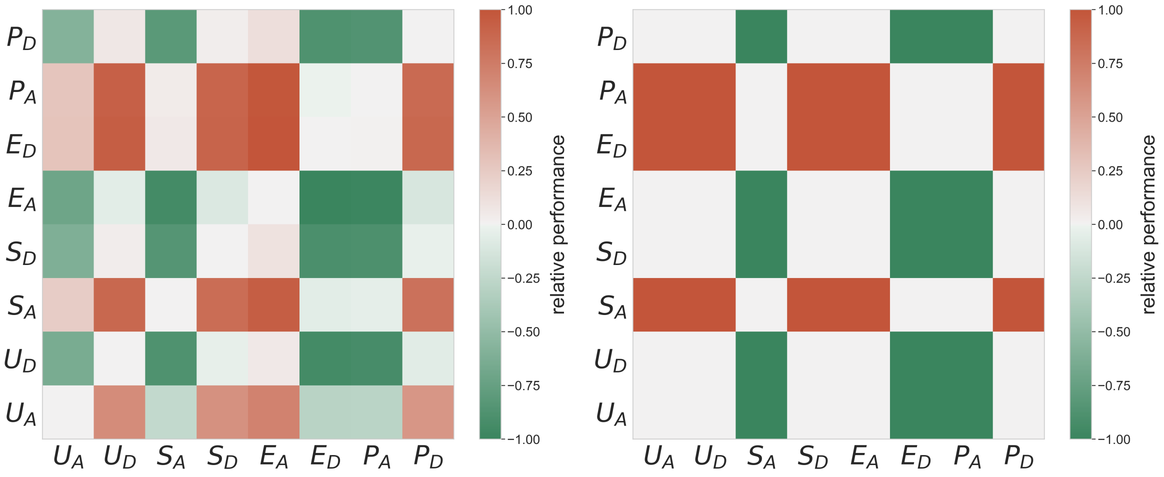
Figure C.11: Relation between Problem Formulations for Median Prioritarian Objectives

A grid cell represents the relation between a row's and a column's problem formulation. Green represents that the row's problem formulation is more desirable than the column's one.

C.4.3. Global Pathways

C.4.4. Regional Pathways

Up to this point in the analysis, we have only considered outcomes on a globally aggregated level. An interesting question would be, whether regional pathways mirror these results or whether they differ. Of course, we cannot present the pathways for all regions and all KPIs in this report. For this purpose, we show a sample with only two regions (US and Africa), four problem formulations (U_A, U_D, P_A, and

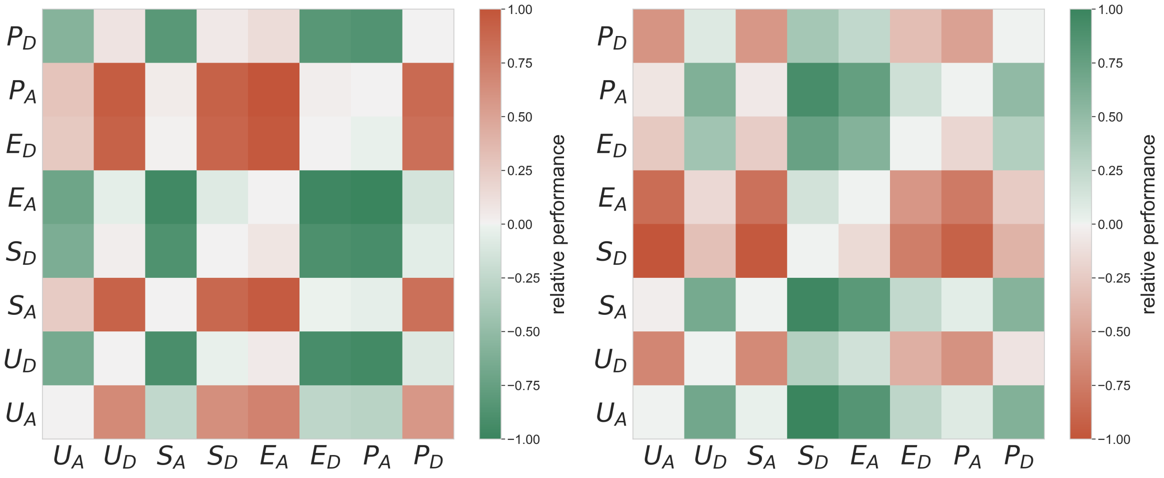


(a) Maximum Distance to Damage Threshold

(b) Population above Damage Threshold

Figure C.12: Relation between Problem Formulations for Median Sufficientarian Damage Objectives

A grid cell represents the relation between a row's and a column's problem formulation. Green represents that the row's problem formulation is more desirable than the column's one.



