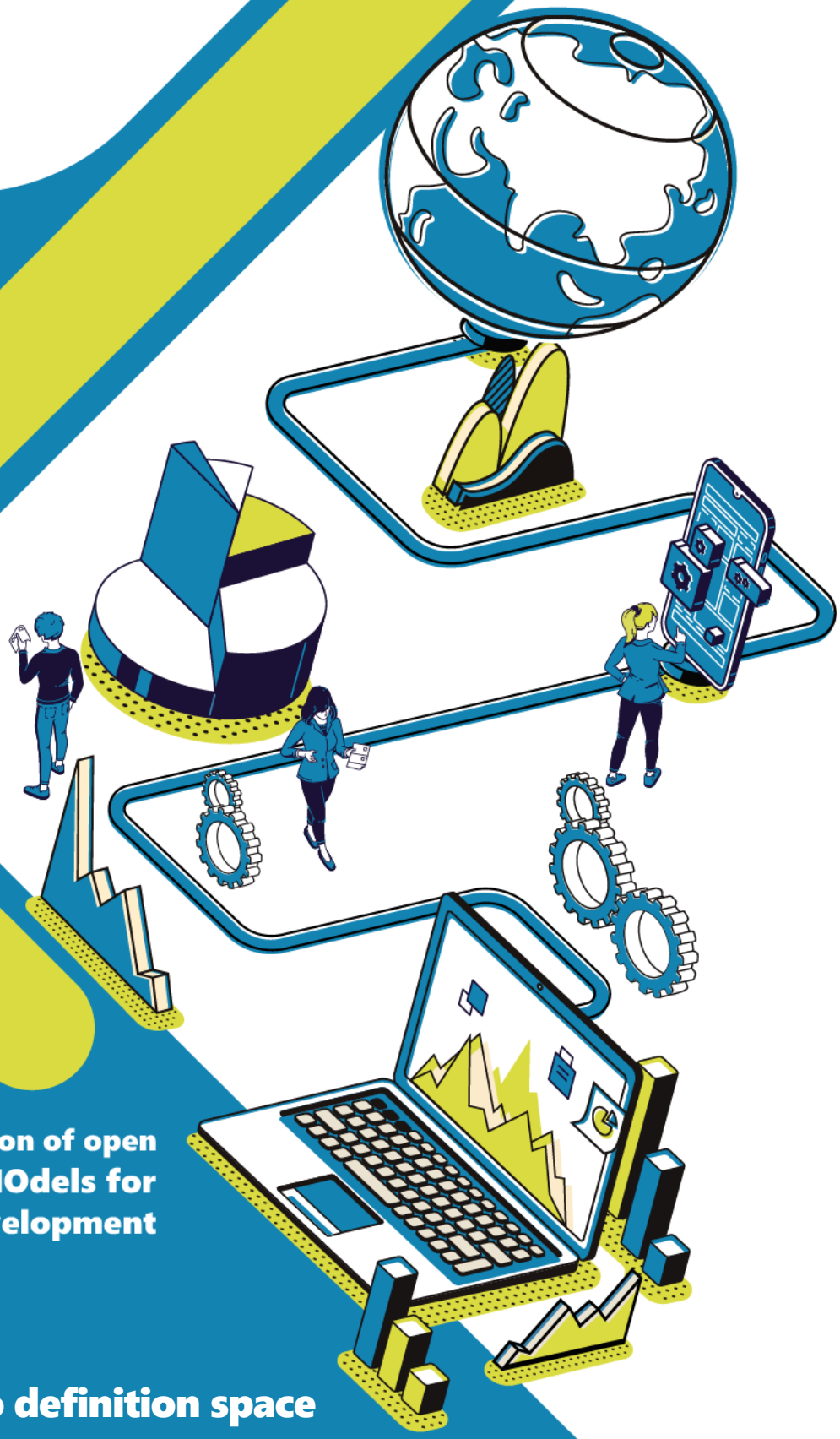
DIAMOND

www.climate-diamond.eu



Delivering the next generation of open Integrated Assessment MOdels for Net-zero, sustainable Development

D5.1 - A novel scenario definition space

WP5 – Apply and Explore

30/01/2025

Funded by
the European Union

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Copyright Message

This report, if not confidential, is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0); a copy is available here: https://creativecommons.org/licenses/by/4.0/. You are free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) under the following terms: (i) attribution (you must give appropriate credit, provide a link to the license, and indicate if changes were made; you may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use); (ii) no additional restrictions (you may not apply legal terms or technological measures that legally restrict others from doing anything the license permits).

Grant Agreement Number	**101081179**		**Acronym**	**DIAMOND**
Full Title	Delivering the next generation of open Integrated Assessment MOdels for Net-zero, sustainable Development			
Topic	HORIZON-CL5-2022-D1-02			
Funding scheme	HORIZON EUROPE, RIA – Research and Innovation Action			
Start Date	December 2022	**Duration**	48 Months	
Project URL	http://www.climate-diamond.eu/			
EU Project Advisor	Silvia Vaghi			
Project Coordinator	Institute of Communications and Computer Systems – ICCS			
Deliverable	D5.1 – A novel scenario definition space			
Work Package	WP5 – Apply and Explore			
Date of Delivery	**Contractual**	31/01/2025	**Actual**	30/01/2025
Nature	Report	**Dissemination**	Public	
Lead Beneficiary	CICERO Senter for Klimaforskning (CICERO)			
Responsible Author	Glen Peters	**Email**	glen.peters@cicero.oslo.no	
	CICERO	**Phone**	+4792891638	
Contributors	Glen P Peters, Jan Ivar Korsbakken, Shivika Mittal, Ben Sanderson, Ida Sognnæs (CICERO); Stephanie Briers, Bianca Vienni Baptista (ETH); Alexandros Nikas, Konstantinos Koasidis (ICCS); Alaa Al Khourdajie (Imperial);			
Reviewer(s)	Maro Malliaroudaki (Imperial); Dirk-Jan Van de Ven (BC3); Maximov Gajardo, Serguey (UCL); Baptiste Boitier (ERASME)			
Keywords	Scenarios, Storylines, Vetting, Infilling, Harmonisation, Climate Assessment			

EC Summary Requirements

1. Changes with respect to the DoA

No changes with respect to the work described in the DoA.

2. Dissemination and uptake

The report offers guidance to consortium members to help them better understand the uncertainties in the scenario pipeline, from scenario development through to scenario assessment. Methods to better characterise and understand the scenario literature and scenario databases are also discussed. The report also serves as a resource for entities outside the consortium (including both researchers and policy/other stakeholders), explaining the sources of uncertainties at different stages in the scenario development and assessment process.

3. Short summary of results (<250 words)

The EU Horizon Europe project DIAMOND aims to enhance six key Integrated Assessment Models (IAMs) for better climate scenario analysis. Task 5.1 focuses on characterising the scenario space and identifying uncertainties within it. It seeks to improve scenario assessments and associated uncertainties by integrating additional data into existing scenarios and addressing gaps in the current scenario literature. The scenarios used in the IPCC are limited in scope and often representative of neither the entire literature nor the broader space of potential scenarios. We demonstrate that the scenario space is not adequately characterised with the current generation of scenarios and, further, that we are unable to know whether the current sampling approach using the literature or databases is sufficient to provide a meaningful statistical representation of the full scenario space. We expect that this will remain the case in the future, simply because scenarios are not a random sampling of physical properties, but rather contemporary snapshots of areas of interest in an ever-evolving research landscape. Within the report, we outline the process for DIAMOND to employ stakeholder engagement and transdisciplinary approaches to develop a new scenario space, producing 'storyline' scenarios that capture key uncertainties in scenario and climate outcomes, with the aim of increasing relevance for scenario users. This report summarises the uncertainties in the scenario pipeline, while the subsequent report associated with this task (D5.2, July 2026) will then analyse how DIAMOND scenarios expand the scenario definition space.

4. Evidence of accomplishment

This report.

Preface

DIAMOND will update, upgrade, and fully open six Integrated Assessment Models (IAMs) that are emblematic in scientific and policy processes, improving their sectoral and technological detail, spatiotemporal resolution, and geographic granularity. It will further enhance modelling capacity to assess the feasibility and desirability of Paris-compliant mitigation pathways, their interplay with adaptation, circular economy, and other SDGs, their distributional and equity effects, and their resilience to extremes, as well as robust risk management and investment strategies. This will be done via integration of tools and insights from psychology, finance research, behavioural and labour economics, operational research, and physical science. The project will develop a transdisciplinary scientific approach to legitimise the implementation process and co-create research questions that stretch the frontiers of climate science, as well as establish vibrant communities of practice to transparently open model enhancements and to develop capacities, thereby lowering the entrance barriers to the established IAM community.

ICCS	INSTITUTE OF COMMUNICATIONS AND COMPUTER SYSTEMS	EL
BC3	ASOCIACION BC3 BASQUE CENTRE FOR CLIMATE CHANGE - KLIMA ALDAKETA IKERGAI	ES
CESAR	KRATENA KURT	AT
CICERO	CICERO SENTER FOR KLIMAFORSKNING	NO
CYI	THE CYPRUS INSTITUTE	CY
E4SMA	ENERGY ENGINEERING ECONOMIC ENVIRONMENT SYSTEMS MODELING AND ANALYSIS SRL	IT
HOLISTIC	HOLISTIC IKE	EL
COMILLAS	UNIVERSIDAD PONTIFICIA COMILLAS	ES
ISINNOVA	ISTITUTO DI STUDI PER L'INTEGRAZIONE DEI SISTEMI (I.S.I.S) - SOCIETA'COOPERATIVA	IT
SEURECO	SEURECO SOCIETE EUROPEENNE D'ECONOMIE SARL	FR
UM	UNIVERSITEIT MAASTRICHT	NL
ESMIA	ESMIA CONSULTANTS INC.	CA
USMF	THE UNIVERSITY OF MARYLAND FOUNDATIION INC	US
UMD	UNIVERSITY OF MARYLAND	US
EPFL	ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE	CH
ETH	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	CH
UNIBAS	UNIVERSITAT BASEL	CH
Imperial	IMPERIAL COLLEGE OF SCIENCE TECHNOLOGY AND MEDICINE	UK
Oxford	THE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY OF OXFORD	UK
UCL	UNIVERSITY COLLEGE LONDON	UK

Funded by
the European Union

Executive Summary

The EU Horizon Europe project "Delivering the next generation of open Integrated Assessment MOdels for Net-zero, sustainable Development" (DIAMOND) is tasked with updating, upgrading, and fully opening six Integrated Assessment Models (IAMs) that are emblematic in scientific and policy processes, improving their sectoral and technological detail, spatiotemporal resolution, and geographic granularity. These IAMs then will be applied to user-driven research questions to explore the new modelling capabilities of the IAMs and expand the 'scenario space'.

The hypothetical scenario space spanning all potential and available futures is difficult to characterise and identify gaps towards fully exploring uncertainties. In DIAMOND Task 5.1, the goal is to therefore better characterise the scenario space and better identify uncertainties. The uncertainties considered cover both the scenario space and the scenario climate assessment pipeline.

The scenario **climate assessment pipeline** takes existing scenarios and supplements them with additional data to provide a more comprehensive climate assessment, since scenarios are often constrained by cumulative CO_2 emissions or radiative forcing as a proxy for climate, and a more comprehensive climate assessment is added later, in a post-processing step.

In the process of developing a climate assessment, several steps are often required, which modify the original scenario(s) to facilitate consistency. Depending on the research question, scenarios may go through a variety of post-processing steps to help select the most relevant to the research question (e.g., vetting against relevant criteria). The climate assessment process might require modification of the scenario output emissions to match the emissions data used in the assessment (harmonisation) as well as supplementing the scenario output with additionally climate-relevant species (e.g., infilling the scenarios with sulphur emissions if these are not provided by the model). After these steps, a simple climate model is used to estimate temperature and uncertainties. All these steps in providing a climate assessment create additional uncertainties, which are poorly constrained. In this deliverable, we outline the key uncertainties and discuss how they may be reduced in further work in DIAMOND.

The hypothetical scenario space is only partially filled. The scenarios assessed in the IPCC process are only a subset of the scenarios available in the literature. Further, they are not a random sample from the literature but based on voluntary submissions leading to an uneven representation of models and scenarios. We provide a more thorough conceptual background to the scenario space and propose ways to better characterise it to help identify gaps. Later in the DIAMOND project, we will assess how the scenarios co-created in the DIAMOND project may modify and fill gaps in the scenario space.

Finally, we outline the process to develop a novel scenario space for the DIAMOND project. This is based on a co-creation process utilising a variety of stakeholder engagement activities. These activities are ongoing, with the results of the process expected to be documented and discussed later in the project. The project will ultimately provide a set of 'storyline' scenarios to span relevant parts of the scenario space and provide a more comprehensive assessment of uncertainties in climate outcomes, reflecting the intrinsic uncertainties in scenario selection, harmonisation, infilling, land use assumptions, and climate response.

This first report of the outcomes throughout this task/process (D5.1, in January 2025) focuses on a systematic review of the scenario pipeline uncertainties and the scenario definition space. The second report (D5.2, due in July 2026) will perform a deeper analysis on the scenario pipeline uncertainties and identify how DIAMOND scenarios broaden the scenario definition space.

Funded by
the European Union

Contents

1 Introduction 8
1.1 Task Description 8
1.2 Scope 8
2 The Scenario Pipeline 9
2.1 Raw scenario data, quality control, and scenario databases 10
2.2 Scenario Selection and Vetting 17
2.3 A climate assessment pipeline 25
2.4 Harmonisation 26
2.4.1 Harmonisation of inputs versus outputs 26
2.4.2 What to harmonise, and past vs. future harmonisation 28
2.4.3 Socioeconomic data (population and GDP) 29
2.4.4 Emissions (for input to climate assessments) 30
2.4.5 Energy 35
2.4.6 Costs, performance, prices, policies, other parameters, and model assumptions 36
2.5 Infilling 36
2.6 Simple climate modelling 39
2.7 Land-use uncertainties 40
2.7.1 Background 40
2.7.2 Geological net zero emissions 41
2.7.3 Modelling land-use change 42
3 Scenario Definition Space 44
3.1 Emissions scenarios 45
3.2 The scenario space 46
3.2.1 Narrative (qualitative) scenarios 47
3.2.2 Quantitative scenarios 48
3.2.3 The hypothetical scenario space 51
3.3 Characterising the scenario space 53
3.3.1 Literature review 54
3.3.2 Quantitative methods 55
3.3.3 Co-designing the scenario space 56
3.4 Scenario assessment in limited scenario spaces 57
4 A plausible set of qualitative scenarios 59
4.1 Background 59
4.2 Co-design process in DIAMOND 61
4.2.1 Storylines based on foresight workshops 61
4.2.2 Co-creating research questions for scenario development 62
5 Bibliography 65

Table of Figures

Figure 1. A schematic overview of the scenario pipeline (assessed in Section 2), without showing the scenario definition space (assessed in Section 3). Harmonisation in the "climate assessment pipeline" refers to output emissions data. A tool has been developed in IAM-COMPACT, an EU-funded research project under the Horizon Europe framework programme, to ensure the quality of the scenarios before they are submitted to scenario database 10
Figure 2. The distribution of model-scenario combinations (scenarios for short) in the AR6 IPCC database across model families and projects, depicting the share of scenarios failing vetting due to insufficient consistency with historical energy and emissions data (yellow), those passing vetting but lacking sufficient data to perform a climate assessment (blue), and those eventually used in the primary assessment of scenarios (green). Source: Peters et al. (2023a) 12
Figure 3. The median and scenario distribution for the variable "global non-biomass renewable primary energy" as well as "primary solar energy" in 2050 under scenario C1 (Below 1.5C with no or limited overshoot) in three cases: M0 (includes results from all models), DM (includes results only from the dominant model), and -DM (excludes results from the dominant model). The entire distribution (grey and brown) shows the distribution across all C1 scenarios, while the colours represent the distribution of the DM (brown) and existing the DM (grey), In these examples, the distribution of the DM is substantially different to the other models. 13
Figure 4. Distribution of scenarios by scenario family in SR1.5 and AR6 database (D. J. van de Ven et al., 2025).. 15
Figure 5. The distribution of variables reported in the scenarios database, in the C1 climate category, for submitted scenarios (top) and models (bottom). While this is shown for C1, the results are similar for other climate categories, with the distributions scaled up or down depending on the numbers of scenarios submitted. 16
Figure 6. Distribution of scenarios in the AR6 IPCC database across model families after conducting historical vetting for primary energy sources: coal, gas, and oil for the year 2020. Additional vetting is performed on scenarios approved for climate assessment (n=1202). 19
Figure 7. a) GDP (MER) and b) population at regional level(R10) for historical years 21
Figure 8. (a) Distribution of scenarios based on the cumulative CO_2 budget between 2020 and 2050 (b) range of % CO_2 emissions reduction between 2020 and 2030 within each cumulative CO_2 budget category. 24
Figure 9. Reproduction of Figure 2 from Kikstra et al. (2022), showing the effect of harmonisation and infilling on emission timeseries and resulting climate characteristics of scenarios in the AR6 Working Group III database. Original caption: Summary statistics (a: emissions, b: atmospheric concentrations, c: global mean surface temperature) over time across all scenarios in the AR6DB that received a temperature classification and across scenarios in AR6 temperature categories C1 and C3. Panel (a) shows emissions as modelled by IAMs ("Native"), after harmonisation ("Harmonized"), and after infilling missing reported emissions ("Infilled"). Panel (b) and panel (c) show climate outcomes per climate model, using the median value of each variable from the climate emulator probabilistic distributions. 32
Figure 10. The global carbon budget showing sources (fossil emissions and *active* land-use change), passive sinks, and the atmospheric increase 40
Figure 11. A schematic showing the translation from the Global Carbon Budget (GCB) definition of anthropogenic land-use change to the national GHG reporting definition. The GCB definition (Step 1) can be split into managed and unmanaged land, using national reporting (Step 2). The managed sources (LUC) and sinks can then be merged to translate to the national reporting definitions defined in UNFCCC (Step 3). In each step the net land emissions are the same. Source: Friedlingstein et al. (2023) 41
Figure 12. Different definitions of net zero emissions (left) lead to different concentrations (middle) and warming

(right). If net zero emissions are defined based on active LUC emissions only, then concentrations will decline, and temperatures stay constant. If net zero includes passive uptake on land and oceans (sinks), then net zero emissions would only stabilise concentrations (as the net flux into the atmosphere is zero) and therefore rising temperatures. Figure from Allen et al (2024). 42
Figure 13. The SSP-RCP combinations placed on the SSP narrative axes of challenges to mitigation and challenges to adaptation, illustrating how primary energy evolves under different SSPs (boxes) and different SPAs (colours). 50
Figure 14. Population projections based on probabilistic scenarios from the UN Population Division (shading showing mean in black, and the interquartile and full range in grey) and the second release of the SSP population data 51
Figure 15. A schematic of a hypothetic scenarios space (light grey). The current scenarios space (dark grey is partially captured by the AR6 Scenarios Database (blue) of which a smaller subset is vetted and assessed. Some scenarios or groups of scenarios are indicated by black dots, which primarily fall within the scenarios assessed by the IPCC. The SSPs are shown in the white box with quadrants. Other prominent scenario exercises are shown as examples (e.g., IEA and NGFS), some of which are assessed also by the IPCC and in the AR6 Scenarios Database. Over time, more of the scenario space is filled, either by infilling with more detail or expanding to new areas, with illustrations showing climate impacts, finance, circular economy, and land use (green). Other scenarios might be further from existing scenarios, such as ones dominated with Artificial General Intelligence (AGI) or nuclear fusion, to show two random examples. The reality is far more complex, but this schematic is to demonstrate that the existing scenario space is limited in extent and key dimensions are not fully explored 53
Figure 16. The quadrants of the Shared Socioeconomic Pathways (SSPs) spanning axes of challenges to mitigation and challenges to adaptation, with a middle of the road at the origin. 60
Figure 17. Hypothetical example of uncertainty in climate, land use, and socioeconomics spanning axes of short-term and long-term climate action 61
Figure 18. The Three Horizons conceptual framework used in the First Foresight Workshop. Figure Source: Adapted to the DIAMOND workflow based on Schaal, Mitchell, Scheele, Ryan, and Hanspach (2023). 62

Table of Tables

Table 1. Vetting criteria and ranges applied to the global scenarios submitted to the IPCC database for the climate assessment (IPCC, 2022a). 18
Table 2. Number of unique vetted scenarios in each category in the IPCC AR6 scenario database. A unique scenario here is defined as implementing a unique scenario definition (e.g., EN_NPi2020_400 in the AR6 scenarios database), while the bottom row shows the model-scenario combinations (multiple models run the same scenario definition in multi-model exercises) 22
Table 3. Different scenario types that exist in the hypothetical scenario space 52

1 Introduction

The EU Horizon Europe project "Delivering the next generation of open Integrated Assessment MOdels for Net-zero, sustainable Development" (DIAMOND) is tasked with updating, upgrading, and fully opening six Integrated Assessment Models (IAMs) that are emblematic in scientific and policy processes, improving their sectoral and technological detail, spatiotemporal resolution, and geographic granularity. These IAMs then will be applied to user-driven research questions to explore the new modelling capabilities of the IAMs. Conceptually, the DIAMOND project will develop new scenarios that can expand and enhance the existing set of scenarios in the literature in two key dimensions: user relevance and model capabilities.

This deliverable represents work undertaken in the first part of the application and exploration activities (WP5) and will be updated 18 months later as more work is undertaken. In this deliverable (D5.1), we define the scenario definition space and outline key uncertainties in the scenario development process leading up to, and including, climate impacts. Through this process we identify some of the key knowledge gaps to be filled in the second deliverable (D5.2).

In the remainder of this introductory section, we outline the task description and scope of the deliverable, before describing the scenario pipeline (Section 2), discussing the scenario definition space (Section 3), and linking to the co-design process (Section 4, as well as work undertaken and reported in WP2).

1.1 Task Description

This task conducts a systematic review of potential sources of bias in the scenario pipeline from policy to emissions to climate outcome, drawing on advances/insights from DIAMOND WP4.

First, we will focus on uncertainties due to harmonisation and infilling assumptions. Using existing tools (e.g., SILICONE) in 'perfect model framework' as well as existing (and new) DIAMOND scenario databases, we will consider biases in infilled data where true data is known. We will explore dependencies of infilled results on training data and represent uncertainties in trajectories as a range of plausible infilled scenarios.

Second, we will explore uncertainty from over-simplistic or over-confident climate response assumptions, building on the work of Task 4.4. We will use perturbed ensembles of OpenSCM climate emulators (CICERO-SCM, FAIR, MAGICC) to quantify both parametric and structural uncertainties.

Third, we will evaluate biases in land management assumptions, arising from insufficiently coupled or biased representations of potential land use productivity. We will draw objective comparisons of productivity and cost assumptions across DIAMOND scenario databases, in the context of larger ensembles of complex land surface models from the LUMIP and C4MIP projects.

We will finally deliver an approach to providing a set of 'storyline' scenarios that span a plausible set of infilled solutions to help expand the scenario space, allowing a native assessment of uncertainties in climate outcome and impacts assessments for a given scenario class, reflecting the intrinsic uncertainties associated with harmonisation, infilling, land use assumptions, and climate response.

1.2 Scope

This first report of the outcomes throughout this task/process (D5.1, in January 2025) focuses on a systematic review of the scenario pipeline uncertainties and the scenario definition space. The second report (D5.2, due in July 2026) will perform a deeper analysis on the scenario pipeline uncertainties and identify how DIAMOND scenarios broaden the scenario definition space.

2 The Scenario Pipeline

The *scenario pipeline* is the broad flow of activities from scenario development or curation through to scenario assessment (Figure 1). We have depicted the scenario pipeline as a linear process, but in practice there may be several branches or feedbacks, and only parts of the scenario pipeline may be used in particular applications. However, our depiction of the scenario pipeline is intended to provide a complete overview, so that all the key elements of the scenario pipeline can be captured. The key components of the scenario pipeline consist of:

- Raw scenario data: the collection of the raw scenario data, which may be the output of a single- or multi-model study addressing a particular research question. The scenario would ideally be generated from a clear scenario protocol that would specify criteria such as input data, key assumptions, methods of implementation, etc. (Korsbakken & Peters, 2023).
- Quality control: The quality control process ensures that the scenario meets various requirements related to variable definitions, sign conventions, unit conventions, and similar. This quality control is done before (or while) the scenario is uploaded to a shared scenario database. Such a quality control process is documented in the I^2AM PARIS platform (Korsbakken & Peters, 2023).
- Scenario database: Ideally, after quality control, the scenario is placed in a database in a harmonised format and with easy access to facilitate further processing and analysis.
- Vetting: The scenario may undergo a vetting process to ensure it meets the specific requirements for its intended use, noting that different applications may require different vetting criteria. The outcomes of the vetting process would be added to the scenario metadata to allow reuse.
- Diagnostics: Depending on the needs, various standardised diagnostics may be run for each model or scenario, particularly to facilitate validation as well as comparability in multi-model studies. The diagnostic data would be added to the scenario metadata to allow reuse.
- Climate assessment pipeline: Depending on the intended research question, the scenario may require a climate assessment (Kikstra et al., 2022). The climate assessment pipeline used in Working Group III (WGIII) of the IPCC's 6th Assessment Report (AR6) performed a relatively comprehensive assessment of climate impacts (temperature) for qualifying scenarios, by applying simple climate models to emissions produced by each scenario, although different approaches may be used. We frame out discussion around the AR6 climate assessment pipeline, which consisted of three broad steps:
    - Harmonisation: It must be ensured that the scenarios' emissions data merges consistently to the historical emissions data used in the climate assessment in the overlapping calibration period. The harmonisation in the scenario pipeline in IPCC AR6 only covered emissions, but in principle other variables could also be included.
    - Infilling: Some scenarios need to be supplemented with additional data that are required by the simple climate model in the next step, such as emissions of black carbon or SO_2, if they are not modelled specifically.
    - Climate assessment: The complete scenario data is put into a Simple Climate Model (SCM) to obtain an assessment of the climate outcomes, often in probabilistic format. The results of this process are *ideally* merged with the scenario data in the database.

The simple schematic of the scenario pipeline (Figure 1) does not capture the hypothetical scenario space covering a potentially infinite number of scenarios but rather illustrates the scenario pipeline to produce a single- or multi-

model scenario that is introduced in the scenario space. Each scenario is a single point in a much bigger *scenario space*. This scenario space is difficult, if not impossible, to fully explore, and is a key uncertainty when drawing conclusions from the available scenarios (a subset of the scenario space). The concept of a scenario space is discussed in Section 3.

In the following section, we outline some of the key characteristics and uncertainties in the scenario pipeline. The uncertainties will be further explored and quantified in the DIAMOND scenario database in D5.2 (due in July 2026). In this section, we discuss scenario pipelines and associated uncertainties in general, with a focus on the scenario pipeline in AR6. We also propose analyses of that uncertainty and potential alternatives to the choices made in the AR6 pipeline, which we intend to carry out in subsequent work in DIAMOND and report in D5.2. We do not propose an improved pipeline for DIAMOND or make recommendations on improving the AR6 pipeline in this deliverable but will build on the proposed analyses to do so in D5.2.

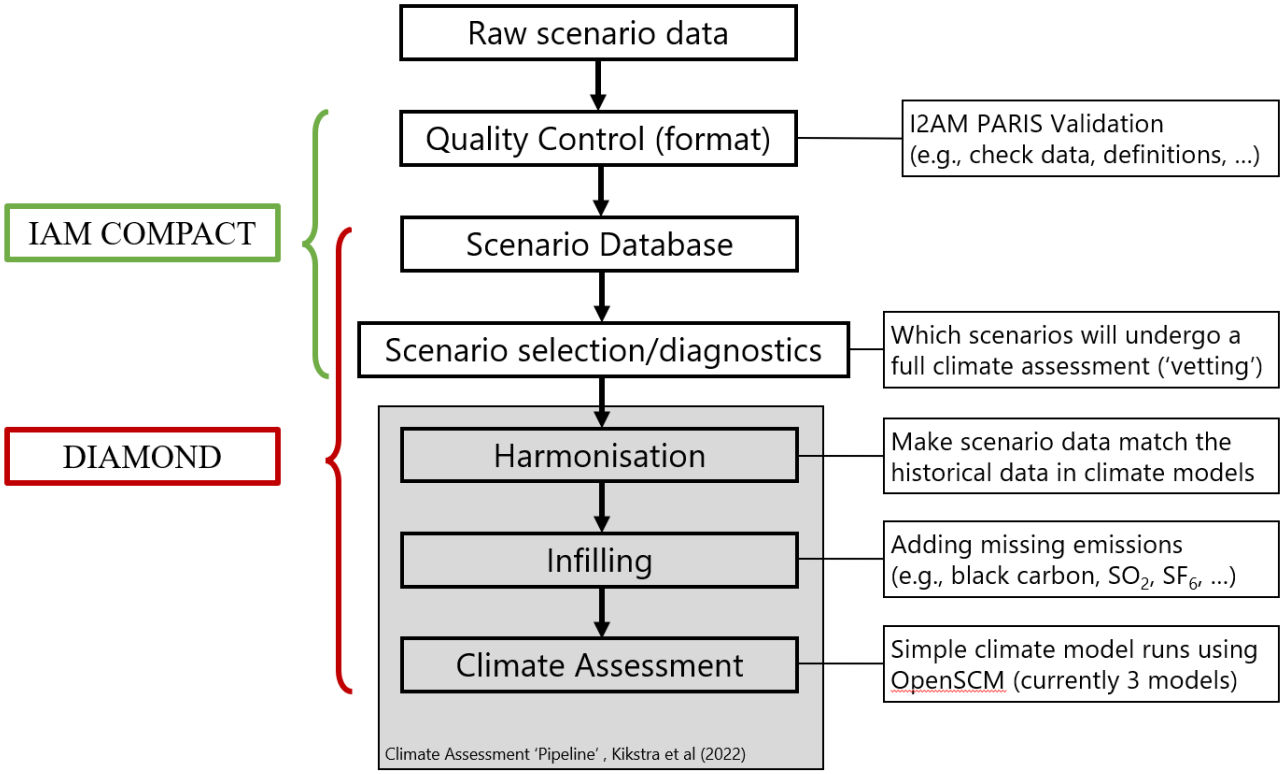


Figure 1. A schematic overview of the scenario pipeline (assessed in Section 2), without showing the scenario definition space (assessed in Section 3). Harmonisation in the "climate assessment pipeline" refers to output emissions data. A tool has been developed in IAM COMPACT[1], an EU-funded research project under the Horizon Europe framework programme, to ensure the quality of the scenarios before they are submitted to scenario database

2.1 Raw scenario data, quality control, and scenario databases

The development of a scenario should be driven by a research question, and this should encompass decisions on input data and key model assumptions necessary to robustly answer the research question. Ideally, a detailed scenario protocol should describe how the research question is translated into a scenario (Korsbakken & Peters, 2023). In many cases, the scenario assessment may be performed using multiple models, either in cascade settings

[1] https://www.iam-compact.eu/

to help expand the details or capabilities of a model (cascade modelling framework), or in parallel to capture a broader action/results space according to different modelling theories and structures and/or reflect important structural uncertainties (or model inter-comparison). To facilitate a robust model inter-comparison, care must be taken on deciding what components are harmonised across models and to which sources (S. Giarola et al., 2021). Once the scenario is generated, it should be available in some form of digital database, even if not formally defined and publicly available.

When compiling a formal (and ideally public) database, it is necessary to perform a minimum level of quality control. The quality control is to ensure the scenario data meets the minimum requirements of the database in terms of variables, definitions, units, and so on. While this is a seemingly obvious step in a database curation process, a formal and standardised process is preferable given the increasing complexity of databases and variables. Primary energy accounting conventions, sign conventions, unit conventions, missing data, etc do not necessarily invalidate a scenario, but the use of consistent conventions is necessary for comparability.

Emissions scenarios have been compiled into some form of database since at least the IPCC First Assessment Report (Cointe, 2024; Peters, Al Khourdajie, Sognnaes, & Sanderson, 2023a). Over time, these scenario databases have grown in number and size. Most scenario databases are now hosted by the International Institute for Applied Systems Analysis (IIASA), using their scenario explorer functionality, which provides easy access to the data and also facilitates interactive exploration of the data using tools designed by the community (Peters et al., 2023a). The IIASA infrastructure is utilised by many EU-funded research projects involving IIASA[2] (D. J. van de Ven et al., 2025). Scenario databases used in IPCC reports are also hosted by IIASA. Alternative database infrastructures include the I^{2}AM PARIS platform[3]. Scenario databases are often constructed around projects and particularly model inter-comparison exercises, also known as model intercomparison projects or MIPs, such as the "Where are we heading" scenarios[4,5] (A. Nikas et al., 2021; Sognnaes et al., 2021); (Nikas et al., 2021). Model inter-comparison studies can also harmonise model inputs and assumptions to enhance comparability (S. Giarola et al., 2021). In the context of the IPCC, scenario database curation is generally based on voluntary submissions (Peters, Al Khourdajie, Sognnaes, & Sanderson, 2023b), comprising single- and multi-model studies, submitted by researchers or research groups interested in having their scenarios featured in the IPCC assessment reports. In this and similar cases, making inter-model and inter-study comparisons may require mitigating inconsistencies by harmonising the model results through post-processing (see Section 2.4) and/or by filtering out unwanted scenarios by using vetting criteria (see Section 2.2).

The scenario databases used in IPCC reports are unstructured, with an uneven distribution of scenarios across projects and models (Peters et al., 2023b) (Figure 2). Simple statistical analysis (means, percentiles) are often performed on these databases (K. Riahi et al., 2022), despite the databases unlikely representing a true statistical sample. Even if the database was constructed to be a representative statistical sample of the existing scenarios, it will remain uneven due to the scenarios that exist but not yet developed (see Chapter 3). The compilation of scenarios into a database is not like compiling estimates of physical quantities. In the physical sciences, increasing the number of estimates of a particularly physical value, will all-else-equal, lead to a better estimate of the mean and distribution (e.g., Law of Large Numbers). Scenarios, on the other hand, are all attempting to assess different research questions and estimate different quantities, and the resulting scenarios are sometimes called an 'ensemble of opportunity' based on the available scenarios(Huppmann, Kriegler, et al., 2018). While it may appear

[2] https://iiasa.ac.at/scenario-ensembles-and-database-resources
[3] https://www.i2am-paris.eu/
[4] https://www.i2am-paris.eu/pr_wwh/landing
[5] https://www.i2am-paris.eu/eu_wwh/landing

that scenarios are converging to a similar value (e.g., 2030 reduction levels), a new experimental design might lead to a different and equally valid outcome(Keywan Riahi et al., 2021). The uneven distribution of scenarios will likely remain a persistent, and unavoidable, aspect of scenario databases and therefore something that needs to be managed and not ignored. The uneven distribution of models and projects across scenarios has been demonstrated to skew scenario results (Figure 2), with the median of different ensembles most closely linked to the median of the model(s) generating the most scenarios (Figure 3).

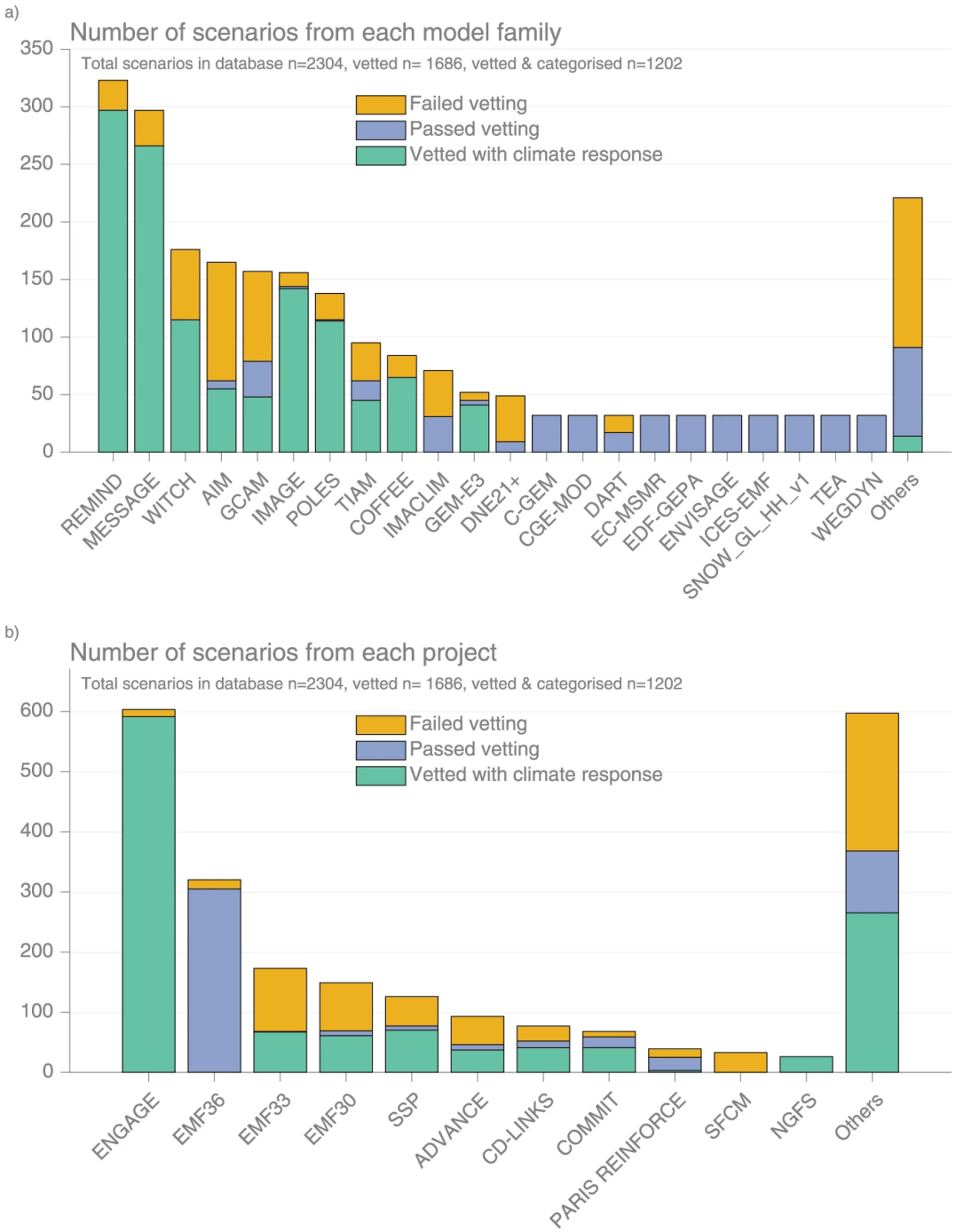


Figure 2. The distribution of model-scenario combinations (scenarios for short) in the AR6 IPCC database across model families and projects, depicting the share of scenarios failing vetting due to insufficient consistency with historical energy and emissions data (yellow), those passing vetting but lacking sufficient data to perform a climate assessment (blue), and those eventually used in the primary assessment of scenarios (green). Source: Peters et al. (2023a)

Funded by
the European Union

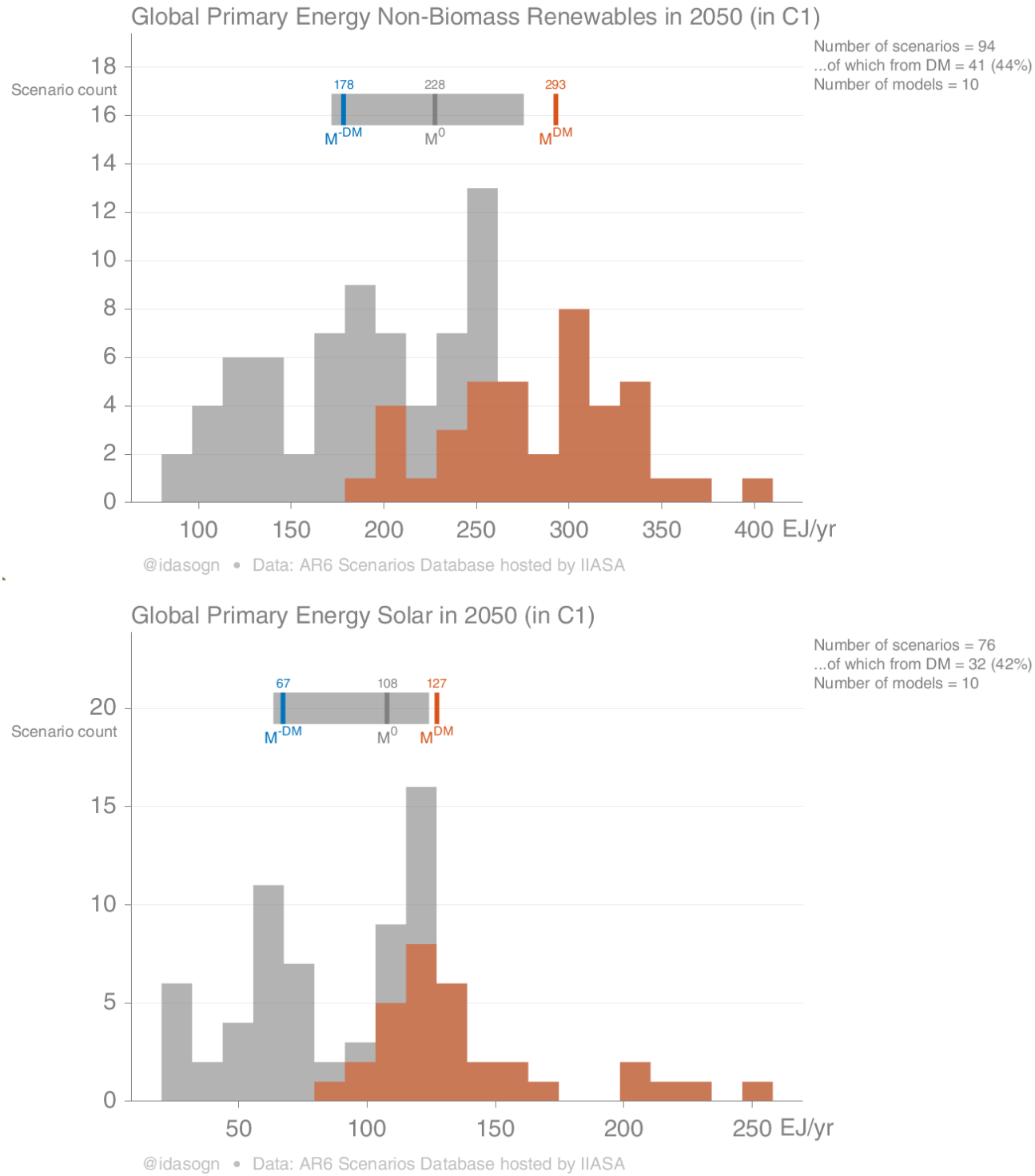


Figure 3. The median and scenario distribution for the variable "global non-biomass renewable primary energy" as well as "primary solar energy" in 2050 under scenario C1 (Below 1.5C with no or limited overshoot) in three cases: M0 (includes results from all models), DM (includes results only from the dominant model), and -DM (excludes results from the dominant model). The entire distribution (grey and brown) shows the distribution across all C1 scenarios, while the colours represent the distribution of the DM (brown) and existing the DM (grey), In these examples, the distribution of the DM is substantially different to the other models.

The uneven distribution across the IPCC scenario database spans many different dimensions, not just models and scenarios—for example, pathways describing a set of socioeconomic inputs. The most common socioeconomic assumptions used in scenarios submitted to the IPCC AR6 database correspond to the so called Shared

Funded by
the European Union

Socioeconomic Pathway (SSP) current trends scenario, SSP2 (Sara Giarola et al., 2024; D. J. van de Ven et al., 2025). Despite the purported strength of the SSP framework to integrate climate research across a variety of socioeconomic backgrounds, this has clearly not happened in practice (Figure 4)(O'Neill et al., 2020; Pirani et al., 2024) Further, since the world is unlikely to be viewed as fitting any particular SSP, with examples of countries mapping to different SSPs, interpretation of SSP outcomes can be ambiguous. Outside of the main SSP studies that focus on the scenario framework and narratives (Keywan Riahi et al., 2017), there have been very few studies comparing outcomes across SSPs—e.g., assessing the mitigation costs associated with meeting various temperature targets under different socio-economic pathways (Keywan Riahi et al., 2017; Rogelj, Popp, et al., 2018). Even if there was an even distribution of scenario exercises across the SSPs, they themselves are also not unique, but rather the outcome of a process involving legacy effects, and particular framings and participants (O'Neill et al., 2017). The SSPs explored "challenges to mitigation" and "challenges to adaptation", but different dimensions could equally be explored, and within those dimensions different story lines (and marker scenarios) could emerge. For example, the SSP1 (sustainability) storyline describes a generally high-growth prosperous society, although others could argue that a sustainability pathway may feature modest growth, or even non-growth (e.g., degrowth, post-growth, or agnostic to growth), and different drivers of change. Indicatively, the well-known Low Energy Demand (LED) scenario (Grubler et al., 2018) is built on SSP2 instead.

The AR6 scenario database also includes individual model runs (single-model studies) or multi model studies with one or two focused scenarios that address specific questions. For example, one such run classified in category C1 (scenarios limiting warming to 1.5°C with low overshoot) is conducted by the model "C-ROADS-5.005" under the scenario "Ratchet-1.5-limCDR-noOS", which focuses on the effects of strict constraints on carbon dioxide removal measures. The results from this scenario assessment show a stringent CO_2 and non-CO_2 emissions reduction trajectory due to limited CDR availability, with the earliest net zero CO_2 date projected in 2037. When combined with the other C1 scenarios (97 in total), it will unlikely have an effect of the distribution percentiles, and thereby the valuable insights provided by this scenario will be lost within the statistical distribution (Al Khourdajie, Skea, & Green, 2024). While any individual model-scenario combination will have a marginal effect on a statistical distribution, some scenarios address particular research questions with insights lost amongst the statistics.

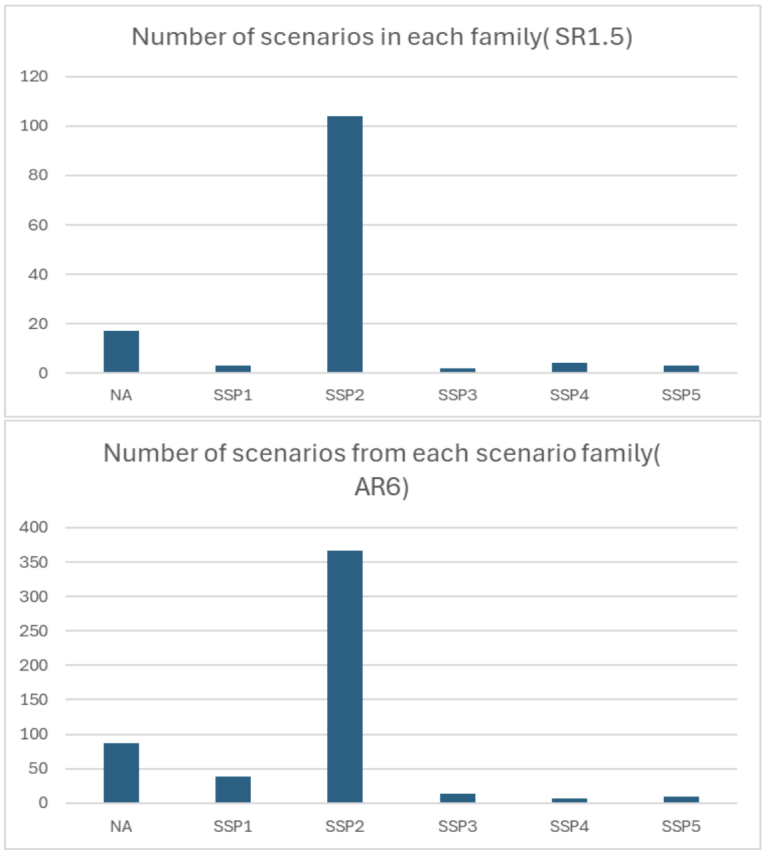


Figure 4. Distribution of scenarios by scenario family in SR1.5 and AR6 database (D. J. van de Ven et al., 2025).

The uneven distribution also varies across variables (Figure 5). The IPCC AR6 scenario database has over 1,400 variables, but no model reports all those variables. Most models report key variables (some of which may be model inputs) like population, GDP, energy use, and CO_2 emissions, while other variables are reported by considerably fewer models. More specific information on variables like water use, technology costs, sub-sector details, labour, etc. are often reported in the AR6 scenarios database by only one or two models, if at all. Of the 97 scenarios in the C1 category, the only variables that all models and scenarios report are CO_2, CH_4, and N_2O emissions, as these are prerequisite for the climate assessment (thus, most of the reported variables are from the climate assessment). While it is no doubt valuable to have models reporting a diverse range of variables, most variables are only reported by few models and the utility of cross-model comparisons is reduced. For example, even though 97 scenarios and 10 models report net CO_2 emissions on land, only 64 scenarios from seven models report afforestation. This illustrates that great care must be taken when analysing different variables from scenario databases. Even in cases where the SSPs were systemically explored, because not all models in the database report every variable, certain indicators have remained unreported in some runs, potentially inflating or deflating variance measures. Acknowledging that not everything must be a prerequisite to making it in a scenario database, this incomplete coverage can hamper cross-model comparisons and obscure the true drivers of variation (Al Khourdajie et al., 2024).