(a) Economic Damages

(b) GWP

Figure C.13: Relation between Problem Formulations for Median Damages and GWP

A grid cell represents the relation between a row's and a column's problem formulation. Green represents that the row's problem formulation is more desirable than the column's one.

P_D), and two KPIs (consumption per capita (`CPC`) and damage per capita (`DPC`)). Figure C.16 depicts this sample[4].

So, are the regional pathways consistent with the global pathways with respect to whether aggregation level or ethical premise is more crucial? If we look at the first row in Figure C.16, we can see the `CPC` pathways for the US spreading more for the disaggregated problem formulations than they do for the aggregated ones. This points in the direction of aggregation being more important. For Africa, however, the pathways have a stronger resemblance within an ethical premise. Furthermore, we can see that the `DPC` pathways for the US and for Africa have also a stronger resemblance within an ethical

---

[4] For a full overview over all twelve regions and all problem formulations, you can find high-resolution images here for CPC and here for DPC.

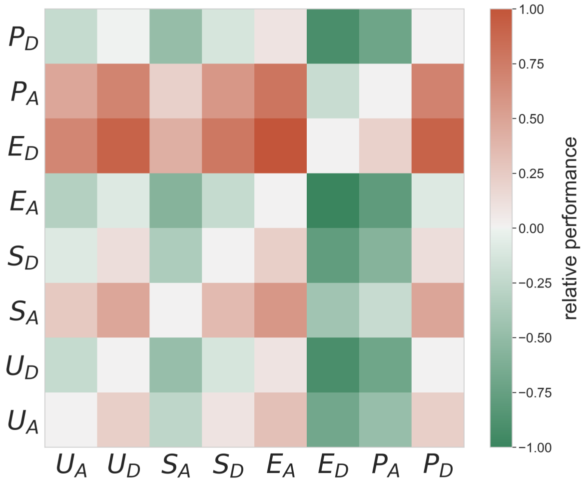


Figure C.14: Relation between Problem Formulations for Median Atmospheric Temperature

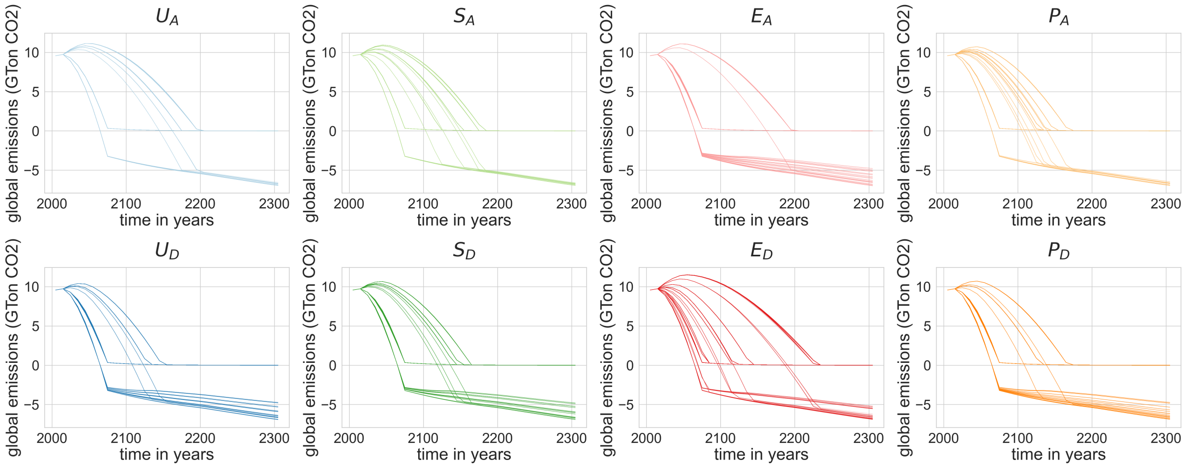


Figure C.15: Global Pathways for Industrial Emissions

Each panel represents the results given the optimal-policies of a problem formulation in combination with 50 vulnerable scenarios. Each combination (policy x scenario) is one line in a panel.

premise. These results are surprising. As `CPC` is strongly connected to `welfare`, we would have expected that they show similar patterns. However, global `welfare` is strongly determined by the ethical premise while regional `CPC` is once stronger determined by aggregation (for the US) and once by the ethical premise (for Africa). The same holds for `welfare loss` and `DPC`. We would have expected the same patterns. Actually, global `welfare loss` is strongly determined by aggregation while regional `DPC` is stronger determined by the ethical premise. Such inconsistent or sometimes mixed results are also found for other problem formulations.