Funded by
the European Union

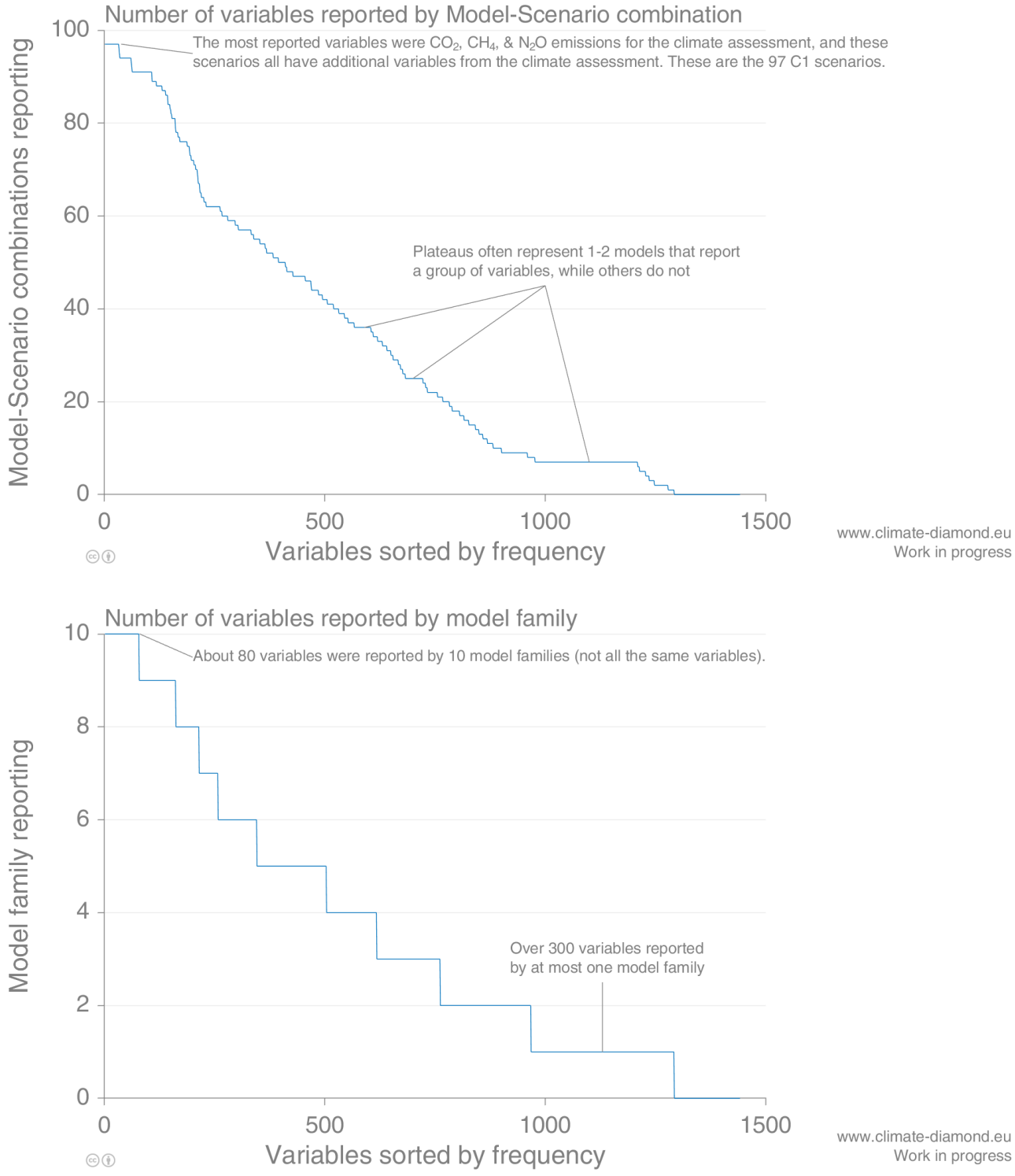


Figure 5. The distribution of variables reported in the scenarios database, in the C1 climate category, for submitted scenarios (top) and models (bottom). While this is shown for C1, the results are similar for other climate categories, with the distributions scaled up or down depending on the numbers of scenarios submitted.

There is clear value in having a scenario database to facilitate analysis (Huppmann, Rogelj, Kriegler, Krey, & Riahi, 2018). The alternative is that individual researchers would need to curate their own databases, duplicating work

and effort. The disadvantage with a scenario database is that it is then easy to perform analysis with the database, even if that analysis has limited statistical meaning. This puts greater pressure on the analyst to ensure they are assessing the relevant scenarios for the questions they wish to address. This could also create opportunities for practices that inadvertently produce biased results. An example is selectively filtering scenarios in a database to align with a specific narrative—for instance, removing all scenarios that include fossil fuels beyond the 2040s and subsequently analysing energy system characteristics in the latter half of the century(Green, Bois von Kursk, Muttitt, & Pye, 2024)). Such an approach risks biasing the analysis, potentially overlooking key trade-offs and synergies within the scenario space. Ultimately, the appropriateness of such practices depends on the specific research question being addressed, with assumptions needed to be stated clearly and transparently. Guivarch et al. (2022) also identified several methodology challenges of using these large ensembles of mitigation scenarios. Many of these issues are discussed below.

An important component of a scenario database curation is the addition of metadata. Some of the metadata should come by default, such as model version, scenario name, publication, contact person, etc. However, other metadata might need to be compiled, such as the model type, model solution method, policy assumptions, technoeconomic assumptions (e.g., technology costs, fossil fuel prices), whether climate impact is included or not, baseline scenario family, socioeconomic assumptions (e.g., SSP), etc. In addition, summary statistics may greatly add user access and duplication, such as variables like year reaching net zero, cumulative emissions, and climate variables (see below). There needs to be a straightforward method to combine different metadata to the scenario data and encourage analysts to publish and share their metadata (Peters et al., 2023b).

2.2 Scenario Selection and Vetting

Once a scenario database is compiled, a subset of scenarios will likely need to be selected to address the research question (Guivarch et al., 2022). In the IPCC AR6 (IPCC, 2022a), there was a heavy focus on providing a climate assessment and ensuring scenarios were consistent with recent emission trends, and so the vetting focused on selecting scenarios falling within a certain bound of historical emissions (Table 1); it is possible, however, to apply vetting criteria to both historical and future data. It is important that the vetting outcomes be added as metadata and not used to remove scenarios, as different research questions may require different vetting. As vetting is dependent on the research question, the type of vetting used depends heavily on the type of analysis to be performed. The IPCC AR6 vetting was clearly an attempt to ensure that scenarios were relatively consistent with historical data. However, depending on the question, different types of vetting may be needed or used. For example, if one wanted to compare how costs of mitigation change by limiting technology deployment in certain scenarios (D. McCollum et al., 2013), performing historical vetting as in AR6 may greatly restrict the scenarios considered with limited gains. If, on the other hand, one wanted to see where the world is heading under current policies (Sognnaes et al., 2021) then vetting would become more important. The importance of vetting could be greatly reduced if modellers strictly adhere to standard protocols and data for model and scenario calibration.

Table 1. Vetting criteria and ranges applied to the global scenarios submitted to the IPCC database for the climate assessment (IPCC, 2022a).

	Reference value	Range (IP range)	Pass	Fail	Not reported
Historical emissions (sources: EDGAR v6 IPCC and CEDS, 2019 values)					
CO_2 total (EIP + AFOLU) emissions	44,251 $MtCO_2\ yr^{-1}$	±40% (±20%)	1848	23	395
CO_2 EIP emissions	37,646 $MtCO_2\ yr^{-1}$	±20% (±10%)	2162	55	49
CH_4 emissions	379 $MtCH_4\ yr^{-1}$	±20% (±20%)	1651	139	476
CO_2 emissions EIP 2010–2020 % change	–	+0 to +50%	1742	74	450
CCS from energy 2020	–	0-250 (100) $MtCO_2\ yr^{-1}$	1624	77	565
Historical energy production (sources: IEA 2019; IRENA; BP; EMBERS; trends extrapolated to 2020)					
Primary energy (2020, IEA)	578 EJ	±20% (±10%)	1813	73	380
Electricity: nuclear (2020, IEA)	9.77 EJ	±30% (±20%)	1603	266	397
Electricity: solar and wind (2020. IEA, IRENA, BP, EMBERS).	8.51 EJ	±50% (±25%)	1459	377	430
Overall			**1686**	**580**	–
Future criteria (not used for exclusion in climate assessment but flagged to authors as potentially problematic)					
No net negative CO_2 emissions before 2030	CO_2 total in 2030 >0		1867	4	395
CCS from energy in 2030	<2000 $MtCO_2\ yr^{-1}$		1518	183	565
Electricity from nuclear in 2030	<20 $EJ\ yr^{-1}$		1595	274	397
CH_4 emissions in 2040	100–1000 $MtCH_4\ yr^{-1}$		1775	15	476

Rows do not sum to the same total of scenarios as not all scenarios reported all variables. EIP stands for energy and industrial process emissions.

The selection of variables for vetting is crucial depending on the purpose of the scenario(s) and the specific questions being addressed using the results. For example, if the goal is to evaluate the feasibility of the energy transition (e.g., the speed of fossil-fuel phase-out), then the scenarios should be vetted based on the relevant energy-related variables (e.g., coal, oil, gas) and how significantly they differ from historical data. In Table 1, the total primary energy variable is vetted for the year 2020 in the IPCC AR6, while other energy-related variables like primary energy coal, gas, or oil are not included in the vetting criteria. If historical data indicates that actual primary energy demand for fossil fuels has increased over the last five years but scenario results predict a drastic reduction in primary energy demand in the near term, the scenario could be considered unfeasible in the near term even though it may feature high long-term feasibility. For this example, if we include historical values for primary energy sources—coal, gas, and oil—in the vetting process for the year 2020, along with total primary energy, only 1,092 scenarios will be eligible for the climate assessment (see Figure 6). A total of 98 scenarios do not meet one or more of these additional vetting criteria.

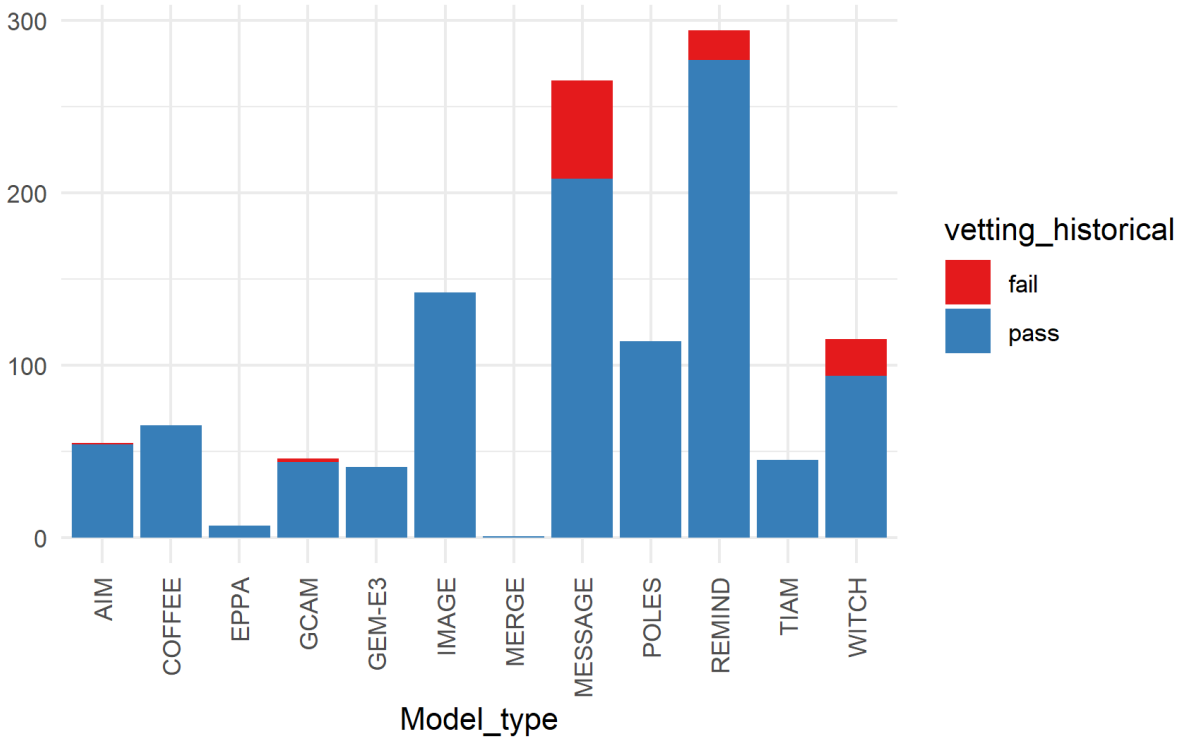
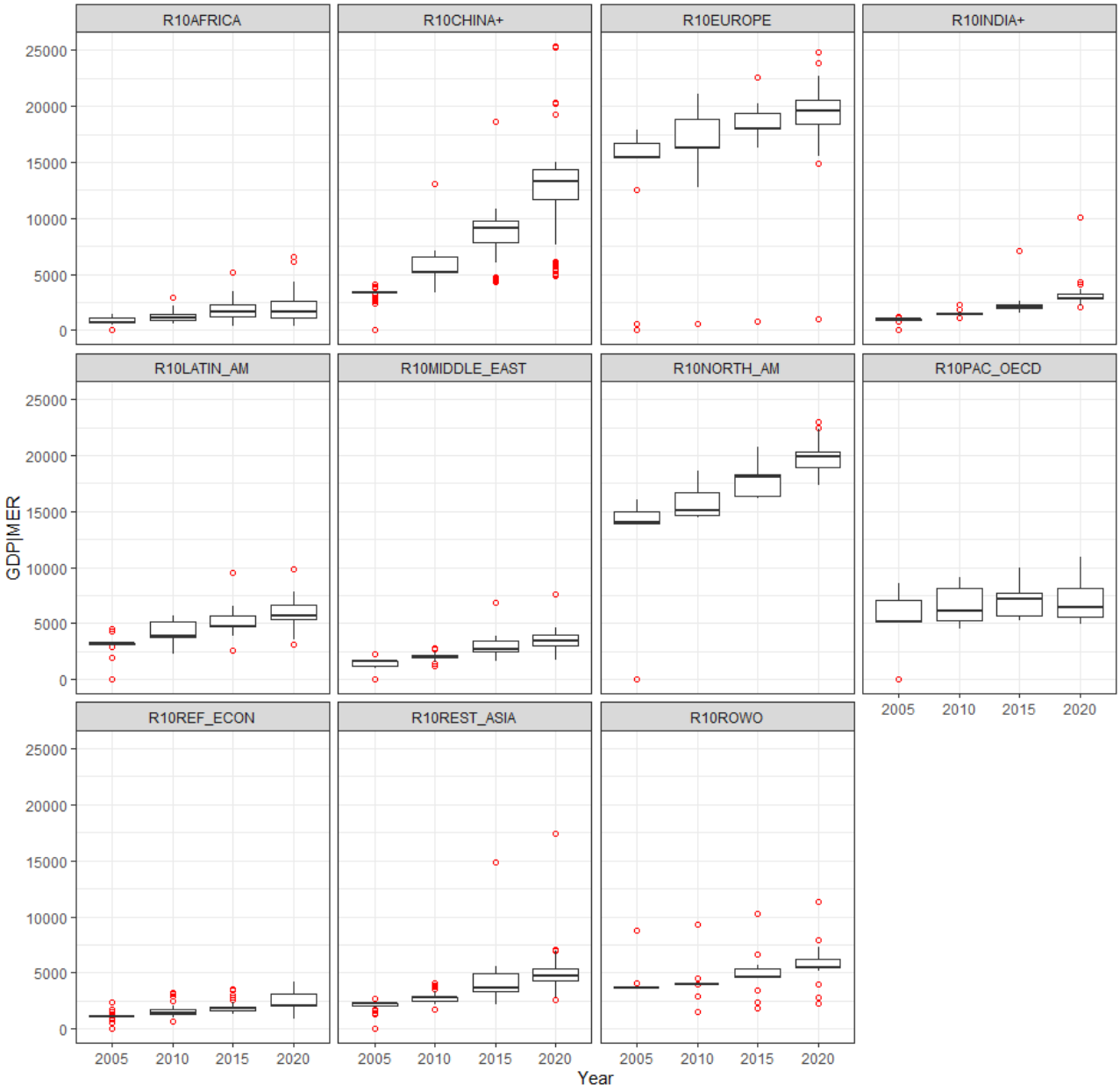


Figure 6. Distribution of scenarios in the AR6 IPCC database across model families after conducting historical vetting for primary energy sources: coal, gas, and oil for the year 2020. Additional vetting is performed on scenarios approved for climate assessment (n=1202).

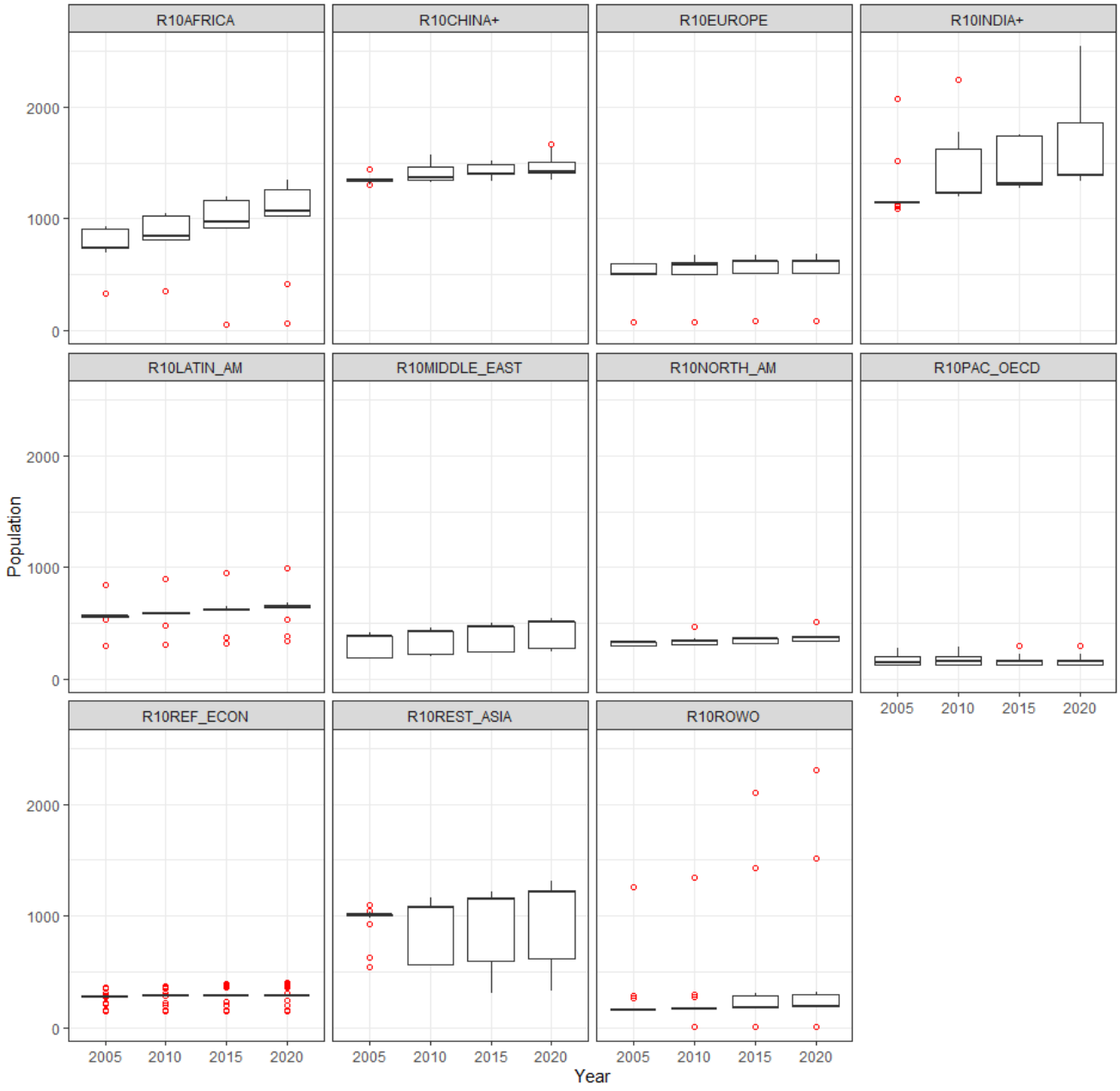
The vetting method is another important consideration. The IPCC AR6 vetting included strict pass or fail criteria for both historical and future data. As an example, a group of scenarios from the PARIS REINFORCE project (Sognnaes et al., 2021) had solar and wind in 2020 59% over the reference value, while the cut off for vetting was 50%. It may not be desirable to have such a strict pass or fail criteria when several dimensions are considered,but rather allow scenarios to break certain criteria if 'on balance' they cover all variables sufficiently well. For example, if a scenario fails marginally on one criterion but meets all others perfectly, is it worse than a scenario which marginally passed on all criteria? A clear concern in the lead up to IPCC AR6 was whether scenarios had consistent renewables growth, and removing scenarios that overperformed on this criterion may not be desirable. This emphasises that the purpose (motivating research question) of the vetting criteria needs to be clearly communicated, in addition to the choices that led to its implementation. In addition, the outcomes of the vetting criteria need to be clearly communicated back to the scenario producer, in case the vetting identified an error in the scenario or scenario reporting.

Vetting of scenarios in AR6 was only conducted at the global level and not at the regional or country level. As mentioned before, if one wanted to understand where the world is heading under current policies (Sognnaes et al., 2021) and to determine whether targets announced by the countries are adequate to meet a global temperature target (D.-J. van de Ven et al., 2023; Vandyck, Keramidas, Saveyn, Kitous, & Vrontisi, 2016) then it would be essential to conduct vetting at the regional level as well. Socioeconomic variables like population and GDP are key drivers used for estimating demand for end-use services in a model; however, these variables were not included as vetting criteria. Figure 7 illustrates several outliers at the regional level, even for the scenarios that passed the vetting process. While these outliers may be related to difficulties or inconsistencies with regional aggregation, it nevertheless highlights the challenges that arise and the difficulties with providing regionally consistent results from global models.

Funded by
the European Union


R10AFRICA
R10CHINA+
R10EUROPE
R10INDIA+
R10LATIN_AM
R10MIDDLE_EAST
R10NORTH_AM
R10PAC_OECD
R10REF_ECON
R10REST_ASIA
R10ROWO
GDPIMER
25000
20000
15000
10000
5000
0
2005
2010
2015
2020
Year

(a)



(b)

Figure 7. a) GDP (MER) and b) population at regional level (R10) for historical years

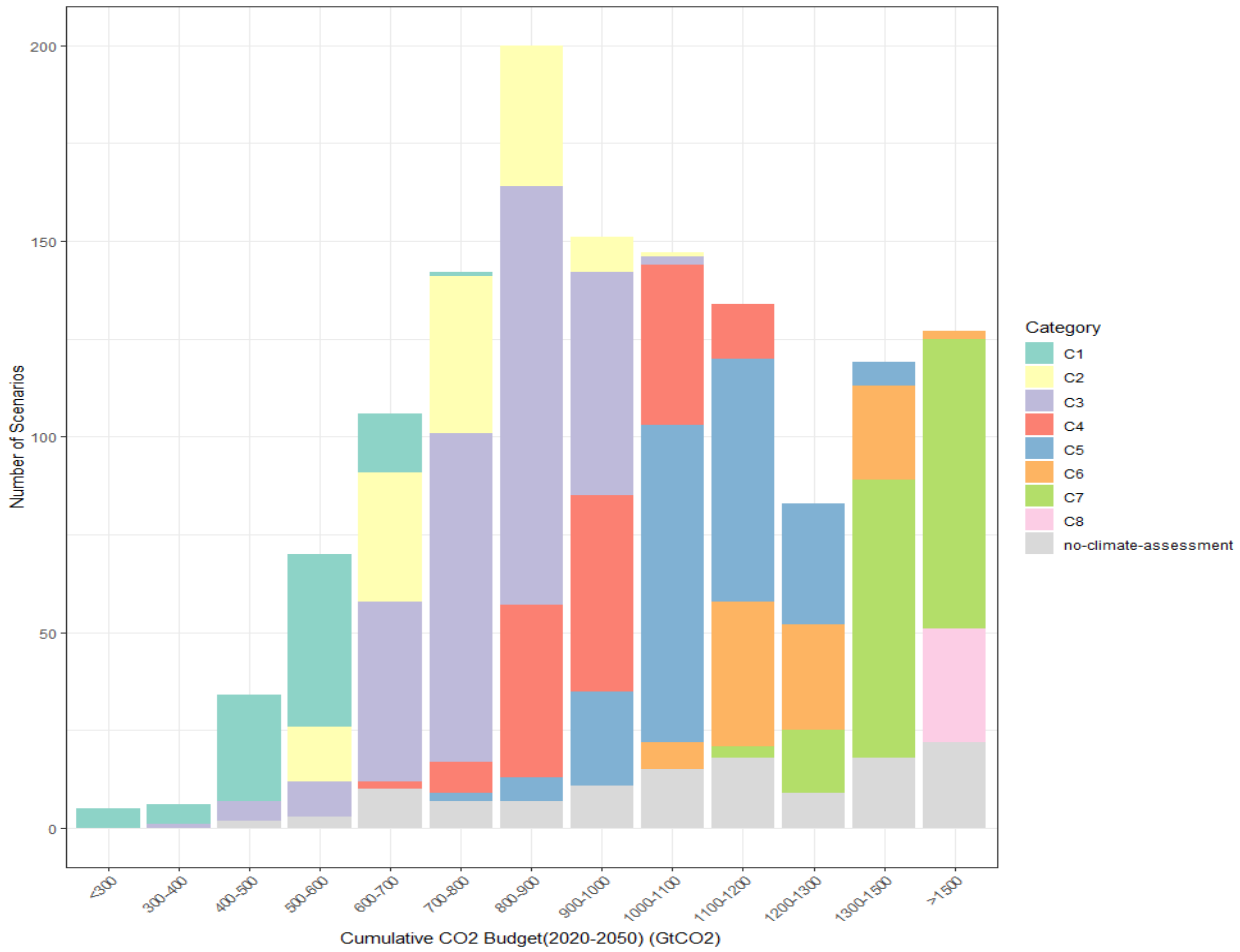
Vetting can also be unnecessarily restrictive. While temperature was not an explicit vetting criterion for IPCC AR6, it was an implicit vetting criterion. Of the 2,304 global scenarios submitted, 1,686 passed vetting, but only 1,202 got a climate assessment (Figure 2). Those that passed vetting but did not have the data for a climate assessment (CO_2, CH_4, N_2O) were implicitly not considered in the IPCC AR6: results were only shown for scenarios allocated in temperature based categories (K. Riahi et al., 2022). However, depending on the research question, it may not be necessary to have a climate assessment, as cumulative emissions may be sufficient. If cumulative emissions to 2050 had instead been used to group scenarios in AR6, instead of direct climate outcomes, then more scenarios could have been considered, potentially leading to different climate outcomes (Figure 8). Additionally, this restrictive vetting process resulted in 148 unique scenario narratives not being assessed, further narrowing the scenario space, and missing out on some potentially insightful trade-offs and synergies to unpack and learn from these scenarios (Table 2).

Funded by
the European Union

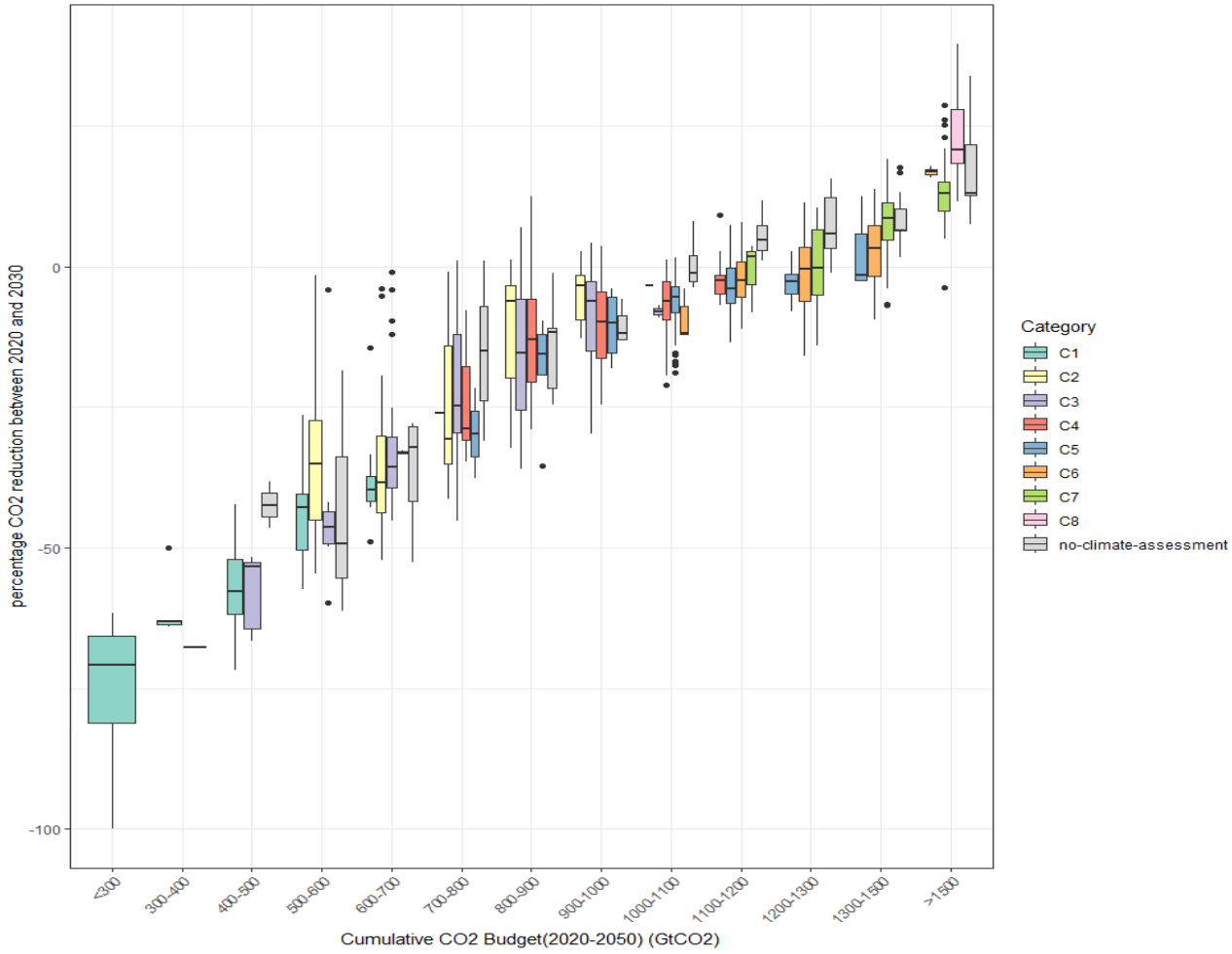
Table 2. Number of unique vetted scenarios in each category in the IPCC AR6 scenario database. A unique scenario here is defined as implementing a unique scenario definition (e.g., EN_NPi2020_400 in the AR6 scenarios database), while the bottom row shows the model-scenario combinations (multiple models run the same scenario definition in multi-model exercises).

Category	C1	C2	C3	C4	C5	C6	C7	C8	w/o climate-assessment
Number of unique scenarios	78	99	177	83	108	47	80	24	198
Number of unique scenarios without overlap	78	81	143	33	80	35	67	13	148
Overlap with C1		18	8	0	0	0	0	0	3
Overlap with C2			26	9	0	0	0	0	6
Overlap with C3				41	9	0	0	0	8
Overlap with C4					19	0	0	0	6
Overlap with C5						12	4	0	5
Overlap with C6							9	0	3
Overlap with C7								11	19
Overlap with C8									10
Total model-scenario pathways	97	133	311	159	212	97	164	29	483

Funded by the European Union


Number of Scenarios
200
150
100
50
0
<300
300-400
400-500
500-600
600-700
700-800
800-900
900-1000
1000-1100
1100-1200
1200-1300
1300-1500
>1500
Cumulative CO2 Budget(2020-2050) (GtCO2)
Category
C1
C2
C3
C4
C5
C6
C7
C8
no-climate-assessment

(a)



(b)

Figure 8. (a) Distribution of scenarios based on the cumulative CO_2 budget between 2020 and 2050 (b) range of % CO_2 emissions reduction between 2020 and 2030 within each cumulative CO_2 budget category.

A particularly useful type of metadata to add to a scenario database is model diagnostics (Dekker et al., 2023) . Model diagnostics could also act as a type of vetting data, depending on the research question. Model diagnostics may represent key model assumptions, model structure, technology assumptions, model behaviour, and so on. These variables may may help explain how results vary across models, such as responses to stylised carbon prices, just as climate modelling has stylised model diagnostic runs—known as the DECK (Eyring et al., 2016). These model diagnostics would need to be estimated from individual model runs but would be useful metadata to add for future analysis, particularly in model inter-comparisons.

Standardised diagnostic scenarios are assessed in various model inter-comparison studies to evaluate how sensitive each model is to different carbon price trajectories. Some modelling teams contribute the outputs from these diagnostic scenarios to a scenario database. Information about model behaviour can be reported in the metadata using key diagnostic indicators from these scenarios. Thus, the diagnostics are model-specific and not scenario-specific. To date, providing this information has not been mandatory in IPCC WGIII processes, unlike in the case of the CMIP database.

In a nutshell, it is ultimately the research question that should dictate the vetting criteria. In this context, vetting criteria should be added as metadata and not used to remove scenarios. There is a strong case to remove, or preferably correct (if a reporting error) or revisit (if a modelling error) scenarios that do not meet the quality control

requirements. But removing scenarios from a database based on vetting criteria excludes those studies from future studies based on alternative, and potentially equally valid, research questions.

2.3 A climate assessment pipeline

Climate assessment of a scenario refers to estimating key climate-related indicators and, in the case of a scenario pipeline, adding those indicators as new variables to the scenario in the database. The climate assessment may require certain emissions data, such as a minimum set of species and a consistent starting level for each of them, which in turn may require that further processing steps be added to the scenario pipeline before the climate assessment itself. Many models consider GHG emissions (CO_2, CH_4, N_2O, f-gases), but there are many other climate-related species that cause warming or cooling of the climate, such as sulphur dioxide (SO_2), black carbon (BC), organic carbon (OC), Volatile Organic Carbon (VOCs), Nitrogen Oxides (NO_x), Carbon Monoxide (CO), etc.

In the case of AR6, climate assessment was done by using simple climate models (SCMs) to add a relatively rich set of climate indicators to the scenario, including timeseries variables for percentiles of global surface temperature, radiative forcing, and probabilities of exceeding 1.5°C, 2.5°C, and 3.0°C of global temperature increase, as well as metadata for time and level of peak warming, median warming in the year of net-zero emissions, and total probability of exceeding key temperature thresholds. Each metric was computed with each of the three SCMs—FaIR, MAGICC, and CICERO-SCM—and added as separate variables or metadata items in the database, although most results in the IPCC report were based on the MAGICC estimates. To facilitate running the SCMs, the AR6 scenario pipeline required a harmonisation step to ensure a consistent starting level of emissions across scenarios, and an infilling step to infer emissions of species that were not included in some scenarios but required to run the SCMs consistently (Kikstra et al. 2022). Models that did not include CO_2, CH_4, and N_2O emissions were excluded from the climate assessment in the report to reduce uncertainty from infilling (see Section 2.5).

A comprehensive climate assessment like the one done in AR6 gives useful detail but may not be necessary in all contexts or for all research questions. It adds barriers to models and scenarios that do not include all the required emissions data for the required years and geographical scope or other details. Thus, many useful models and/or scenarios may not be assessed if a climate assessment is used but is not essential. In a climate context, it may be sufficient to provide the required data for estimating CO_2 emissions to be considered relevant, due to the strong relationship between temperature change and cumulative CO_2 emissions, through the transient climate response to cumulative emissions (TCRE). If required, this could be combined with an uncertainty range for the impact of non-CO_2 forcers taken from an existing database of scenarios that had sufficient detail to undergo a full climate model-based assessment (Sognnaes et al., 2021). Many scenario comparisons are in any case based on cumulative CO_2 emissions (Keywan Riahi et al., 2021), with a climate assessment in terms of temperature added later. Thus, the addition of the climate assessment is really assessing questions of how different models respond to non-CO_2 emissions. A deeper consideration of alternatives to a full climate assessment may diversify studies in terms of the range of models used and scenarios explored.

It is thus important to evaluate, for a given research question, what type of climate assessment is needed. In some cases a detailed and comprehensive assessment is needed (Kikstra et al., 2022), in other cases CO_2 emissions may suffice (Sognnaes et al., 2021) (Böhringer, Peterson, Rutherford, Schneider, & Winkler, 2021). The more comprehensive the climate assessment, the closer the scenario analysis may be to estimating climate outcomes, but also the more restrictive in terms of the models and scenarios that can be considered. Thus, there is an important trade-off between a robust climate assessment and the diversity of scenarios available for the assessment.

Funded by
the European Union

In the following three subsections, we discuss harmonisation of emissions data, infilling of emissions data, and the climate assessment, which were the three main steps of the full climate assessment pipeline in AR6.

2.4 Harmonisation

The term "harmonisation" in general refers to taking different things that disagree or are somehow inconsistent with one another and to some extent bringing them into agreement or making them more compatible. In the context of integrated assessment modelling and scenarios, the term has been used to refer both to harmonisation of model *inputs*/assumptions and to harmonisation of model *results* produced, using possibly non-harmonised inputs, by adjusting the results to agree with a given set of reference data. Only the latter is strictly a part of the scenario pipeline discussed here. It is noteworthy, however, that harmonising model inputs and assumptions can reduce the need for and the impact of both vetting and harmonisation of results, and thus the bias and uncertainty that may be added by the pipeline processing.

The scenario pipeline employed in the climate assessment framework of the WGIII contribution to the IPCC AR6 (Kikstra et al., 2022) only applied harmonisation to the emission variables in the submitted model results, since the outcome of the harmonisation was only used in the climate assessment (which was based only on emissions). However, harmonisation in general is significantly broader and may need to be applied to other types of variables depending on the analyses that the results will be used for, including socioeconomic data (e.g., population, GDP), costs, energy, and other variables, in addition to emissions. Indeed, it may be appropriate to harmonise some types of modelling assumptions like GDP, population, although this is usually only possible at the modelling stage and rarely by post-processing results in the scenario pipeline.

Harmonising only emissions data from different scenarios may be sufficient in contexts where only climate outcomes are being compared among models, and the drivers behind different emissions levels and resulting climate outcomes are not being investigated. However, in studies that make broader comparisons between scenarios or seek to explain differences between scenarios or models, a broader set of quantities may need to be harmonised to be able to isolate the effects of differences in model behaviour, differences in scenario design, and simply differences in input or calibration data.

In the subsequent sections, we consider different types of harmonisation, focusing on the scenarios assessed by WGIII for the IPCC AR6.

2.4.1 Harmonisation of inputs versus outputs

As discussed above, "harmonisation" in the context of scenarios may be used to refer to different things, without an explicit statement of the differences between the different uses. The ones that are relevant to the discussion here are:

1. **Harmonisation of inputs and assumptions (Type 1).** This refers to establishing a common reference set of input/calibration data and modelling assumptions and ensuring that models adhere to these reference data and assumptions. In some cases, there is some flexibility to perform a minimum check that specific data and assumptions used by a given model or the endogenous results that it produces are *sufficiently* consistent with (i.e., within an acceptable range of) the reference data. This is typically done in the context of a multi-model project, or by establishing a prominent reference data set that several projects elect to use, such as the Shared Socioeconomic Pathways (SSPs). Giarola et al. (2021) provided a much-cited example of this, from the EU Horizon 2020 project PARIS REINFORCE, and systematically investigated the individual impact on model results of successive, increasingly comprehensive harmonisation steps. That work was cited in the discussion of this meaning of "harmonisation" in Annex III of the WGIII contribution

to AR6 (IPCC, 2022a).

2. **Harmonisation of model results through post-processing (Type 2).** This refers to modifying parts of the model results, after the models have been run, to make them conform to a set of reference data—typically a set of historical values for a subset of variables. This is, in some sense, a way to compensate for the lack of Type 1 harmonisation when performing post-analyses on model results from different studies, which may have used diverse, or even partially conflicting, input data and assumptions and/or may have been performed at different times (meaning that the different studies had had access to different sets of historical statistics). In the WGIII contribution to the IPCC AR6, this type of post-processing harmonisation was applied specifically to emission variables in the model results submitted to the WGIII scenario database, by scaling emissions of each species for each region to match a selected dataset for 2015, and then using a methodology that combined scaling and additive offsets (depending on species and the characteristics of the emission time series per scenario and region) to adjust the full emission time series accordingly. The harmonisation methodology was described by Gidden et al. (2018), and its application in the AR6 scenario pipeline was reported in Kikstra et al. (2022), and its use in the ScenarioMIP experiment in Coupled Model Intercomparison Project 6(CMIP6) was described in Gidden et al. (2019).

Of the two, Type 1 harmonisation seems to be the one to have appeared first in an IPCC context. It was used at least as early as in the context of the IPCC Special Report on Emission Scenarios, or "SRES", in 2000 (IPCC, 2000), where it is listed in the glossary as the only meaning of the term. The first appearance of Type 2 harmonisation in an IPCC report appears to be in AR6, in connection with the scenario pipeline, citing and building on the then-recent publications by Gidden et al. (2018); Gidden et al. (2019).

In addition, the word "harmonisation" occurs in different, not directly related contexts in the AR6 WGIII report and previous IPCC-related publications. This includes (but is not limited to) coordination of policies and climate actions between governments, emission accounting conventions, and numerous types of data not directly related to the scenario pipeline described here. These other uses of the term will not be discussed further.

Only Type 2 harmonisation is part of the scenario pipeline itself. Type 1 harmonisation, by definition, must be done during the design of a modelling study, before the scenarios are implemented and run by models, and therefore before the start of the scenario pipeline described in this section. However, the two types share connections, and it is therefore useful to discuss both here. In particular, the need for harmonisation type 2 can be greatly reduced if the modelled scenario results that are included in a database have been predominantly produced with data and assumptions harmonised according to Type 1.

Furthermore, although any quantitative variable in principle can be harmonised according to Type 2 in a scenario pipeline, the model-specific correlations between variables will not necessarily be preserved. It is therefore best applied only to a single type of variable that is used in further processing or analysis (e.g., emissions variables for climate assessment, in the case of AR6). Harmonisation of Type 1 is usually required if it is necessary to consistently harmonise several different types of variables, or to align qualitative scenario assumptions. It is therefore needed in analyses that compare models or scenarios across a range of different types of variables, to ensure that any observed differences across models and scenarios can be attributed to the models and scenarios themselves, rather than to inputs and assumptions.

As for how to perform harmonisation, S. Giarola et al. (2021) discuss potential methods for harmonisation according to Type 1, providing a stepwise example from the PARIS REINFORCE project. For Type 2, many different procedures could in principle be used. The most relevant here is the approach based on the software package ANERIS as described in Gidden et al. (2018). That approach was applied to emission variables in the scenario pipeline for the climate assessment in the IPCC AR6 WGIII scenario database and is discussed under harmonisation

of emissions in Section 2.4.4.

One could argue for a general preference to Type 1 harmonisation, to largely avoid the need for Type 2 harmonisation. Type 2 harmonisation modifies a subset of output variables of an IAM (e.g., emissions) in a post-processing step. Therefore, any model-specific correlations with other variables (such as energy demand or technology deployment) are disrupted. This can alter system-level trade-offs that were originally embedded in the internally consistent model run, potentially biasing analyses that rely on the relationships among multiple variables.

2.4.2 What to harmonise, and past vs. future harmonisation

The main purpose of any type of harmonisation in the context of model inter-comparisons and scenario ensembles is usually to improve comparability between model results. More specifically, to ensure as far as possible that differences between model results and scenarios are caused by real differences in model dynamics or scenario narratives, and not because of inconsistencies in input data or assumptions that should logically be the same across scenarios. At the same time, harmonisation must take care *not* to remove differences between results that reflect the kind of model diversity or scenario uncertainty that the scenario ensemble is supposed to capture (e.g., by customising the input data or scenario implementations ad hoc in each IAM owing to a priori known behaviours of specific IAMs). It is meant to flesh out diversity among IAMs (in terms of structure, theory, solution approach, etc.) rather than hide it.

What to harmonise thus depends on what the model results will be used for. It may be appropriate to harmonise some variables and not others, and it may be appropriate to harmonise the variable values for only a certain range of *years* and not others.

Historical values of any variable should logically be the same across all scenarios (other than possibly counterfactual ones). In principle, therefore, values of all variables up to some historical cutoff year (such as the last year of available statistics) would be candidates for harmonisation. By the same logic, *future* values of exogenous (input) variables should be harmonised if they should explicitly be the same across scenarios (i.e., control of an experiment), but not otherwise, or the harmonisation would most likely remove model and scenario variability that would be relevant for comparison studies.

Which variables could be harmonised on future values depends on the scenario narratives. Global (and often national) population is frequently a candidate, since many climate scenario narratives do not differ in ways that should affect global population numbers (but might affect migration). Harmonising future GDP trajectories might also be appropriate, if the scenario narratives do not imply that they should be different or are not detailed enough to infer how GDP should evolve differently in different scenarios. Furthermore, technology cost and performance parameters, energy prices, and other cost parameters may require harmonisation, if the scenario narratives do not explicitly imply that they should vary, or if they are not modelled endogenously by any models. These parameters are especially relevant because they often greatly affect outcomes and so might create a great deal of non-transparent variance between models if different models use different assumptions(Krey et al., 2019)).

Nevertheless, which variables it makes sense to harmonise in practice depends both on whether it is model inputs or outputs that are being harmonised (Type 1 vs. Type 2 in Section 2.4.1), and not least what is practically feasible.

When harmonising model inputs and assumptions (Type 1), it can make sense to harmonise any type of variable, subject to the discussion above. But, if harmonisation is only done by adjusting model *outputs* as part of a scenario pipeline, it is only useful and sensible to harmonise variables that will be used directly in subsequent analysis, and only a single type of variable or sets of variables that do not correlate significantly with one another. How changing one variable affects a different variable is generally not a priori known without rerunning each model, so

harmonisation of Type 2 can only change the harmonised variables themselves. For the same reason, harmonisation of Type 2 typically introduces additional uncertainties and possible biases, since it is usually not possible to guarantee that the harmonised results are consistent with the original model assumptions or with the non-harmonised variables. The harmonisation in the IPCC AR6 climate assessment pipeline was applied specifically to emissions variables since they would subsequently be used by simple climate models to assess the climate outcomes of each scenario. For that purpose, other variables were not relevant. Nevertheless, the procedure introduced some uncertainty, given that it was not possible to know how consistent the harmonised emission trajectories had been with the original model assumptions. The harmonisation could therefore potentially have biased the statistics of the climate outcomes in the final assessment.

Even when harmonising inputs (Type 1), full harmonisation may not be feasible for many variables because of model design constraints. For example, if models use very different technology categories with no easy mapping between them, it may not be feasible to fully harmonise technology costs and parameters, or even determine whether parameters used in the different models are mutually consistent or not. Or, in the case of GDP or energy consumption or emissions, the variables may be modelled endogenously by some models and be exogenous inputs in other models. Harmonisation may thus require some models to calibrate to the harmonised variable values by adjusting the values of other, exogenous input data or parameters, which may be too time consuming and not always feasible with available resources. The only solution in such cases may be to check consistency and accept (and transparently report) differences within a certain level.

In the following subsections, we briefly consider how and when different types of variables can be harmonised.

2.4.3 Socioeconomic data (population and GDP)

When model inputs can be harmonised (Type 1, in Section 2.4.1), historical values of socioeconomic variables should almost always be harmonised as far as possible, given their potentially high impact on emissions, energy use, and other variables that are commonly of interest in analyses. Future values (projections) are frequently also harmonised, given that many scenario studies explore topics for which demographics and/or GDP are background variables and not themselves under study, in which case the studies need a suitable default value for socioeconomic variables. The Shared Socioeconomic Pathways (SSPs) were developed for and are frequently used to provide future values. Though, how well the SSPs are incorporated in the models, particularly at a regional level, still needs assessment (Figure 7).

As mentioned in 2.4.2, when not harmonising inputs but instead adjusting outputs (Type 2 harmonisation), it will usually not make sense to harmonise socioeconomic variable values. In climate scenarios, these variables are important mainly because of their effects on other variables of interest, but that effect is of course not captured by post-adjusting the socioeconomic variables. If harmonised socioeconomic assumptions are required and this has not been done at the model calibration and input data preparation stage, the only option will typically be to impose vetting requirements on the socioeconomic variables, and either exclude scenarios that do not pass the vetting or require models to rerun the scenarios with harmonised assumptions.

Socioeconomic variables were not harmonised in the AR6 scenario pipeline. As shown in Figure 7, this can be a significant source of uncertainty and unintended variance, especially at the regional level. Emissions in the AR6 climate assessment were harmonised to a common set of historical reference values. But, if the variation in historical emissions between models were caused by differences in historical population or GDP figures, which also would affect many other variables, the Type 2 harmonisation may have made emissions highly inconsistent with other variables in the database.

2.4.4 Emissions (for input to climate assessments)

Emissions variables are the main determinant of climate outcomes in climate scenarios. Harmonising historical emissions is therefore crucial to ensure a consistent starting point for climate assessment. It is therefore one of the most common quantities to harmonise, both when harmonising model inputs before running models, and when harmonising model outputs as part of an assessment pipeline. It was the only type of variable that was included in the harmonisation step in the IPCC AR6 WGIII climate assessment pipeline, since the purpose of that pipeline was to perform SCM-based climate assessment and process the emissions data in the scenarios to be sufficient and suitable for running the SCMs.

So, the harmonisation applied to modelled emission outcomes in AR6 (Type 2, see Section 2.4.1) was necessary to ensure that variability in climate outcomes between different scenarios as far as possible was due to actual differences in the emission trajectories of each scenario, with as little contribution as possible from inconsistent starting points, i.e., from inconsistencies in assumed historical emissions. But that type of harmonisation necessarily requires a number of discretionary assumptions, which introduce a hidden amount of uncertainty in the final climate outcomes produced in the climate assessment.