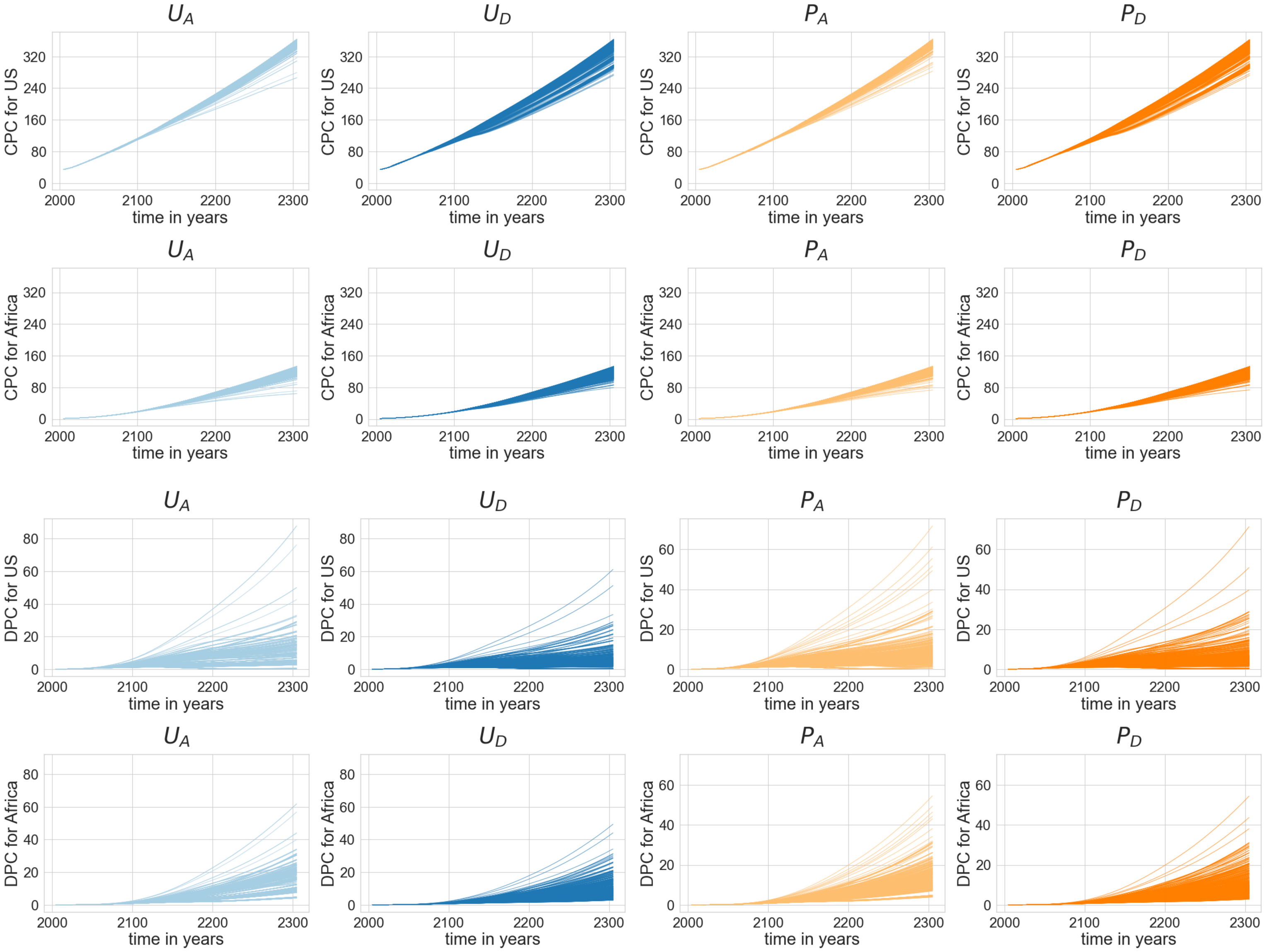


Figure C.16: Regional Pathways: 2 Regions, 4 Problem Formulations, 2 KPIs

Each panel represents the results given the optimal-policies of a problem formulation in combination with 50 vulnerable scenarios. Each combination (policy x scenario) is one line in a panel.

C.4.5. Robustness Analysis

Figure C.17 shows parallel axes plots for all 8 problem formulations. The axes represent the robustness metrics which are corresponding to the three selected KPIs. We have Hurwsicz's optimism-pessimism score for `welfare`, the 90th percentile minimax regret for `welfare loss`, and Starr's domain criterion for `atmospheric temperature increase`. Higher values are more desirable and the limits of all axes are shared between the panels such that the results are better comparable.