The emissions harmonisation in AR6 was done using a methodology implemented by the software package ANERIS, which aligns modelled emissions with historical emissions data by applying ratio-based scaling and/or additive offsets that ensure the emissions after the harmonisation match historical reference data in a base year, and optionally reduces the scaling and/or offsets over time so that the harmonised emissions match the original modelled emissions by a future convergence year. This procedure introduces four types of decisions that must be made to fully specify the harmonisation for each combination of variable and scenario:

1. **Reference emissions data:** What set of historical emission figures the scenarios are supposed to match after harmonisation
2. **Base year:** In which year the scenarios are made to match the reference data
3. **Adjustment type:** Whether scenario results are adjusted to match the reference data in the base year by scaling with the ratio (multiplicative) or offsetting with the difference (additive) between the two
4. **Convergence:** Whether and how the adjustment is reduced over time, including in which year the harmonised variable values again match the original scenario values (the convergence year), and how the scaling ratio or offset is reduced between the base year and convergence year

The choice of each of the above can be justified by certain reasonable criteria; however, both these criteria and in turn these decisions could in principle be different, meaning that these four decisions introduce uncertainty in the form of arbitrariness that is not quantified in the scenario database itself. In the following 4 subsections, we outline what choices were made in AR6 and discuss what room there is for making different choices, and how significant the impact of doing so can be expected to be. The descriptions of what choices were made in the AR6 WGIII harmonisation (Type 2) process is based on Kikstra et al. (2022)—see there for more details.

2.4.4.1 Reference emissions data

AR6 used the same sources for historical emissions data as in CMIP6 and Reduced Complexity Model Intercomparison Project (RCMIP). A large share of the data came from the Community Emissions Data System (CEDS), which in turn uses a number of other sources. The details are described in Gidden et al. (2019).

The effect of choosing a different emissions dataset while keeping other choices constant is potentially straightforward: each emission variable in the reference year will either get scaled by the ratio between the new

and old reference data, or shifted by the difference between them, depending on whether the variable is harmonised through scaling by ratios or offsetting by differences. The effect of harmonisation on an emission variable is therefore proportional to the ratio or the difference between the native scenario value and the reference data value for that variable in the reference year (and gets smaller over time if convergence is used, see above). This further means that the possible variability in the impact of harmonisation that is associated with different choices of harmonisation data, is simply proportional to the spread in emission values between the different possible harmonisation data sets.

One complication is that, in AR6, the choice of the adjustment method (scaling versus offset, and convergence) was done using a decision tree that depends on the gap between model and reference variable values in the base year (see Sections 2.4.4.4 and 2.4.4.5, and references therein). Changing reference data can therefore result in a different adjustment method being used for some variables and scenarios, which means that it can have a less straightforward effect than described in the previous paragraph for individual variables and scenarios.

The uncertainty associated with choosing reference data is not insignificant. Even for fossil CO_2 emissions, estimates of recent historical emissions can vary by 5-10% (Andrew, 2020), and substantially more for some countries or when disaggregating by sectors. For land-use change and non-CO_2 emissions, the effect can be substantially larger, in excess of 50%.

It is unknown how much the reference data choices used for harmonisation would affect the temperature outcomes of the AR6 database pipeline. Given that the variance in unharmonised emissions in the original model results are likely to reflect to some extent the variance in the available emissions data, the most likely magnitude of the effect may be seen already by looking at the effect of the existing harmonisation in the AR6 scenario pipeline. The original article by Kikstra et al. (2022) shows how large this effect is for different emission types, and how the effect differs by climate scenario categories (see Figure 9).

Funded by
the European Union

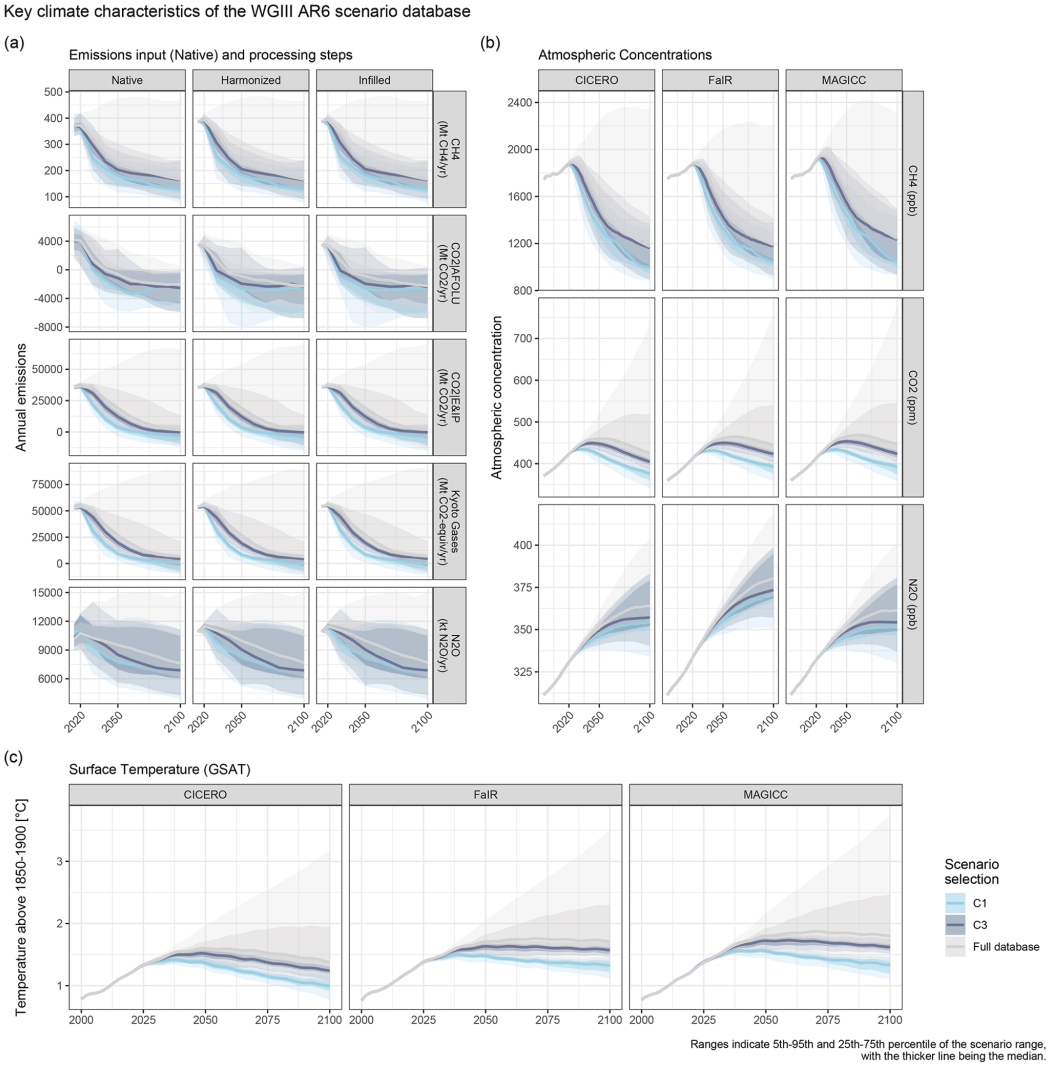


Figure 9. Reproduction of Figure 2 from Kikstra et al. (2022), showing the effect of harmonisation and infilling on emission timeseries and resulting climate characteristics of scenarios in the AR6 Working Group III database. Original caption: Summary statistics (a: emissions, b: atmospheric concentrations, c: global mean surface temperature) over time across all scenarios in the AR6DB that received a temperature classification and across scenarios in AR6 temperature categories C1 and C3. Panel (a) shows emissions as modelled by IAMs ("Native"), after harmonisation ("Harmonized"), and after infilling missing reported emissions ("Infilled"). Panel (b) and panel (c) show climate outcomes per climate model, using the median value of each variable from the climate emulator probabilistic distributions.

2.4.4.2 Base year

In AR6, emissions were made to match the reference data in 2015. For models that only had reported emissions in 2010 and 2020, a combination of those were used as proxy for its 2015 emissions. The choice of base year is somewhat arbitrary and, in principle, any year for which historical data exists could be used. However, if the base year is a future, modelled year and not based on historical data for some models, it might result in a very large adjustment for those models, which might mean that the overall adjusted results do not reflect well the original model results.

Choosing a different base year can change the size of the adjustment for a given variable and scenario, if the difference between the model-native and reference-data variable value is different in the new base year and can, therefore, vary considerably by variable and scenario. Changing the base year could also result in a different adjustment method for some variable/scenario combinations, in the same way as when changing reference data.

The effect of changing the base year is therefore not straightforward to predict.

2.4.4.3 Adjustment type

The harmonisation methodology in ANERIS used in AR6 adjusts the scenario emissions values to match the reference data in the base year in one of two ways:

1. Scaling: The emissions are scaled by multiplying with the ratio between the reference and scenario values
2. Offset: The emissions are shifted by adding the difference between the reference and scenario values

In both cases, the size of the scaling or offset for a given variable is calculated using only the scenario and reference values in the single base year. The scaling or shift is applied to the entire timeseries for that variable, but the size of the adjustment may be reduced over time and go to zero in a defined convergence year (see "Convergence").

The difference between the two adjustment types is small for variables whose values change little over time, but correspondingly big for variables that do change considerably. Although the choice is in principle arbitrary, one or the other adjustment type can be more or less appropriate depending on the variable values. For example, scaling is problematic for variables where the scenario value is very small relative to the reference data, or if either value is zero. Similarly, additive offsets are not likely to be appropriate for variables where the scale of the variable values change monotonously and significantly over time, such as in the case of most CO_2 emission variables in mitigation scenarios or conversely in high-growth/low-mitigation scenarios.

Gidden et al. (2018) and AR6 (Kikstra et al., 2022) use scaling in most cases, except when scenario values are zero in the base year. AR6 also used an offset for AFOLU CO_2 emissions to deal with both high interannual variability and high uncertainty in the existing historical data, which resulted in rather large shifts in negative emissions in mid-century for high-mitigation scenarios. The default decision tree for choosing adjustment type and convergence (described in the next subsection) is given in Figure 1 of Gidden et al. (2018), and custom choices made in AR6 are described in Table 1 of Kikstra et al. (2022).

It is not clear which adjustment types are preferable, and choices can appear arbitrary. The choices were carefully considered in both Gidden et al. (2018 and 2019) and in the AR6 climate assessment, so there is no a priori reason to believe that making different choices would yield results that are more appropriate in some way.

2.4.4.4 Convergence

In AR6 and in the default decision tree introduced in Gidden et al. (2018), the scaling ratio or offset for most variables and scenarios was linearly reduced from the base year to a convergence year, in which the adjustment reaches zero so that the harmonised variable values again match the original scenario values. The default convergence year was chosen to be 2080, although 2100 was used in cases of large discrepancies between scenario values and reference (greater than a factor of 2), and 2150 (past the final model year) for variables with high interannual variability in the historical data. Furthermore, the adjustment was *not* reduced over time (i.e., was fixed) for variables that go negative during the scenario period, since doing so likely would result in the harmonisation flipping the sign of the variable for some years, which was seen as undesirable. Finally, the adjustment was also left constant for variables where confidence in the reported model values was seen as being low, which mainly applied to F-gases. Details of the choices are specified in Figure 1 of Gidden et al. (2018) and Table 1 of Kikstra et al. (2022).

The choice of convergence year and whether to let the harmonised results converge to the original scenario results at all is seemingly one of the most complicated and, in some sense, most arbitrary of the harmonisation choices. It can also have a very significant effect on the result, since the convergence procedure in general will change the

shape of the original timeseries, the apparent relative mitigation speed of different GHG species and different sectors and regions, and important parameters of interest like zero-crossing year or whether or not the variable changes sign at all, unless done carefully. Gidden et al. (2018) contains a discussion of several issues that may arise, with illustrative examples.

The choice of convergence methods made in AR6, based on Gidden et al. (2018), was made with careful consideration and involvement of the IAM community. Nevertheless, the choices made can clearly be questioned and could have been made differently. The main motivation for letting harmonised emissions converge back towards the original modelled results towards the end of the scenario period is to ensure that the statistics of the final-year results reflect the structure of the original model results and don't bear a heavier imprint from the choice of reference data, and can be justified by assuming that the model converges towards a state that is determined more by the conditions at the end of the scenario period and the structure of the model itself, rather than the exact starting point. But this will not necessarily be true in models or for variables with high path dependence, or where the trajectory scales more or less proportionally with the starting value. As an extreme but simple example, for a scenario where a model simply reduces the emissions of a certain species by 90% from the start year to the final year, but is harmonised to a reference data set that is 45% higher than the scenario value in the base year, the most reasonable harmonisation choice might well be to let the emission trajectory be scaled up by 45% for the entire time period, and thus preserve the original reduction ratio and relative reduction speed. Letting the harmonised values converge to the original by the final year would result in a final value that is 31% lower and an emission trajectory that goes down 31% faster than the original model logic would suggest.

Given these potential issues with convergence and convergence-related choices, it is important to investigate what the impact would be of different reasonable choices for convergence, and in particular of not letting harmonised variables converge to the original results at all (i.e., setting a convergence year in the infinite future), and of letting all variables converge in the same year.

2.4.4.5 Further comments on effect of harmonisation

As shown in Kikstra et al. (2022), the total impact of harmonisation on global outcomes in the AR6 WGIII scenario database is rather limited, but this is to some extent due to the dominance of fossil CO_2 emissions, for which the mean and median absolute adjustment sizes are just 2.9% and 2.3% or the original values, respectively. For AFOLU CO_2 emissions and non-CO_2 species, the effect can be much larger. Even if the collective effect of harmonisation across a group of scenarios is modest (e.g., climate category C1), there can be much bigger effects for individual scenarios.

The effect of harmonisation on AFOLU CO_2 emissions for C1 scenarios (the highest mitigation category) in AR6 has some particularly striking and potentially problematic characteristics. Because of the high uncertainty and high interannual variability in the historical data, an offset was used rather than scaling, and the adjustment was made to decrease very slowly with time (convergence year 2150). Because some scenarios originally had much higher AFOLU emissions than the reference data in 2015, and many of them exhibit large negative AFOLU emissions in mid-century, the offset had the effect of substantially increasing the magnitude of negative emissions for the lower range of the scenarios, with 5th-percentile AFOLU emissions in 2050 going from just under −6 Gt CO_2 to almost −8 Gt CO_2. The median on the other hand had the opposite effect, with peak negative emissions moving somewhat closer to zero, and the overall shape of the AFOLU CO_2 emissions timeseries becoming flatter. Given the attention on negative emissions and afforestation as ways of reaching net-zero or net-negative emissions, these kinds of changes may have some significance in policy-relevant debates.

The analysis in Kikstra et al. (2022) did not consider whether the effects of harmonisation could have moved some scenarios from one climate category to another, since climate assessment in AR6 was only done after

harmonisation. However, it is not unreasonable to think that that may have happened. In particular, older mitigation scenarios that assumed strong global climate policies starting before 2015 would most likely have lower CO2 emissions in 2015 than the reference data and would not have been meant to match real-world emissions in 2015. The harmonisation procedure would have resulted in the entire CO_2 timeseries from the base year to the convergence year being adjusted upwards. Due to convergence, the harmonised scenario would still be a strong mitigation scenario, with even steeper emission reductions than the original, but both cumulative and peak CO_2 emissions would be higher, which could result in a different climate categorisation than what would normally be associated with the original emission levels. The number of scenarios for which this happens cannot be computed using the AR6 climate assessment pipeline itself because of the requirement to harmonise and infill emissions but could be approximated by an alternative assessment where climate categorisations are estimated by some other method than running SCMs. This could include using TCRE with appropriate error bars to estimate warming levels based on cumulative CO_2 emissions and comparing the peak or final temperatures with and without harmonised CO_2 emissions. Another possible method would be to simply run the SCMs on each scenario without infilling emissions, but with and without harmonisation, and then comparing the results. The absolute temperatures in this assessment would not be accurate, and scenarios could not be compared to one another because of differences in which species they include. But the difference in temperature outcomes with and without harmonisation for a given scenario could still be sufficient to assess whether or not the harmonisation process changes the temperature outcome or even may have changed the climate categorisation of that scenario. Further, some models may be able to report native climate assessments, based on different climate assessment pipelines.

Although we have identified a number of potential issues related to choices that must be made in a harmonisation procedure, it should be stressed that the decision tree in Gidden et al. (2018) and the version of that applied in AR6 were the result of lengthy and careful consideration. It is therefore not given, or even likely, that other choices would be more appropriate, at least not in general. Our aim here has nonetheless been to enumerate what the issues can be, with a somewhat different angle than that used in the previous publications. It is also worth exploring what the effect of different choices might in fact be, again to a greater extent than has been done in the existing literature.

We plan to conduct some experimental analyses to assess uncertainties in the harmonisation process and report results in D5.2.

2.4.5 Energy

Energy consumption per fuel source is perhaps the most important and directly relevant category of variables for determining greenhouse gas emissions, and thus of particular importance to climate scenario databases. To ensure a consistent starting point, it is therefore usually highly desirable to harmonise historical energy consumption, if possible, by both sector and energy type as well as regionally. Harmonising future trajectories is less commonly useful, given that energy consumption is usually modelled endogenously by models that represent it, and modelling energy consumption and resulting emissions is often a major part of climate-related modelling studies.

Even when harmonising model outputs (Type 2, see 2.4.1), energy consumption is somewhat unique in that it can often be harmonised consistently together with emissions, provided that emission variables are disaggregated by energy type, in particular if they are disaggregated in the same way as the energy variables also along other dimensions. In most cases, one can simply scale or shift the energy variables in a way that preserves the ratio between emissions and energy consumption for each energy type, and the result will generally be consistent or close to consistent with the original model dynamics. If the emissions and energy variables use different disaggregation by sector or region, this procedure will introduce some uncertainty, if the model uses region- or sector-specific emission factors, but may still be accurate enough to be useful in some contexts.

The climate assessment pipeline in the AR6 WGIII database did not harmonise energy consumption along with emissions. It, therefore, did not introduce additional uncertainty in the energy-related variables in the database, but care must be taken to use non-harmonised emissions when analysing any quantity that depends on both emissions and energy consumption, such as emission intensities. As mentioned in Table 1, the vetting stage in the AR6 pipeline required primary energy consumption and electricity production from nuclear, solar, and wind in 2020 to fall within certain bounds, and somewhat more controversially also for future values (in 2030). Such vetting criteria potentially reduce uncertainties related to discrepancies in model assumptions about energy consumption, but they may also lead to biases.

2.4.6 Costs, performance, prices, policies, other parameters, and model assumptions

Many other parameters may have a high impact on modelling outcomes and would ideally be harmonised to the extent that they should not vary by scenario. This includes, in particular, variables and assumptions related to costs and performance, such as technology costs and performance parameters, energy and commodity prices, policies, and elasticities or other demand-related parameters. For models that represent land use change, data and assumptions about land areas and their usage will also be of key importance. Most such parameters are important for climate scenarios, because of their potential high impact on emissions and energy consumption but are not processed further in later steps in a typical scenario pipeline. They are therefore only suitable for harmonisation in model inputs (i.e., Type 1 harmonisation, see 2.4.1).

However, these parameters tend to be highly model-specific, in terms of representation to begin with, granularity and boundaries between them (e.g., technology types, commodities/material types, sectors, etc.), and whether they are required as exogenous inputs or modelled endogenously. Harmonising them and even just selecting a suitable reference data set can be challenging. The process will frequently require at least some modelling teams to tune their models and compare aggregated variables to aggregated harmonisation values, which can both be highly time-consuming and introduce both uncertainties due to mismatches in categories and room for errors. For similar reasons, even performing vetting on these parameters would in most cases be very challenging, and most likely result in many models not being assessed owing to their lack of representation of the variables that are selected for vetting.

No cost, performance, or price variables or policies were harmonised in the AR6 scenario pipeline. Variance in these parameters, especially future trajectories, can to some extent be considered part of desirable model and scenario diversity. But discrepancies in *historical* costs, prices, and performance parameters—which are not necessarily transparent in the database—can potentially be a significant source of unexplained variance in the model results.

2.5 Infilling

Not all models report emissions for all the species that are required for the climate assessment. For the models that did not report the necessary variables, and require a climate assessment, then the missing data needs to be estimated. Infilling refers to the process of inferring reasonable values for the missing data based on other data points that the model has reported. This process inevitably involves adding information that does not actually exist in the models, and so necessarily adds uncertainty as well a significant amount of arbitrariness associated with the choice of method and assumptions used to perform the infilling.

To infill emissions of a missing species *A*—i.e., the "follower" variable (Lamboll, Nicholls, Kikstra, Meinshausen, & Rogelj, 2020) for a model that does not include it—one typically infers a correlation between emissions of species *A* and emissions of a different species *B*—the "lead" variable that *is* reported by the model—in a set of scenarios

Funded by
the European Union

produced by different models that *do* include both species *A* and *B*. The correlation, along with some computation procedure, is then used to construct possible emission timeseries for *A* based on emissions of *B* in scenarios produced by the first model.

There are many different procedures that can be used for this purpose, but in all cases it is a form of generalised regression analysis, which is then subject to the fundamental limitation of all types of regression: it requires that the statistical relationship between the lead and follower variables in the scenarios including both can also reasonably be assumed to hold in the scenarios that are to be infilled. Unless there is a direct physical correlation between the lead and follower variables (like products that are mainly produced by the same processes), this assumption is not likely to hold if there are major differences either in scenario design or in model dynamics that affect the emissions of either variable. For example, methane emissions may follow a very different pathway if coupled to fossil fuel use or agricultural activity. Conversely, the infilling procedure will potentially cause some part of the correlation structure between the lead and follower emission species in the scenarios used for infilling to become amplified in the processed results by being artificially transferred to model/scenario outputs that did not have any such structure to begin with. If the infilled emission species have a significant impact on the final climate outcomes, this further implies that the statistical correlations between climate category and other variables (including non-emission variables) could become biased in highly non-transparent ways.

Ideally, infilling would be done by carefully reviewing the scenarios and species to be infilled and then using expert judgement to select a structurally similar set of scenarios to use as data, the emission species to use as the lead variable, and the computation procedure for inferring the follower variable values. This has been done in individual studies—see, e.g.,(Schaeffer et al., 2015)—but is obviously not scalable to an assessment pipeline for a large scenario database like that of AR6.

The currently most widely used method for automated, scalable infilling is the Python package SILICONE (Lamboll et al., 2020), which includes several different modules (called "crunchers") that each implement a different method for inferring infilled follower emission species form lead species and a changeable data set of scenarios, referred to as the "infiller database". This package was used in AR6, with the database compiled from model scenarios that individually or in combination included all the emission species that were required for the climate assessment. These scenarios were taken from the harmonised AR6 database itself, where possible, but supplemented with scenarios from CMIP6 for species that were not included in a sufficient number of model scenarios in the AR6 database (Kikstra et al., 2022).

Infilling will inevitably to some degree result in imposing statistical correlations from the infiller database on the infilled scenarios, even if these correlations potentially conflict with the logic of the infilled scenario or with the most likely outcomes, if one were to extend the model that produced the scenario to include the infilled emission species natively. The resulting bias can be mitigated somewhat by ideally matching the infilled scenario to scenarios in the database that share similar narratives and assumptions, which may not be scalable. Or, alternatively, one needs to use a large and diverse database with a representative set of scenarios that represent all species natively, combined with an infilling method that uses a broad aggregate of database scenarios rather than algorithmically selecting a single "closest" scenario (depending on definition), so that at least the infilled scenarios will not bear a heavy imprint of the correlations from a single or small number of possibly biased database scenarios. This, however, results in a trade-off: if the infilling is based on a single or a small number of scenarios from the database, the result will inevitably be biased by the characteristics of those few scenarios; but, since the emissions trajectories are timeseries rather than single-point values, basing the infilling on a statistical aggregate of database scenarios instead may result in a trajectory with a kind of shape that no model would naturally produce. As discussed in Lamboll et al. (2020), some methods of this type can also be highly sensitive to exactly which scenarios are included in the database, result in unnaturally jagged timeseries, or change seemingly

chaotically in response to small changes in the database. Different choices of seemingly reasonable infilling methods may also give very different results. The choice of infilling method and data are therefore potentially large sources of additional uncertainty in the temperature assessment that should be investigated.

In AR6, infilling of most species was done using the "quantile rolling window" method (QRW), which uses an aggregate of many (potentially all) scenarios in the infiller database to infill each missing species. For each year, the infiller database is used to construct a function that maps from values of the lead variable in that year to values for the follower variable (which is to be infilled). The function is defined by first defining values of the follower variable at a number of discrete values of the lead, usually corresponding to quantiles of the lead variable (AR6 uses 10 points, the deciles of the lead variable), and completes the function by linearly interpolating between those points. The value of the follower variable at each point is determined by taking a weighted median of the follower variable over each scenario in the infiller database, with the weights determined by a function of how far the value of the lead variable is from the point value. The weighting function peaks at the point value, and decays quickly enough to become small for scenarios where the lead variable is more than one value point (one decile in the case of AR6) away from the given value point—see Lamboll et al. (2020) for details. For the choices in AR6, the decay ensures that the weighted median for the follower variable is dominated by values from the database scenarios that belong to the same decile for the lead variable as the lead-variable value in the scenario to be infilled.

Infilling of F-gases was done differently because of the limited number of model scenarios that reported them natively. These gases were instead infilled using the "RMS-closest" (Root Mean Square) method. This method does not use an aggregate of scenarios but instead infills the follower variable directly with values from a single "closest" scenario from the infiller database. The "closest" scenario is defined as being the one where the lead variable has the smallest sum-of-squares difference to the lead variable in the scenario to be infilled, where the sum is taken over all modelled years (or timesteps in general). Since the closeness is a function of the entire timeseries of the lead variable in the scenarios, a change in the lead variable value in any year in the scenario to be infilled can yield a different database scenario being selected as the closest one, so that a change of lead variable value in any single year can change the infilled values for the follower value across *all* years, and there is in principle no correlation between the size of the change in the lead variable and the resulting change in the follower variable.

These two methods used in AR6 both have pros and cons, and different biases. The QRW method used for most species by design results in a trajectory for the infilled species that has fewer extreme values and changes between time steps than the follower variable in the database scenarios, and that is potentially more middle-of-the-road than *any* of the database scenarios. This means that, although the QRW method is very stable to small changes in the infiller database scenario, it may not be suitable in situations where there are clear splits in how the infilled species behaves in distinct groups of scenarios. As a hypothetical example, take an infiller database where scenarios with relatively high CO_2 emissions are split into two groups with very different CH_4 emissions: 1) weak mitigation scenarios that try to compensate for lack of CO_2 emission reductions by aggressively mitigating methane emissions; and 2) laissez-faire scenarios with no mitigation of any kind, where emissions mostly follow GDP. High CO_2 emissions would correlate with *low* methane emissions in the first case, and *high* methane emissions in the second. But if both types are equally represented and the database is used to infill methane based on CO_2, a high-CO_2 scenario would then be infilled with methane emissions in the middle between the two extremes, which would correspond to no scenario in the database, and probably not with any kind of logic in the infilled scenario itself.

The RMS-closest method avoids this "unnaturally conservative" behaviour, since the infilled emissions always come from an actual modelled scenario. But it then of course suffers from maximal bias, in the sense of being determined by the behaviour of just a single model scenario from the infiller database, which could be governed by a scenario

logic that is not at all applicable to the infilled scenario.

Because of the high uncertainty that can be introduced by infilling, especially for large databases where manual scenario matching is not feasible, in AR6 it was chosen not to infill any of the major greenhouse gases (methane, nitrous oxide, or carbon dioxide—including AFOLU and process emissions) for the climate assessment presented in the WGIII report. All included scenarios were required to include these natively, as well as to report all emissions through 2100. Only relatively minor gases such as sulphates, organic compounds and F-gases were subject to infilling, although there remain some scenarios with infilled emissions for methane and nitrous oxide despite not being used in the climate assessment.

To some extent, weaknesses like those mentioned above will always be present for any infilling method that does not carefully choose infiller database scenarios that match the scenario logic and model dynamics of the scenario to be infilled. This makes it important to explore the difference in infilled emissions and review the temperature climate assessment that can result from different choices of infilling method, and also to explore alternatives to infilling that might either be more robust or that at least could give a more transparent contribution to the uncertainty in the final climate assessment.

We will conduct various sensitivity experiments on the infilling process to assess uncertainties and report findings in deliverable D5.2.

2.6 Simple climate modelling

In IPCC AR6, climate response uncertainty to scenarios was represented by parametric ensembles using three SCMs implemented in the common OpenSCM framework (CICERO, FaIR, and MAGICC). This process provided consistency throughout the report, connecting aspects of the physical climate assessment in the Working Group I (WGI) report to the scenario assessment in the WGIII report, but this process could be improved in future assessments and DIAMOND has been engaged with developing strategies for better documentation, consistency, and linkages between simple models as they are treated in the scenario pipeline. Key activities to help better estimate parametric and structural uncertainties include:

- *Improving OpenSCM pipeline:* DIAMOND/CICERO is one of the teams co-developing and testing the AR7 OpenSCM pipeline, involving a rapid climate assessment pipeline that combines information from the CMIP7 Fast Track/Rapid Evaluation Framework with coordinated experiments from SCMs.
- *Improving idealised tests for SCMs:* DIAMOND/CICERO has coordinated joint experiments for climate reversibility, which have been conducted both in Earth System Models and in SCMs (Sanderson et al.). These experiments, which were tested in a pilot project and will now be included in the CMIP7 fast track activity, cleanly test model responses to emissions rather than past exercises that have focussed on experiments with predefined concentrations.
- *Coordinated calibration/structural sampling:* Past assessments have seen each simple climate model taking independent strategies for model calibration. This means that the roles of data constraints, harmonisation, model structure, and calibration strategy cannot be clearly linked to future policy-relevant assessments (such as remaining carbon budgets). DIAMOND has developed a calibration strategy for CICERO-SCM (Sandstad et al., 2024), and is looking at how this type of activity can be coordinated across different SCMs.
- *Common software stack*: There are ongoing activities to reduce duplication between SCMs as they are represented within the common OpenSCM tool. DIAMOND/CICERO-SCM are working with FaIR and MAGICC developers to implement models with key common dependencies, allowing for robust

ensembles that can explore diverse climate representations and uncertainties.

A key activity leading into the second report of this activity (D5.2) is a more comprehensive uncertainty assessment that brings together uncertainties in harmonisation and infilling, together with the structural and parametric uncertainties from multiple SCMs via OpenSCM. This analysis aims to reflect the fact that IAMs producing only a subset of emissions (therefore requiring infilling) should be subject to larger uncertainties compared to a model that provides all emission species, with minimal if any (Type 2) harmonisation required.

2.7 Land-use uncertainties

2.7.1 Background

The carbon cycle community splits land into two components, *active* and *passive*. The *active* component is where there is a change in land use (e.g., from cropland to forestland) or where there is harvest (Figure 10). It is possible for other types of management to be included but are generally not modelled and the effects are likely much less important (e.g., thinning, fertilisation, management of grazing animals, etc). The *passive* component is a response to the changing climate and atmospheric composition—a feedback on the climate system (Figure 10).

In a simplified representation of the carbon cycle, the *active* components ($\alpha(E_F + E_L)$) and is exogenous to the system (an input), while the *passive* component is given by βC and represents an endogenous response (output),

$$\frac{dC}{dt} = \alpha(E_F + E_L) + \beta C$$

The Global Carbon Budget (GCB), Dynamic Global Vegetation Models (DGVMs), Bookkeeping Models (BMs), and Integrated Assessment Models (IAMs)—essentially all scientific research that makes the backbone of IPCC Assessment Reports–operate on this principle.

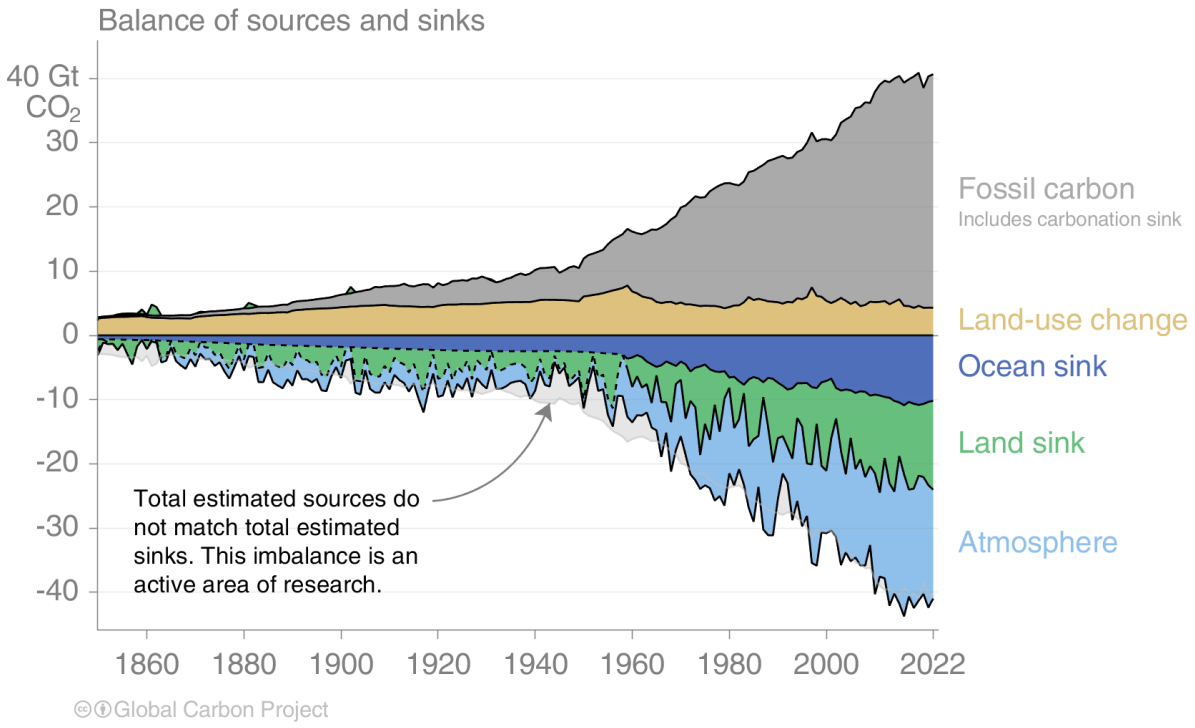


Figure 10. The global carbon budget showing sources (fossil emissions and *active* land-use change), passive sinks, and the atmospheric increase.

In contrast, the IPCC reporting guidelines used as the basis for National GHG Inventories (NGHGIs) use a different

definition of *anthropogenic* emissions and removals (Grassi et al., 2017; Grassi et al., 2018). They argue that it is not possible to split the active and passive emissions and removals, and therefore define anthropogenic to include both, but only on land that is defined as *managed* ("land where human interventions and practices have been applied to perform production, ecological, or social functions"). The IPCC reporting differentiates land processes into *direct* (active), *indirect* (passive), and *natural* (passive, such as disturbances). For the land defined as managed, direct, indirect, and in many cases natural, fluxes are included as anthropogenic. The 'managed land' area is also larger than the area where active processes occur. Thus, the IPCC reporting definitions capture a significant portion of the removals that IPCC assessment reports would allocate to non-anthropogenic activities.

The differences between definitions are more than just academic. More than one-half of the passive sink is defined to be managed based on country reporting. The active land-use change from the global carbon budget is translated from a net emission of over 4GtCO$_2$ per year to a net removal of over 3GtCO$_2$ per year, a difference of around 7GtCO$_2$ per year (Figure 11). The different estimates represent different definitions, and not errors. Numerous studies have now evaluated and assessed these differences and there are operational methods to translate between the definitions (Friedlingstein et al., 2023). However, translation shows that we can move between different definitions; the risk lies in the case net zero is defined with one definition (from IPCC assessment reports) but met with another definition (NGHGIs).

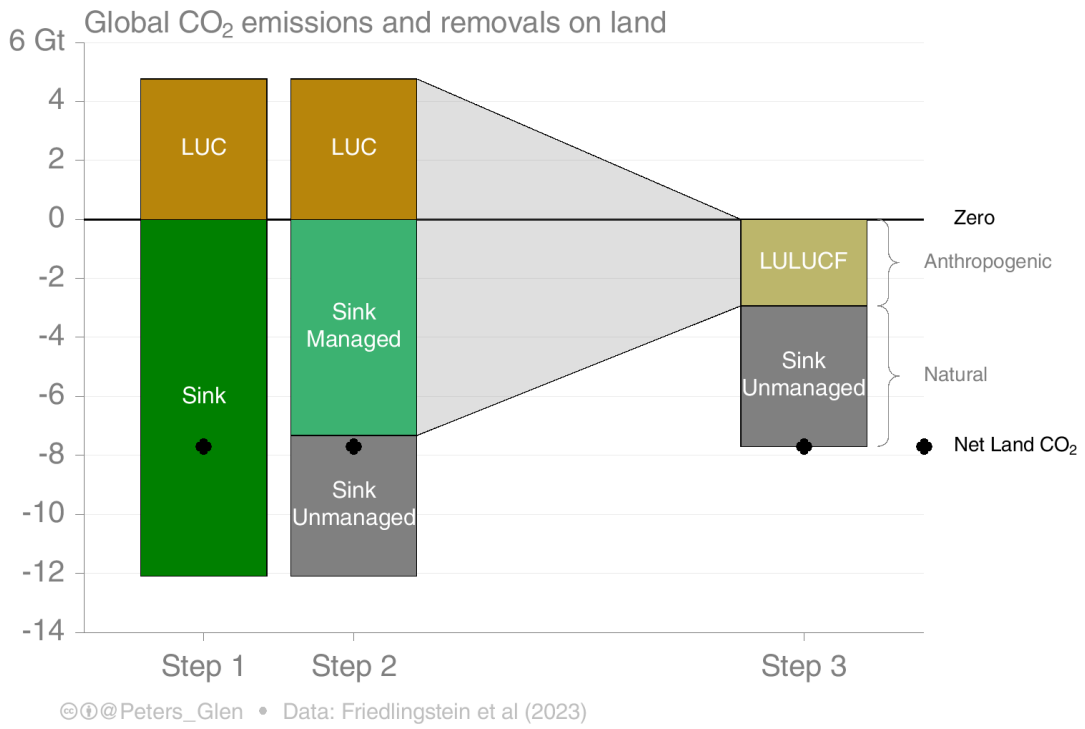


Figure 11. A schematic showing the translation from the Global Carbon Budget (GCB) definition of anthropogenic land-use change to the national GHG reporting definition. The GCB definition (Step 1) can be split into managed and unmanaged land, using national reporting (Step 2). The managed sources (LUC) and sinks can then be merged to translate to the national reporting definitions defined in UNFCCC (Step 3). In each step the net land emissions are the same. Source: Friedlingstein et al. (2023)

2.7.2 Geological net zero emissions

The goal of achieving net zero emissions, whether CO$_2$ or GHGs, has been defined based on the carbon cycle definition of land use emissions, active emissions, and removals. However, if net zero emissions are achieved based on NGHGI definitions, using passive uptake, then 'net zero emissions' will not stabilise temperature (Allen et al.,

2024). Allen et al. (2024) compares two scenarios, one that meets net zero CO_2 emissions using only active emissions and removals, and an extreme example that allocates all passive removals to anthropogenic. After net zero CO_2 emissions (active) are reached (black line), passive removals continue (green line) leading to falling CO_2 concentrations and stable temperature (Figure 12). This is how net zero emissions have been defined. If, on the other hand, passive removals are allowed to offset and balance active emissions (dotted lines), then the CO_2 emissions remain constant, and temperature increases. In this extreme example, by 2150, an additional 0.5°C or warming could be achieved, simply by adhering to the wrong definition of net zero emissions.

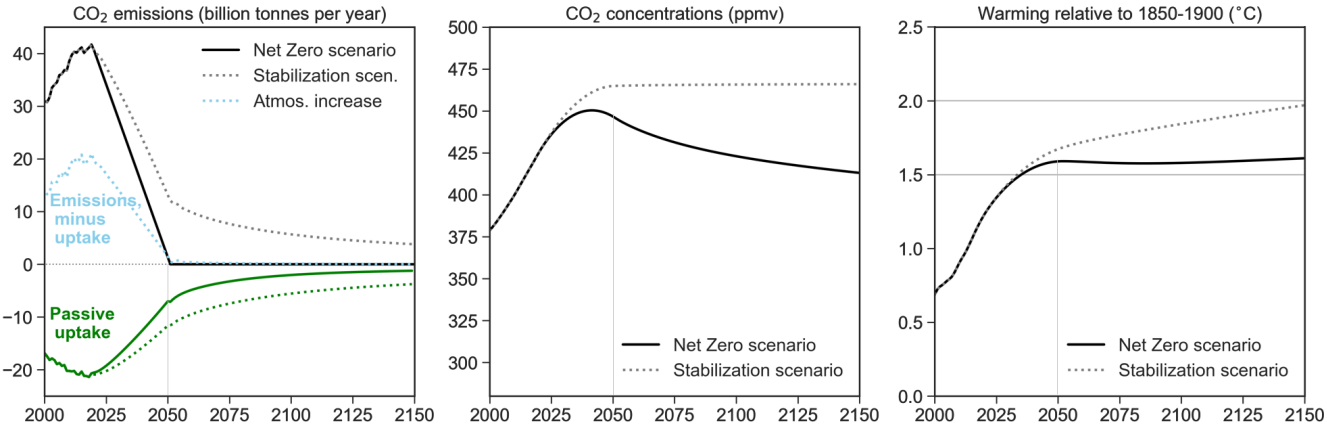


Figure 12. Different definitions of net zero emissions (left) lead to different concentrations (middle) and warming (right). If net zero emissions are defined based on active LUC emissions only, then concentrations will decline, and temperatures stay constant. If net zero includes passive uptake on land and oceans (sinks), then net zero emissions would only stabilise concentrations (as the net flux into the atmosphere is zero) and therefore rising temperatures. Figure from Allen et al (2024).

Because countries have different relative shares of forests and CO_2 removal on land, the differences at the country level can be more significant. Sweden, for example, already has net zero GHG emissions using the NGHGI emissions, despite the fact it continues to have significant active emissions. Whilst Sweden does not argue it should be allowed to offset all its emissions with LULUCF, this may change politically over time and other countries may view this differently. Currently, the EU allows all its LULUCF removals to offset active emissions, which has implications for Swedish interpretations, but the EU additionally has a separate LULUCF target, essentially making it more like two targets (one for emissions and one for removals).

2.7.3 Modelling land-use change

There have been emerging issues with how land use is modelled in a climate and policy context, but these multiple definitions bring new challenges. The Global Carbon Budget (GCB) now provides separate estimates of the different definitions (Friedlingstein et al., 2023), and the OSCAR simple climate model has the flexibility to translate results from IAMs between the different definitions (M. J. Gidden et al., 2023). However, questions remain on how to model land use from both a technical and a policy context. To date, IAMs have not included climate feedbacks onto land use modelling, but this is becoming an increasingly difficult assumption to maintain. Some land use models that incorporate climate feedback into land use modelling primarily focus on changes in crop yield due to increasing levels of CO_2 in the atmosphere and do not consider the impact on crop yield due to extreme events. Climate models provide gradual temperature changes in the long term and their impact on agriculture, which is considered as chronic risks, but often fail to capture the effects of inter-annual variability in weather—such as fluctuations in temperature and precipitation—and how this variability affects agricultural yields and food prices. In many regions of the world, LULUCF removals are weakening due to combined demographic, harvest, and emerging disturbances. Unless the land is modelled dynamically, these effects will not be represented in IAMs and

give policymakers potentially incorrect information. Additionally, IAMs have limited policies on land even when modelling active management, but the passive removals need to be considered more explicitly for IAM assessments to remain relevant.

D5.2 will perform a more systematic review of the challenges with modelling land within IAMs and ESMs and discuss how to identify key knowledge gaps. It will also take stock of the modelling capabilities in DIAMOND to identify where progress can be made in the context of stakeholder-driven research questions.

3 Scenario Definition Space

Scenarios are one of the key tools used in our understanding of the climate problem, its consequences, and how to mitigate or adapt to the consequences. Scenarios have also been a core part of IPCC processes, starting with the First Assessment Report (Bernthal et al., 1990). Because of their prominence, it is perhaps unsurprising that the role of scenarios in the IPCC has been critiqued over the years (Pedersen et al., 2022). There has long been a history of assessing how scenarios assessed by the IPCC have performed over time (Pedersen et al., 2021; Peters et al., 2013; D. J. van de Ven et al., 2025). However, here we focus specifically on the space of available scenarios, how well they cover the literature and hypothetical scenario space, and the relevance of scenarios, particularly in an IPCC context of 'policy-relevant and yet policy-neutral, never policy-prescriptive' (Hermansen, Boasson, & Peters, 2023).

There have been several critiques that some types of scenarios or narratives are overlooked (Pedersen et al., 2022). These 'missing scenarios' can be identified through literature review of studies that identify gaps (Anderson & Peters, 2016; Fuhrman et al., 2024; Matthew J. Gidden et al., 2023; Grubler et al., 2018; Kanitkar, Mythri, & Jayaraman, 2024; van Ruijven, 2016), quantitative analysis (Schweizer & O'Neill, 2014), expert elicitation (Grant, Hawkes, Mittal, & Gambhir, 2021), and so on. Questions also arise on the purpose of scenarios and how they are used (Peters et al., 2023a). Through the IPCC process there has been a focus on scenarios covering a wide range outcomes, such as low and high scenarios (Pedersen et al., 2022). Whilst the IS92 scenarios [6]had a specific research question around no policy intervention (Houghton, Callander, & Varney, 1992)—in response to the strong political backlash on the Supplementary Assessment 1990(SA90) scenarios (van Beek, Hajer, Pelzer, van Vuuren, & Cassen, 2020)—later IPCC assessment cycles have only stated that scenarios should consider a wide range of outcomes (Pedersen et al., 2022). Notably, the IPCC has identified no specific research questions or focus areas in SRES or later IPCC reports. Since AR5, scenario curation has been left up to the research community (Moss et al., 2010; Skea, Shukla, Al Khourdajie, & McCollum, 2021), though, at least indirectly, the IPCC Special Report on Global Warming of 1.5C(Rogelj, Shindell, et al., 2018) had a specific research question on deep mitigation pathways. Scenarios are compiled into a scenario database by the research community (Huppmann, Rogelj, et al., 2018), representing a range of research questions developed by individual modelling groups or consortia, with very little stakeholder engagement (A. Nikas et al., 2021).

The current state-of-the-art to define the hypothetical 'scenario space' is through voluntary submissions to a scenario database. Due to the process of curation, the scenario databases are not comprehensive and represent a mosaic of research questions (Peters et al., 2023b), often leading to the term 'ensemble of opportunity' (Huppmann, Rogelj, et al., 2018; Peters et al., 2023a). Based on the scenario databases, IPCC authors provide a range of database outputs on policy-relevant quantities (K. Riahi et al., 2022), such as emission reduction levels in 2030, net zero years, or shares of coal, oil, and gas in 2050, to mention a few. However, few of the scenarios have been constructed specifically to answer those questions, making it unclear how robust the answer is. Additionally, key assumptions such as cost-effective pathways using globally harmonised and uniform carbon prices, might not conform to the assumptions many policymakers would expect (Kanitkar et al., 2024). The prevalence of scenarios with large-scale carbon dioxide removal (CDR) (Anderson & Peters, 2016; Fuss et al., 2014; van Vuuren, Hof, van Sluisveld, & Riahi, 2017) arguably have become performative (Beck & Mahony, 2017, 2018) and could even reduce the variety of scenarios considered. The closeness of the IPCC and the scenario community (Cointe, Cassen, &

[6] IS92 scenarios refer to the six scenarios (IS92a to IS92f) developed by IPCC in 1992 for climate assessment

Nadaï, 2019) has also not alerted the scenario community over what is important to perceived as important to policymakers, as perhaps illustrated by the dearth of scenarios considering equity (Kanitkar et al., 2024). Given the lack of real structure behind scenario curation and a lack of motivating research questions, researchers themselves have decided what is important to policymakers and have essentially limited access to potentially relevant but underexplored areas of the scenario space.(Koasidis, Koutsellis, Xexakis, Nikas, & Doukas, 2023)

Given this context, we aim to address what emission scenarios are, define the scenario space, and discuss how the scenario space can be characterised and enriched. This initially builds on earlier work showing the challenges of using scenario databases to answer policy relevant questions (Guivarch et al., 2022). We then discuss the importance of clearly linking to research questions and subsets of relevant scenarios. We conclude with suggestions for how the scenario literature can evolve to provide much more policy-relevant analysis to feed future IPCC reports.

3.1 Emissions scenarios

The type of scenarios assessed by the IPCC has changed over the last 30 years. The original focus was on baseline scenarios in a process of scenario curation driven by the IPCC, without including climate policies (Houghton et al., 1992; IPCC, 2000). After the Special Report on Emission Scenarios (SRES) (IPCC, 2000), the IPCC's mandate to develop scenarios was contested van Beek et al. (2020) and the scientific community began developing scenarios with the IPCC having no formal role (Pedersen et al., 2022)—although many individuals were involved in both communities blurring the separation (Cointe et al., 2019). The Fifth Assessment Report considered a broad range of scenarios, but most of them were mitigation scenarios (Clarke et al., 2014). Likewise, the Sixth Assessment Report included a broad range of scenarios but focussed primarily on low-end scenarios in the assessment (K. Riahi et al., 2022). As the context of climate policy has changed, so has the demand for scenarios. The recent focus on equity (Kanitkar et al., 2024) is arguably a response to scenarios becoming more relevant and getting closer to what is needed: deep mitigation.

Scenarios have long been used in different contexts, often related to strategic planning or risk assessment. A scenario is often defined as a series of potential events or futures, that can be used to assess, qualitatively or quantitatively, how the events or futures may be managed. Here, we focus specifically on the use of scenarios in IPCC assessment reports.

The IPCC originally started with a rather generic definition of "emissions scenarios" (Houghton et al., 1992):

> *"Projections of future emissions based on specific assumptions about key determinants, such as population, economic growth, technological change, land-use trends or emission control policies."*

The IPCC "Climate Change 1994" report, which assessed the IS92 scenarios, laid out a method to assess scenarios (IPCC, 1994). They laid out four purposes of scenarios (reference scenarios, intervention scenarios, mitigation analysis, and input to negotiations) which they then evaluated against a range of criteria. They found the three most important criteria to be comprehensiveness, detailed output, and diversity (a range of views). Three other criteria were mentioned as important (reproducibility, sensitivity analysis, and internal consistency), and a range of others were also assessed (documentation, disaggregation of data, reference scenarios, a range of policy actions, and temporal calculations).