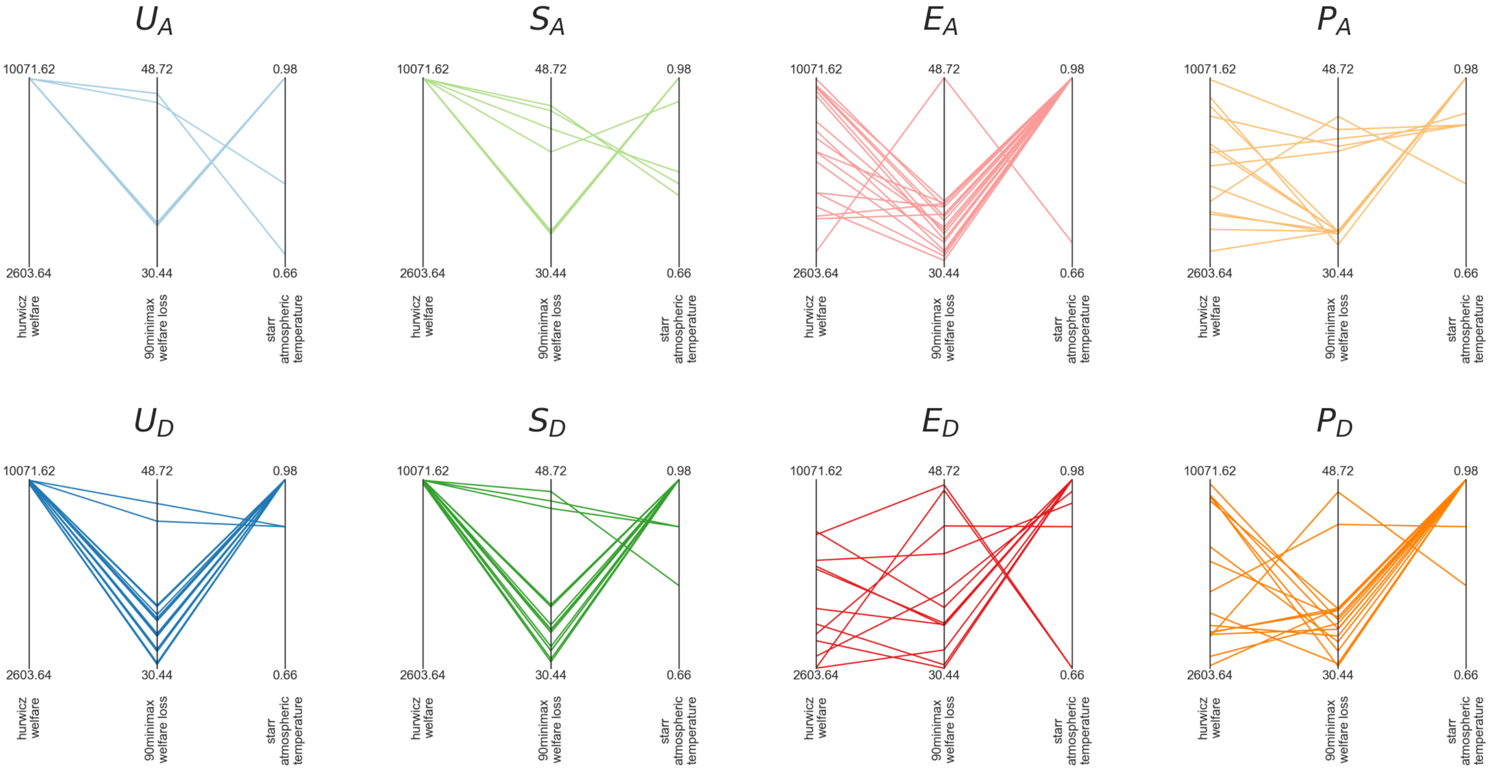


Figure C.17: Robustness: Welfare, Welfare Loss, Atmospheric Temperature Increase

Each panel to one problem formulation. Each line represents a point in robustness space for one policy given a problem formulation. On each axis, top values represent the higher robustness degrees. Each of the objective and their robustness values refer to the year 2105.

Again, the results can be split into two groups, utilitarian and sufficientarian in one group and egalitarian and prioritarian in another group. The trade-offs between the robustness metrics are very similar within the first group. And also again, the level of aggregation can subgroup them even further. The robustness patterns for U_A and S_A look more similar than compare to any other problem formulation. The same holds for U_D and S_D. Interestingly, U_A, U_D, and S_D have a few optimal policies that exhibit a high degree of robustness across the three KPIs. Other problem formulations show a robustness trade-off, with two robustness metrics being high while the other one is low, and vice versa.