In the intervening years, starting with the IPCC SRES, the definition of scenarios changed to the current definition (IPCC, 2022a):

"A plausible description of how the future may develop based on a coherent and internally consistent set of assumptions about key driving forces (e.g., rate of technological change, prices) and relationships."

The definition now encompasses difficult to define terms, such as "plausible", "coherent", and "internally consistent". The IPCC does not define these terms, leaving it up to the scenario producer or scenario user to define them. These terms are contentious, as different views may exist on whether high-end scenarios are plausible (Hausfather & Peters, 2020), Similar arguments could be used on low-end scenarios (D.-J. van de Ven et al., 2023).

The current IPCC definition of emissions scenarios includes terms that were originally seen as criteria for evaluating a scenario's purpose. "Internally consistent", "coherent", and "plausible" could mean many things in different contexts. "Internally consistent" has been used in the context of soft-linking models (Meadows, 1981), ensuring scenario narratives are consistent (Schweizer & O'Neill, 2014), and could naturally apply when certain relationships and/or coherence are maintained. One could argue a key application of scenarios is to look for futures that are not "internally consistent", "coherent", or "plausible" as these are the ones that are difficult to manage and address, as originally done in the IPCC (IPCC, 1994). IAMs have been poor at capturing the dynamics of key technologies, which also depend on policy, such as nuclear (Keepin & Wynne, 1984), solar (Jaxa-Rozen & Trutnevyte, 2021), or CCS (Kazlou, Cherp, & Jewell, 2024), to name a few. While these issues could arguably be addressed with model or assumption improvements, real-world developments are often outside what earlier generations of scenarios saw as "plausible".

There is arguable value in retaining a rather generic scenario definition and rather linking more strongly to purpose and criteria in the assessment process (IPCC, 1994). Comprehensive scenarios are likely needed for some purposes (e.g., input to climate models), but for example scenarios to evaluate border carbon adjustments in the industrial sector meet another purpose and different criteria—arguably, equally important from a policy context. There may be value in assessing what may be seen as unfeasible (or implausible) pathways, to identify risks and strategies to minimise them (Hausfather and Peters, 2020). Ultimately, it is a question of the purpose of the scenario, and how the purpose changes over time. Originally, it was argued the IS92 scenarios were produced to evaluate baseline scenarios without policy intervention (IPCC, 1994). It is clear in the intervening years that the purpose of scenarios has changed. Scenarios now consider several other dimensions: along with the Kyoto Protocol came a surge in political (and scientific) interest in *mitigation* (as largely reflected in AR4), and along with the concrete and increasingly stringent temperature targets and the debate on negative emissions the notion of feasibility gained traction (as largely reflected in AR5); and scenarios are increasingly used and inscribed in the Paris Agreement and NDCs. The criteria against which to assess scenarios needs continuous revision, given the passage of time and new knowledge generated. This understanding seemingly remains a rather significant gap in the research literature.

3.2 The scenario space

Here we advance the discussion further to develop what may be termed the "scenario space". In statistics, the concepts of a population (all possible members of a group) and a sample (a selection from the population) are at the centre of any empirical analysis. Here, we propose the existence of a hypothetical "scenario space" covering all possible scenarios (i.e., the population of all scenarios). Then the question arises whether a sample of that scenario space is statistically representative, in that it would return statistical outcomes reflecting the entire population, and indeed whether such an approach is even feasible. While seemingly abstract, this is a highly relevant point. Are statistical summaries of the scenarios submitted to the IPCC Scenario Database (Peters et al., 2023a) reflective of the population of all scenarios—such as, level of coal in 2050 in a 1.5°C-consistent pathway? Would adding more scenarios to the database make it more representative? What does it mean to have an *inclusive* scenario space, covering a wide range of views?

In the physical climate sciences, one could expect that adding more observations improves the overall estimate. If estimating sea level rise, one could imagine a rather bounded uncertainty space (limited example by the laws of physics and observations) that, over time, converges to an accepted value (and range), and that can be later confirmed with observations. Does the same hold for emissions scenarios? This is much harder to discern. The IIASA scenarios of the early 1980s were largely dominated by nuclear (Keepin & Wynne, 1984), not considering the rapid growth of solar and wind, nor the emergence of then unknown technologies such as CCS and CDR. The scenario outcomes, while relevant at the time, are now meaningless. The world changed, in ways not depicted in those scenarios. Then what value can be placed on statistical ensembles of the current generation of scenarios, given they will be likely outdated as the world changes in ways not depicted in those scenarios? While the scenarios themselves may remain useful and provide useful information, summary statistics based on them may have limited value (Huntington, Weyant, & Sweeney, 1982).

We break down our discussion into different types of scenarios, to identify how scenarios may miss parts of the solution space and, within this discussion, therefore reflect the notion that a sample of the scenario space is unlikely to be representative, in a statistic sense, of the scenario population. The implications of assuming a scenario database is representative are then discussed later.

3.2.1 Narrative (qualitative) scenarios

A key element of a scenario, and perhaps the driving element, is the scenario storyline. This is defined by the IPCC as a "narrative description of a scenario (or family of scenarios), highlighting the main scenario characteristics, relationships between key driving forces and the dynamics of their evolution". Qualitative scenarios may be developed in a deliberative process with a cross section of interests, often trying to span different dimensions of a problem. Perhaps the most well-known type of qualitative scenarios is the quadrant or 2x2 matrix approach, often utilised in business(Amer, Daim, & Jetter, 2013).

In climate science, the Shared Socioeconomic Pathways (SSPs) are an example of a quadrant scenario design (O'Neill et al., 2017), spanning two axes of 'challenges to mitigation' and 'challenges to adaptation'. There are four narratives spanning the extremes of the quadrants, plus a 'middle-of-the-road' narrative (SSP2) at the centre. At this stage, the scenarios are just narratives and are not yet quantified. The narratives are not unique, nor are the axes of the quadrant. SSP1, known as 'sustainability', is a high-growth and equal world, but a different constellation of interests might derive a low-growth and equal world (e.g., degrowth or agrowth). The SSPs may not be exhaustive either, with narratives potentially missing relevant aspects (e.g., for finance or labour); for the example of economic growth above, post-growth notions (such as degrowth or agrowth) are not considered in the SSP framework at all(Walker Wood, Richter, & Atkins, 2024). And even though the SSPs consider equity, this does not appear to be the case when quantifying the scenarios (Kanitkar et al., 2024; Skeie, Peters, Fuglestvedt, & Andrew, 2021). The SSPs also assume that mitigation and adaptation actions/policies as well as climate change itself are independent dimensions in the scenario space, despite knowing that either and both change the background socioeconomics.

The SSPs are designed to answer questions spanning the challenges of mitigation and adaptation, but not all (and perhaps very few) research questions need to span those dimensions. Indeed, despite the perceived importance of the SSPs, they are barely cited in the IPCC AR6 WGIII report (Ipcc, 2022b) outside of cross references to the IPCC AR6 WGI application of the SSPs (e.g., SSPx-y naming convention). While it is stated that "Most of the scenarios in the AR6 database are SSP-based" (Box TS.5), this is since the vast majority of scenarios apply SSP2 baseline assumptions for socioeconomics (Figure 4), not because the studies compare across the different SSPs. The comparison of outcomes across SSPs is almost exclusively done in (Keywan Riahi et al., 2017), and around one-third of those scenarios did not pass the AR6 vetting due to their older vintage (Figure 2). If correctly

implemented, as in Riahi et al (2017), each SSPs would change a multitude of assumptions in a model, far beyond just switching to different socioeconomics. The additional workload would not be justified, unless it was required by the research question. The most representative study in the IPCC AR6 database, the ENGAGE study(Keywan Riahi et al., 2021), used SSP2 socioeconomics, but has a research question designed to investigate overshoot. These examples are not to downplay the importance of the SSPs but illustrate that the SSPs are now primarily used to provide baseline socioeconomics (SSP2) and not explore the interactions between the challenges of mitigation and adaptation.

The Network for Greening the Financial System (NGFS) scenarios are also based on the quadrants, spanning axes of transition and physical risks (NGFS, 2024). They are instead exclusively based on the middle-of-the-road (SSP2) socioeconomics, indicating that different socioeconomics is not a dimension the NGFS scenarios were designed to consider. They comprise seven different scenarios that sit in the different quadrants of the scenario space, with three in one quadrant alone. The multiple scenarios in a single quadrant essentially add a third dimension to the NGFS, representing different climate outcomes (1.5°C or 2°C) or policy approaches (low demand). Unsurprisingly, the NGFS scenarios are designed in a way to address issues the Network find relevant and important, but those scenarios may not be relevant for other scenario users.

The International Energy Agency (IEA) also produces scenarios relevant for climate. The IEA only explores one socioeconomic dimension (not an SSP), but spans different policy storylines motivated by climate policy (Stated Policies, Announced Pledges, Net Zero in 2050) (IEA, 2024). Many energy companies and other organisations produce scenarios as well, each with their own narratives and context, even though many share similar elements (Skea, van Diemen, Portugal-Pereira, & Khourdajie, 2021). Qualitative scenarios do not need to be based on quadrants but could be based on a process to simply address relevant challenges (Böttcher et al., 2024).

Qualitative scenarios, even if based on clearly defined quadrants, are unlikely to be unique nor representative. The framing, context, timing, and attendees of a process may lead to different outcomes. The SSPs were largely built on the SRES scenarios, drawing on some expert meetings((Schweizer & O'Neill, 2013; Wilbanks & Ebi, 2014)), but were ultimately developed by the authors of the SSP narratives (O'Neill et al., 2017). The SSPs would be expected to look different if they involved other or more authors, including from developing countries, different user communities, different research communities, and so on. The SSPs share many elements of the earlier SRES scenarios and other earlier scenario exercises (van Vuuren & Carter, 2013)), indicating that legacy is also an important factor. In the context of the scenario space, it is important to recognise that the SSPs, NGFS, IEA, or other exercises will only look at one segment of the scenario space and the perspective is not unique.

A transdisciplinary approach underpinned by transdisciplinary principles (Stephanie Briers & Vienni-Baptista, 2024) is therefore useful, if not essential, in a scenario process to ensure inclusivity and a variety of views(Alexandros Nikas et al., 2020). Some transdisciplinary principles that will pave the way to robust and inclusive scenarios are making scenarios much more contextually specific (Hoffmann, Deutsch, Klein, & O'Rourke, 2022), embracing plural perspectives (Vienni-Baptista, Fletcher, Lyall, & Pohl, 2022) in scenarios, and allowing scenarios to integrate diverse knowledge and worldviews (Bornemann & Christen, 2020). These, among other principles, can ensure that legitimacy of scenarios is increased, also increasing the sense of ownership over scenarios and therefore their increased use as tools to support decision-making (Norström et al., 2020).

3.2.2 Quantitative scenarios

In climate science, qualitative scenarios are *generally* turned into quantitative scenarios (Alcamo, 2008). To illustrate this, three steps in the SSP quantification process are highlighted (Keywan Riahi et al., 2017) . First, models are used to develop quantitative population and GDP projections based on the SSP narratives (also education, urbanisation, etc.). Second, these population and GDP projections are used to estimate end-use demand that are

inputted into Integrated Assessment Models (IAMs) to generate pathways for energy, emissions, land use, and so on. The initial pathways for the quantified SSPs were baseline scenarios, which assumed zero climate policy (effectively, a carbon price of zero). Third, Shared Policy Assumptions (SPAs) can be added to the baseline scenarios (Kriegler et al., 2014), as a third dimension to the SSP framework, which in the quantified SSPs were selected to reach different levels of radiative forcing in 2100, known as the Representative Concentration Pathways (RCPs)—which, in turn, link to different temperature levels. The applied policies are generally globally harmonised carbon prices, sometimes with delayed implementation (Kriegler et al., 2014). The SSP-SPA combination, in turn, leads to SSP-RCP pairs; for example, SSP1-19 is a scenario using SSP1 ('sustainability') and reaching 1.9W/m^2 in 2100 (~1.5°C) and SSP5-85 a scenario using SSP5 (fossil fuel development) and reaching 8.5W/m^2 in 2100 (~4°C). The SSP framework allows one to assess how mitigation and adaptation may change in different SSPs (Keywan Riahi et al., 2017; Rogelj, Popp, et al., 2018); however, most scenarios developed today are based on middle-of-the-road (SSP2) socioeconomics (Figure 4) and do not assess sensitivities to different SSPs (Peters et al., 2023a).

The scenario quantification process can serve many functionalities. The quantification process can be a method to assess for coherence and internal consistency of the scenario narrative (Schweizer & O'Neill, 2014), although this element is little discussed. The quantification process provides a variety of scenarios for application in climate science (Keywan Riahi et al., 2017) or other areas. In the IPCC context, the quantified scenarios are assessed, and summary statistics of selected scenarios are used to feed key messages in the IPCC Summary for Policymakers (SPMs). However, preceding each quantified scenario, there is often a specific research question or scenario exercise, which is often lost when scenarios are mixed into a database (Peters et al., 2023a)—the scenario exercise connecting the research question to the quantified scenario may even be omitted in the scientific publication underpinning it, as there is often much larger (academic, policy, etc.) interest in the outcomes in response to a research question than in the process leading to the question or the process of quantifying a qualitative narrative into a scenario.

The quantified SSPs were initially run to generate a matrix of SSPs (5x), forcing levels or RCPs (6x, including 1.5°C), across six selected IAMs (6x). Of the 180 possible SSP-RCP-IAM combinations, not all were run—and among those, not all were necessarily feasible (Keywan Riahi et al., 2017; Rogelj, Popp, et al., 2018). Some scenario combinations were found to be infeasible (e.g., reaching 1.9W/m^2 in an unequal world, SSP4, or reaching 8.5W/m^2 without SSP5 socioeconomics). These findings could indicate that the scenario was not "plausible", or that the IAM did not have sufficient flexibility to allow a solution—albeit some combinations are largely established as infeasible in the literature(Ritchie & Dowlatabadi, 2017). Not all scenario combinations were attempted, indicating either a lack of capacity or an expectation that the run would be infeasible. For the scenario combinations that had feasible solutions, the resulting outputs reveal information on mitigation costs of different climate targets with different socioeconomic assumptions and models. However, even among these feasible runs, some variables were missing, likely due to varying model capabilities. This variability in data availability introduces challenges in asserting that these outputs collectively cover the scenario space. For example, variables such as hydrogen and solar consumption, geothermal energy supply, or oil with CCS had limited reporting across models in the SSP database (Al Khourdajie et al., 2024). These gaps highlight how model design and reporting inconsistencies can leave parts of the scenario space underexplored, emphasising the need for careful contextualisation of results before drawing broad conclusions.

Depending on the Shared Socio-economic Pathways- Shared climate Policy Assumptions (SSP-SPA) combination, a range of alternative pathways were developed (Figure 13). For primary energy, all pathways in SSP5 are higher than those in SSP1, regardless of the level of mitigation. This demonstrates the important role that socioeconomics plays in dictating scenario outcomes, with those factors assessed for other scenario outcomes such as emissions, temperature, or mitigation costs. This was a key application of the SSP framework, though, arguably the framework

has had a less notable impact in other areas of study, such as adaptation research (O'Neill et al., 2020).

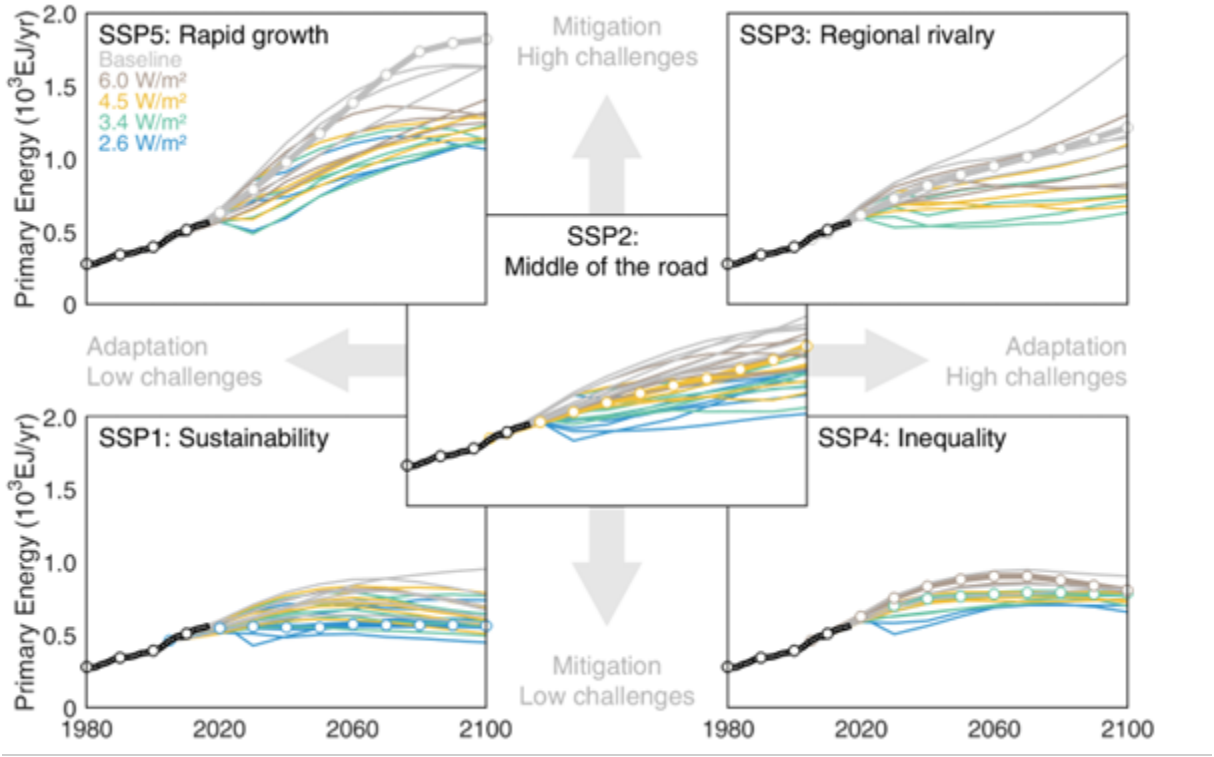


Figure 13. The SSP-RCP combinations placed on the SSP narrative axes of challenges to mitigation and challenges to adaptation, illustrating how primary energy evolves under different SSPs (boxes) and different SPAs (colours).

While the SSP framework has led to interesting findings, the original goal was to facilitate a greater linkage across mitigation and adaptation and speed up the scenario generation process (Moss et al. 2010). One could question how successful the process was on both criteria: modelling focus has predominantly remained on mitigation(van Maanen et al., 2023), while an assessment of how slower or faster the scenario generation process would have been without the SSP framework is no straightforward task. Nevertheless, a selection was made of the potentially 180 individual quantified scenarios across SSPs, RCPs, and IAMs deemed to be of the most relevant for climate research in an exercise known as ScenarioMIP (O'Neill et al. 2016) and IPCC AR6 WGI is entirely populated by the assessment of the climate outcomes of those scenarios. While that means the scenarios have extensive citation and use, that is because of the important position they were given in the IPCC process, not necessarily that the scenarios were more relevant than alternative scenarios. The SSPs were an important input in the IPCC Special Report on Global Warming of 1.5°C (SR15) but were less relevant in IPCC AR6 WGIII. By far, the biggest application of SSPs is throw ScenarioMIP, and the assessment of physical science outcomes (IPCC AR6 WGI). The IPCC AR6 WGIII report in fact barely mentions the SSPs but focuses the assessment on other scenarios that were deemed more relevant. Of the 126 scenarios from the original SSP exercise submitted to the database, only 70 passed the 'vetting' by the IPCC author team and received a climate assessment. Ultimately, the SSPs comprised 70 of the 1,202 climate assessed scenarios. Despite the dominance of the SSPs in WGI, they have very little role in WGIII, bringing into question their role as a framing concept for scenarios more broadly, including in the future use of ScenarioMIP.

Al Khourdajie et al. (2024) examine what drives the differences in future energy scenarios by trying to attribute variations in key variables to three main factors: how ambitious climate policies are, which economic and social futures were assumed (as in which SSP), and which model is used. Using Shapley-Owen decomposition analysis,

they show that a significant challenge emerges from missing data across model outputs. For example, out of all variables studied, oil with CCS is only reported by three models, while hydrogen deployment data is missing from 49 out of 126 possible model outputs. Similarly, data for geothermal energy is absent from 48 models, and industrial sector energy use is missing from 42 models. These reporting gaps create challenges for robust cross-model comparisons and make it harder to understand what truly drives variation in these variables. The limited reporting of certain variables may indicate technological uncertainties, modelling challenges, or differing priorities across modelling.

The SSPs were a selection of scenarios based on five alternative narratives of the future world. It has already been noted these narratives are not unique, nor comprehensive. To illustrate this, when quantified, the SSPs do not span a range of probabilistic population projections from the United Nations Population Division (Giarola et al. 2024) (Figure 14). Indeed, while being labelled 'middle-of-the-road', SSP2 is also below the median projection of other scenarios. When combined with the dominance of SSP2 socioeconomics in the scenario database, the database is clearly not representative of population and thereby of other factors as well. Furthermore, as shown for primary energy (Figure 13), SSP2 across a range of IAMs and climate outcomes does not span the full range of the scenario space (SSP1, 3, 4, 5). It is therefore hard to argue that the IPCC AR6 WGIII scenarios database spans even the existing literature very well, not for population, energy, nor any other variable. Even if the database captures the range of outcomes (maximum-minimum) it is hard to make the case the statistical distribution in between is at all meaningful.

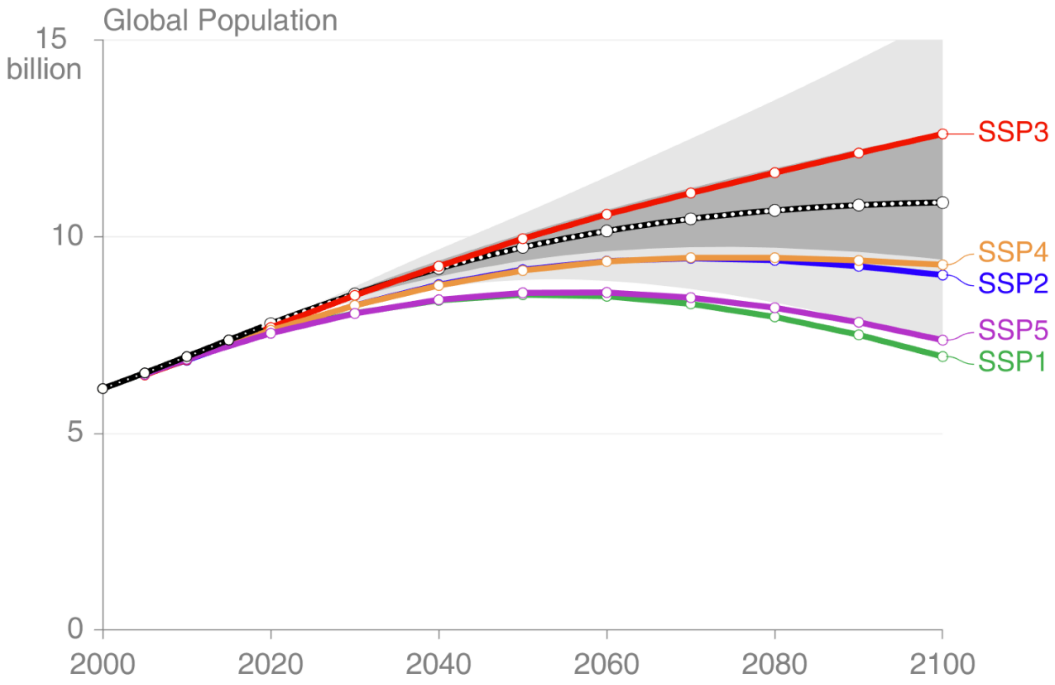


Figure 14. Population projections based on probabilistic scenarios from the UN Population Division (shading showing mean in black, and the interquartile and full range in grey) and the second release of the SSP population data.

3.2.3 The hypothetical scenario space

An infinite number of scenarios could be envisaged, and here we conceptually illustrate how this hypothetical 'scenario space' could be visualised (Figure 15). In the context of the IPCC, some scenarios may be published in academic or grey literature and could potentially be assessed: these may come from multi-model comparisons (using the same scenario protocol) or single-model studies. The IPCC has for decades either curated, or

encouraged the curation of, a scenario database (Peters et al., 2023a). The scenario database therefore captures, at least, a part of the scenario space. However, scenarios are based on voluntary submissions. Some scenarios in the academic or grey literature are not submitted to the database but are assessed regardless—for example, the IEA scenarios[5]. In AR6, the IPCC assessment was entirely focused on scenarios submitted to the database and not scenarios in the broader literature. Thus, an IPCC assessment will cover a part of the known literature, but not necessarily all of it, despite the IPCC mandate to assess the full breadth of available scientific literature. Then there exist scenarios adjacent to the existing published literature but applied with a broader range of models or modelling frameworks (Gambhir, Ganguly, & Mittal, 2022; Koasidis, Nikas, & Doukas, 2023). To illustrate this, around 50 modelling frameworks submitted scenarios to the IPCC AR6 database, but the 1.5°C scenarios (C1) originated from only nine model frameworks, indicating parts of the scenario space that may not be explored across a diverse range of model frameworks. And finally, there is also a large part of the scenario space which is not explored: either because a known scenario has not been produced (e.g., scenarios of general artificial intelligence) or because of scenarios can't cover unknown unknowns (e.g., one hundred years ago computers were unknown). The definition of "scenario" itself may also limit the size of the scenario space. It should be clear that the hypothetical scenario space will never be fully populated, and therefore great care is needed when assessing this scenario space (Guivarch et al., 2022).