C.4.6. Convergence of Directed Policy Search for All Problem Formulations

In Figure C.18, we can see that the egalitarian disaggregated problem formulation does not show a convergence line for `seed 1` in `reference scenario 2` and `reference scenario 3`. This is due to the fact that the data is incomplete. The optimization processes failed in the former case due to a fluke in the ema_workbench.

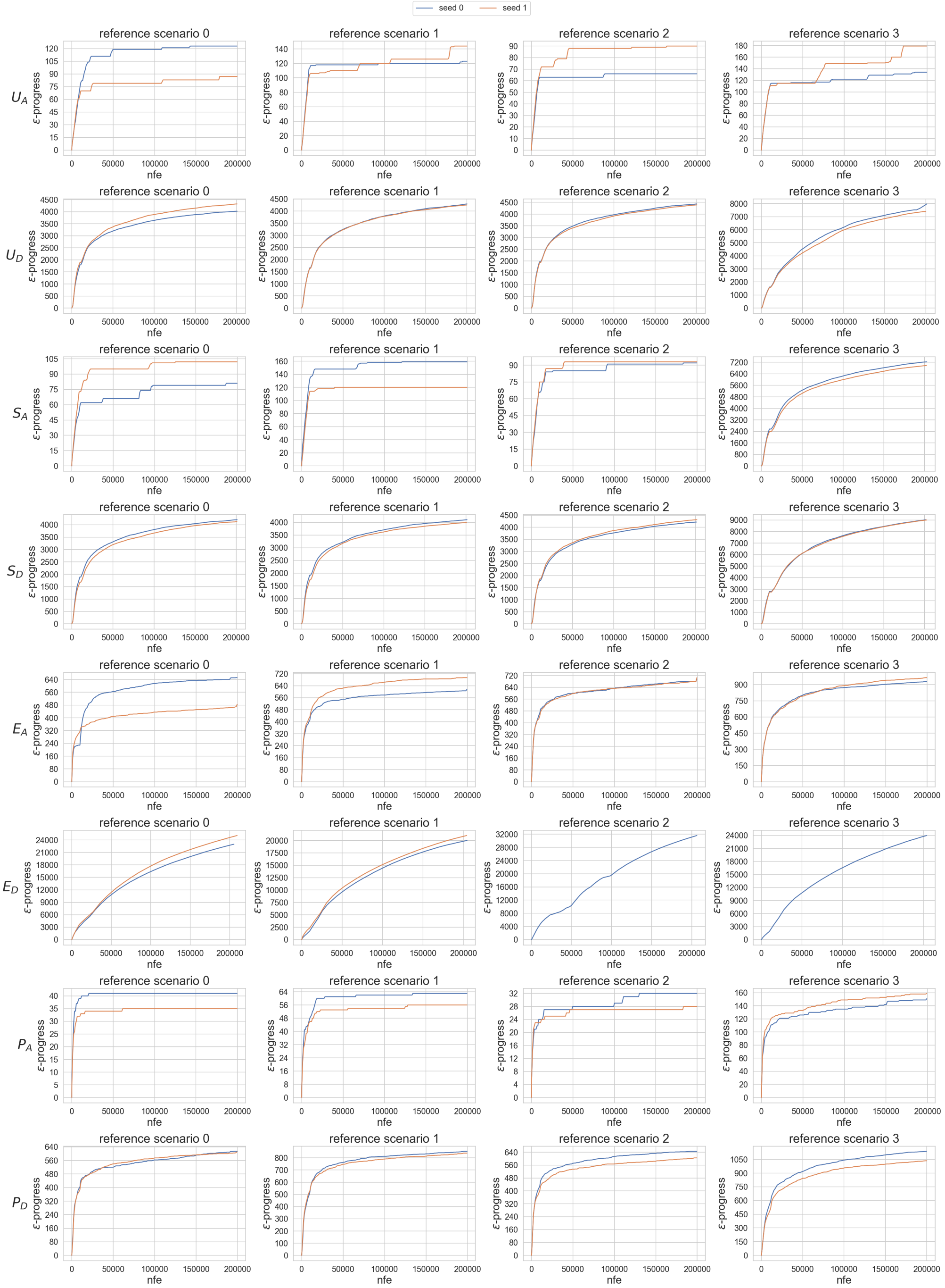


Figure C.18: Convergence of Directed Policy Search via ϵ-Progress

Each row represents a problem formulation. Each column represents the corresponding reference scenario that has been used during the optimization. Each subplot contains two lines for which a different seed has been used.