Table 3. Different scenario types that exist in the hypothetical scenario space

Considered scenario type	Examples
Published and in the IPCC-assessed scenarios database	Scenarios in the multi-model comparisons projects
Published and not in the IPCC-assessed scenarios database	Several scenarios from the International Energy Agency[7]
Scenarios that are minor variations to the published literature	Slight variations in costs of key technologies that change results, but not markedly
Scenarios that we know exist but have not been performed (or performed by only very few models)	Scenarios that endogenously integrate climate impacts onto the land sector and affect CDR and its impact on mitigation pathways, Scenarios that assess the impact of AI on mitigation cost and power-sector transition
Scenarios that we do not yet know exist	100 years ago, it was unknown that computers would exist; futures we may view as highly unrealistic (unlimited cheap energy, space travel, etc.)

[7] The AR6 Scenarios Database has three IEA scenarios: One scenario from the World Energy Outlook 2021 (NZE), one from the World Energy Outlook 2020 (SDS), and one from the Energy Technology Perspectives 2020 (B2DS). These scenarios are from different years or different models and are not comparable. In WEO2021, other scenarios could have been submitted, including STEPS, APS, and SDS. Likewise, WEOs from other years had many other scenarios that could have been in the AR6 Scenario Database.

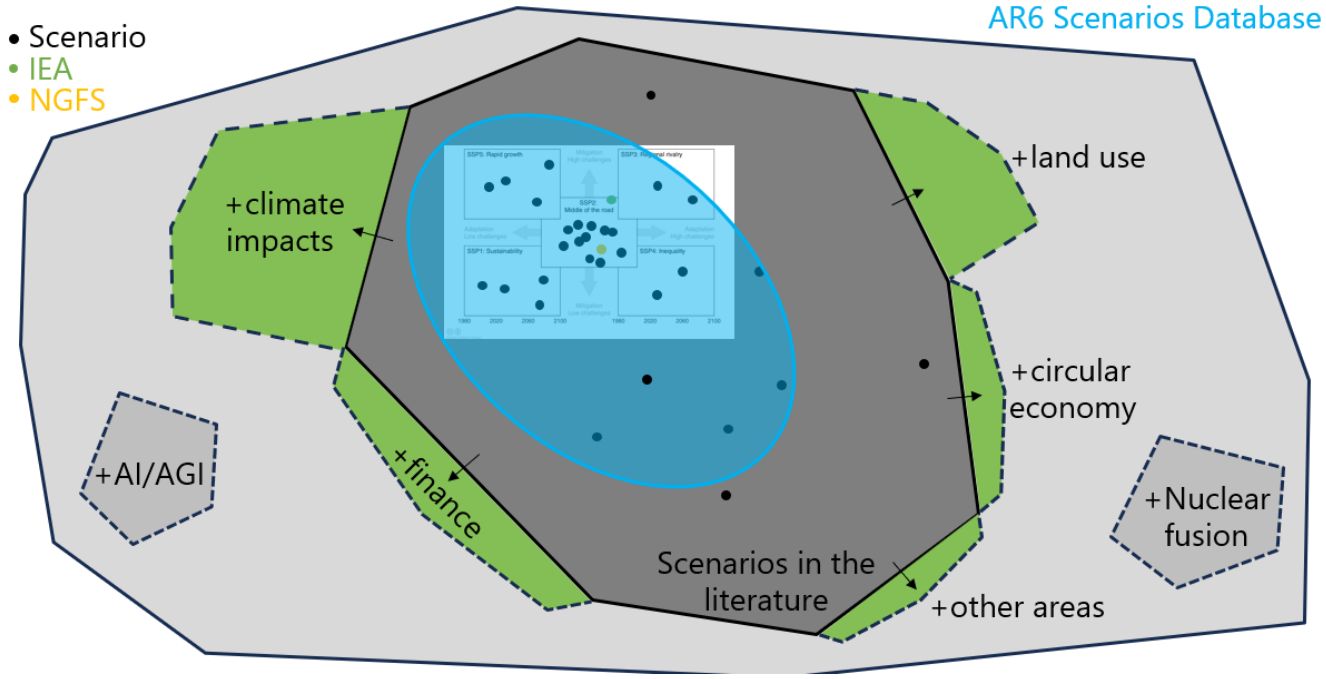


Figure 15. A schematic of a hypothetic scenarios space (light grey). The current scenarios space (dark grey is partially captured by the AR6 Scenarios Database (blue) of which a smaller subset is vetted and assessed. Some scenarios or groups of scenarios are indicated by black dots, which primarily fall within the scenarios assessed by the IPCC. The SSPs are shown in the white box with quadrants. Other prominent scenario exercises are shown as examples (e.g., IEA and NGFS), some of which are assessed also by the IPCC and in the AR6 Scenarios Database. Over time, more of the scenario space is filled, either by infilling with more detail or expanding to new areas, with illustrations showing climate impacts, finance, circular economy, and land use (green). Other scenarios might be further from existing scenarios, such as ones dominated with Artificial General Intelligence (AGI) or nuclear fusion, to show two random examples. The reality is far more complex, but this schematic is to demonstrate that the existing scenario space is limited in extent and key dimensions are not fully explored.

Even within the known scenario space, it is important to point out that there are a multitude of gaps (Sections 4.1, 4.2). The IAMs or models used to generate scenarios, the variables (outputs) reported, and the associated analysis are also likely to be non-exhaustive. It is not clear how or whether one can define the boundaries of the scenario space that could usefully be explored by models, or how the coverage of the scenario database over such a space should be defined or quantified.

3.3 Characterising the scenario space

Acknowledging that decision makers are to some extent, directly or indirectly, involved in the definition of research questions and thus of many scenarios stemming from these questions, for robust decision making in the context of climate adaptation and mitigation, it is important that we know the scenario space is sufficiently sampled for the given research question. How well, and what part, of the scenario space needs to be well sampled will depend on the research question. The global GHG emission reductions required to meet a given climate target may best be answered using one part of the scenario space, while an assessment of financial flows may require different models and a different part of the scenario space (Alexandros Nikas, Frilingou, Al Khourdajie, & Gambhir, 2024).

Here we outline three broad ways that the scenario space could be characterised using: literature review, quantitative analysis, and qualitative methods. Conceptually, one could think of a taxonomy of scenario types and spaces. Our objective is not to limit the scenario space, but to better classify it.

3.3.1 Literature review

One way to identify missing parts of the scenario space is through literature review. There are several policy relevant areas where it is known that the literature is lacking, such as, circular economy (Alexandros Nikas et al., 2022; Pauliuk, Arvesen, Stadler, & Hertwich, 2017; Sovacool et al., 2020), land-based mitigation (Roe et al., 2019), climate finance (Hong, Karolyi, & Scheinkman, 2020), labour (Ram, Osorio-Aravena, Aghahosseini, Bogdanov, & Breyer, 2022), post-growth (Hickel et al., 2021), extremes (D. L. McCollum, Gambhir, Rogelj, & Wilson, 2020), behaviour (Hahnel, Chatelain, Conte, Piana, & Brosch, 2020), to name just a few areas. In many of these cases, models were not capable of assessing many of the most relevant and/or pertinent issues and model development is therefore needed. Expanding models to address these issues gives them the capability of answering a wider range of research questions (Pye et al., 2021). While parts of the scenario space may not be fully elaborated, transdisciplinary processes and principles, including stakeholder engagement, can be used to co-create and refine the research questions to provide more relevant information within the scenario space (Stephanie Briers & Vienni-Baptista, 2024). It is unknown how incorporating these dimensions and scenarios based on them will change the scenario space: for example, how will the emission reductions required in 2030 change based on better modelling of finance, labour, or the circular economy? Calcaterra et al. (2024) showed that how emissions change at the regional level in the baseline scenario by using better cost of capital estimates instead of uniform cost of capital across all regions in the Integrated assessment models.

There are many other areas where gaps have been identified in the existing scenario. Recent papers have identified that equity has not been adequately assessed in the IPCC process (Kanitkar et al., 2024), with suggested pathways forward with alternative scenarios (Ranjan, Kanitkar, & Jayaraman). While the SSPs do include equity, particularly through the "challenge to adaptation" axis, these narratives are insufficient or poorly implemented in models so that there is no clear distinction on key aspects of equity (Skeie et al., 2021). Since most global coal consumption occurs in developing countries, it is expected that the inclusion of equity would lead to greater reductions in oil and gas in the developed countries, with less reductions in coal in the developing countries (Bauer et al., 2020; Muttitt, Price, Pye, & Welsby, 2023). Thus, the inclusion of equity may meaningfully change key characteristics, both globally and regionally, of mitigation pathways.

Climate impacts are rarely included in mitigation studies. While they are included in cost-benefit analyses comparing the trade-off between climate impacts and climate mitigation, mitigation studies (IPCC WGIII) rarely consider climate impacts. In fact, the SSPs are explicitly designed not to include climate impacts, rather allowing other communities to assess the climate impact. While this may make sense in some contexts (e.g., comparing across 2°C scenarios with the same assumed climate impacts), it does not make sense when comparing scenarios with different warming levels. The high-end scenario, SSP5-85, leads to warming of around 4°C in 2100, with zero impacts on climate change, a highly unlikely outcome (Hausfather & Peters, 2020), but this scenario was given the "highest priority" (O'Neill et al., 2016). Including climate impacts in mitigation scenarios would change mitigation pathways and costs, as the chunk of the GDP spent on climate impacts and adaptation would crowd out investment in low-carbon technologies. The exclusion of climate impacts in mitigation scenarios clearly leaves a big gap in the scenario space.

Since the emergence of CDR in scenarios, particularly since the publication of the IPCC AR5, there has been considerable debate on its role in scenarios (Fuss et al. 2014; Anderson and Peters 2016; van Vuuren et al. 2017). While CDR has several roles in mitigation pathways (K. Riahi et al. 2022), the use of large-scale CDR does lead to slower short-term mitigation and thereby constitutes a risk of moral hazard. Some scenarios have specifically been designed to avoid large-scale CDR (and in many cases CCS). Scenarios based on SSP1 generally have lower energy demand (Figure 1) and thus less need for CDR, but low-demand scenarios more broadly lead to lower levels of CDR (Grubler et al., 2018; van Vuuren et al., 2018). Degrowth scenarios were noted in the IPCC assessment report

but not included in the database.(Kikstra et al., 2024) Despite this, most scenarios are still at the high end of CDR feasibility levels (Grant et al., 2021). Whilst zero or low CDR is represented in the scenario space, those scenarios are far outweighed by scenarios with high levels of CDR, clearly biasing statistical ensembles to lower mitigation rates in the short term (Anderson et al., 2023).

Another well-known weakness of the current generation of scenarios is the types of policies implemented. By far, scenarios are dominated by the idealised implementation of global carbon pricing, despite the fact this is unlikely to be a reality. While many current policy and NDC scenarios explicitly model real-world policies (Roelfsema et al., 2020; Sognnaes et al., 2021; D.-J. van de Ven et al., 2023), extensions behind the current policy time frame (often 2030) often revert to standardised policy assumptions (Sognnaes et al., 2021). Relatively few scenarios model policies over long-time frames, with some rare exceptions (Morris, Hone, Haigh, Sokolov, & Paltsev, 2023; van Soest et al., 2021). The IEA uses carbon pricing in addition to other policies, with carbon pricing generally used in power and industrial sectors and other policies elsewhere, to provide a more nuanced and realistic policy portfolio (IEA, 2024). Policy and governance are generally poorly represented in the current generation of IAMs (Bertram et al., 2024; Peng et al., 2021). Scenarios will become more policy-relevant if they are able to represent more realistic policies (considered to be) implemented by countries in the context of political and social constraints.

Literature review is useful to clearly identify gaps in the scenario space. Most of these gaps are not yet filled and, when filled (and assuming evenly distributed with other scenarios), a given scenario database may become *more* representative but will likely remain a non-representative sample of the entire scenario space.

3.3.2 Quantitative methods

To supplement literature review, quantitative methods can be used to help characterise the existing scenario space. We separate the discussion into current approaches (statistical summaries) and emerging methods (becoming more prevalent in the literature).

3.3.2.1 Summary statistics

One potential avenue to quantify the gaps in the scenario space is the use of the IPCC scenario database. Groups voluntarily submit scenarios to a scenario database (Peters et al. 2023a), which is then used in the IPCC assessment process (K. Riahi et al. 2022). This scenario database spans a limited part of the scenario space, as previously discussed, and is often referred to as an 'ensemble of opportunity' because of this (Huppmann et al. 2018). Using the AR6 scenario database, it is possible to get an idea of the coverage. In the global database, 919 scenarios were submitted from 109 unique models. These models, after adjusting for version numbers and hybrid versions, leads to around 50 unique model families. After a vetting process and the allocation of a climate category, only 1,202 from 14 model families made it. Within that distribution, 297 came from one model (family), and 591 from one study (ENGAGE, Riahi et al 2021). The summary statistics used in IPCC assessment reports are thus heavily affected by the dominance of this one model and one study. In terms of socioeconomics, Majority of the scenarios use SSP2 socioeconomic assumptions. Clearly, the summary statistics indicate significant bias in the types of submissions in the database, and therefore its representativeness.

Based on the scenarios database, it is common to report various summary statistics, most notably percentiles (median, 25-75, 5-95). These summary statistics report what is available is represented in the database, but unless the database is a representative statistical sample of the wider scenario space, the summary statistics may reveal little about the underlying statistical properties of the entire scenario population. This has been demonstrated for the AR6 Scenarios Database, where the uneven representation of models and project (Figure 2) is found to significantly skew the statistical distributions (Sognnaes & Peters, 2025). The dominant model or project is found to considerably change policy relevant outcomes (e.g., Figure 3), with most key findings in the IPCC AR6 WGIII

Summary for Policy Makers more indicative of the dominant model than the underlying scenario dataset (Sognnaes & Peters, 2025).

3.3.2.2 Emerging quantitative methods

In recent years there has been interest in classifying the scenario space using higher-level statistical and mathematical methods (Guivarch et al. 2022). Key methods employed include Sobol decomposition (Dekker et al. 2023), clustering (Minx et al. 2024), regression trees (Achakulwisut et al. 2023), Shapley-Owen decomposition (Al Khourdajie, Skea, and Green 2024), and similar. While all these methods have advantages and disadvantages, they all share the weakness that they can only characterise the existing scenario space and not the hypothetical scenario space. Nevertheless, by exploring the characteristics of the existing scenario space, gaps in the literature are identified, allowing a better overall quantification of parts of the scenario space that may be missing or poorly explored. Scenarios are generally classified in a rather simplistic way (K. Riahi et al. 2022): temperature levels, policy, and socioeconomic and model metadata. Higher-level statistical and mathematical methods can help give a greater granularity of the scenario space, and more clearly identify the strengths and weaknesses of different scenarios, models, and missing parts of the scenario space.

3.3.3 Co-designing the scenario space

A notable approach to characterising the scenario space is through diversifying exactly who contributes to shaping the scenario space and to what end. The gap between how scenarios are used, by whom, and to what purpose (i.e., in response to what needs) on the one hand, and how scenarios are created on the other hand, represents how increased involvement of diverse scientific and societal actors would benefit scenario development and transparency. Here, a transdisciplinary approach, guided by key transdisciplinary principles and competences (S. Briers & Vienni Baptista, Forthcoming) offers a meaningful process and position from which to reimagine scenario spaces. Defining a representative sample of actors involved on such processes is also extremely difficult, because it is also essentially impossible to define the full population of stakeholders, as it is for scenarios. However, by being as inclusive as possible, a wider range of perspectives may be drawn in, helping to expand the scenario space, particularly in the areas of relevance.

Transdisciplinary approaches help to understand and address highly contextual and complex problems based on the integration of different knowledge types and worldviews from diverse societal actors to develop legitimated (Lang et al., 2012) and transformative solutions (Lawrence, Williams, Nanz, & Renn, 2022). This approach proposes a transformative shift towards inclusive stakeholder engagement to address key limitations of traditional, expert-driven approaches to evidence-based policymaking (Herbig, Briers, & Vienni-Baptista, forthcoming), which have remained unresolved even through more progressive approaches, such as participatory modelling (Hedelin et al., 2021; Moallemi et al., 2021). While participatory modelling engages non-academic actors or stakeholders for feedback and validation of specific assumptions(McGookin et al., 2024), a transdisciplinary approach moves towards co-production of knowledge in science and policy (Galende-Sánchez & Sorman, 2021; Alexandros Nikas et al., 2020). In consolidating a scenario space, a transdisciplinary principle represents guiding orientations to integrate co-production at the level of the project, the process, or the outcomes (S. Briers & Vienni Baptista, Forthcoming). Recognising that knowledge flows across communities, the transdisciplinary approach ensures a policy-relevant yet grounded flow of comprehensive and comprehensible data to be integrated in the scenario database and therefore the scenario space (Stephanie Briers & Vienni-Baptista, 2024). By applying methods that incorporate diverse perspectives into the knowledge production process, building on an advanced transdisciplinary approach centred around inclusivity, DIAMOND seeks to support the development of a scenario space that is not only scientifically sound but also socially robust, transparent and equitable (Herbig, 2024).

The first of the three phases of a transdisciplinary research design involves familiarising with the problem space and all the relevant actors (Stephanie Briers & Vienni-Baptista, 2024), and then jointly framing the problem that needs addressing (Schäfer & Kröger, 2016). The outcome of such a jointly framed problem is a set of stakeholder-informed research questions that guide the second phase, where solutions are co-produced and then legitimated and co-owned in the integrative final phase of the transdisciplinary approach, towards achieving transformative change (Jahn, Bergmann, & Keil, 2012). Here, scenario development would benefit from a joint problem framing approach that includes diverse actors to co-produce research questions underpinned by the specificities of the jointly framed problem. Such questions can take a step towards scenario development, assessment, and analysis that are societally relevant and practically operable. Moreover, the role of scenarios, including how they influence decisions, is also an area that can be better understood through the early engagement of user groups, such as policymakers and climate activists. Through such early engagement, we can place scenarios within the complex ecosystem of how decisions are actually made, shifting how we conceive of scenarios from decision-making tools to decision-supporting mechanisms.

However, to ensure societal relevance and operability, inclusivity in the kinds of actors that are invited and able to contribute is a vital but challenging task (Herbig, 2024). A detailed process of actor familiarisation and engagement preparation is needed, reviewing who is overlooked and who is key to invite to formulate research questions. Though this takes time and commitment from all team members, with actors often needing to be personally engaged with and sometimes convinced of their valuable knowledge, the involvement of such actors can greatly benefit the development of a scenario space. Through involving diverse actors in the development of scenarios, different needs and perspectives can be meaningfully included in scenarios, rather than non-scientific actors merely validating scenarios that are embedded within a fairly fixed framework (Braunreiter, van Beek, Hajer, & van Vuuren, 2021). Braunreiter et al. (2021) calls for rethinking such modelling and scenario spaces in a participatory way to allow for an increase in the imagination of what is desirable, plausible, and equitable in projected futures.

Through embracing a transdisciplinary approach in scenario development, the key principle of plurality (S. Briers & Vienni Baptista, Forthcoming) becomes central in rethinking the nature of the majorly used and accepted scenarios, embracing diverse perspectives that sit side by side in "diverse unison" (Bornemann & Christen, 2020). Such a process can pay off in ensuring that co-created research questions are pushing the boundaries of innovation in scenario development, their use, and the consolidation of an integrated scenario space, rather than reproducing questions that fit neatly into existing scenario frameworks.

3.4 Scenario assessment in limited scenario spaces

We have demonstrated that the hypothetical scenario space is poorly characterised with the current generation of scenarios, and that we cannot know whether our current sampling approach using the literature or databases is sufficient to provide a meaningful statistical representation of the full scenario space. We expect that this will remain the case in the future, simply because scenarios are not a random sampling of physical properties, but rather contemporary snapshots of areas of interest in an ever-evolving research landscape. Thus, given that any sample of the scenario space is unlikely to be representative, how do we assess scenarios in the context of an IPCC or other context/assessment to provide policy-relevant information?

It is worth reflecting on how the use of scenarios has evolved through the IPCC assessment reports. After an initial set of early scenarios were contested for assuming hitherto inexistent international conventions(Girod, Wiek, Mieg, & Hulme, 2009; van Beek et al., 2020), the first official scenarios documented in the First and Second Assessment Reports focused on a broad range of baseline scenarios (IPCC, 1994). The IS92 scenario assessment (IPCC, 1994) was framed around a very targeted question "Which issues and uncertainties about key variables may critically affect future net greenhouse gas emissions and aerosol concentrations?". The later SRES process did not have an

explicit research question but justified that there was a need for "new emissions scenarios that assume no *additional* climate policy initiatives" (IPCC, 2000). Since the SRES, the scenario generation process has been left up to the community (Moss et al. 2010). The current practice (Fifth and Sixth Assessment Reports) is that a call for scenario submissions to a database is issued for the community(van Beek et al., 2020), and these scenarios are assessed (K. Riahi et al. 2022). The community is already preparing for IPCC AR7[8], but ultimately, the assessment of scenarios is done by the IPCC authors drawing on the available literature. Disadvantages of the scenario database approach are that scenarios are submitted that answer very different research questions, that not all scenarios in the literature are submitted, vetting and scenario selection may exclude certain modelling groups, and that some research communities may not be motivated to submit scenarios (Peters et al., 2023a).

One could wonder why the scenario space is so poorly explored, given the vibrant and influential modelling community (van Beek et al. 2022). First, the scenario database, which constitutes the most prominent line of evidence of the IPCC assessment, has a variety of barriers to submission (Peters et al., 2023a). Second, arguably, much of the support for modelling and scenario development falls to bigger research groups, potentially crowding out opportunities to diverse approaches to address the climate challenge. Much of the scenario literature related to climate mitigation stems from a core group of closely linked people and institutes, like the IPCC and IAMC (Cointe et al., 2019). Ensuring greater diversity in the research questions asked, and the scenario methods used to answer them, will likely require lowering the entrance barriers for new participants and give them stronger incentives to contribute. Nevertheless, even with broad participation from diverse communities, the scenario literature (and databases) will continue to be dominated by questions that are relevant to the research community, and not the broader scenario space.

We argue that it is necessary to refocus scenario assessments around key research questions. It is the research questions that should determine socioeconomic assumptions, model selection, vetting criteria, and ultimately scenario selection. However, the current practice, where all submitted scenarios are treated as equally valid for assessing every mitigation indicator, while seemingly widening the uncertainty space, ultimately leads to an unrefined approach. It combines scenarios that were never designed to inform the same indicators, thereby diluting the relevance and rigour of subsequent analysis. While individual studies will address specific research questions, the collection of these studies into a database no longer give answers to an identifiable and meaningful research question. Rather, the database gives a mosaic of scenarios addressing different issues. Building on the IS92 assessment (IPCC 1994), we would suggest revisiting the purpose of scenarios and the criteria they should meet to address specific research questions. While compiling scenarios into a database has merit (Peters et al. 2023), the nature of the database makes it increasingly important to select scenarios for specific questions (Guivarch et al., 2022). More prominent linking of the scenarios in a database to explicit research questions, through metadata, could be helpful in this regard.

There is also very little information on how scenarios are used. While scenarios have had an influential role in the IPCC context (Pedersen et al., 2021; van Beek, Oomen, Hajer, Pelzer, & van Vuuren, 2022), there have seemingly been few attempts to gauge how scenarios are used (Lahn, Sognnaes, & Hermansen, 2025). There remains potential, however, to better mobilise users to help frame the research questions, and thereby the tools and methods used to address those questions (Stephanie Briers & Vienni-Baptista, 2024). We recommend a much more concerted effort to determine how scenarios are used, how they may or may not influence decisions, and to identify what types of scenario analysis will be most useful to stakeholders in the future.

[8] Submission Process | Scenario Compass Initiative

4 A plausible set of qualitative scenarios

This section is primarily a description of the conceptual framework for qualitative scenarios (narratives/storylines) but will be more comprehensive in the second report of the underlying task (D5.2, in July 2026). First, we provide an overview of qualitative scenarios and their role in scenario analysis, describe the workshops so far undertaken in DIAMOND, and describe the process for the remaining foresight and co-creation workshops. The outcomes and assessment of these latter workshops will be described in the final version of this deliverable.

4.1 Background

Scenarios are often developed based on storylines or narratives, termed here as qualitative storylines. A qualitative scenario (or narrative) may be developed in a deliberative process with a cross section of interests, often trying to span different dimensions of a problem. In the climate context, the most well-known of these are the SRES (IPCC, 2000) and SSPs(O'Neill et al., 2013) . Elements (and influence) of the SRES can be clearly seen in the SSPs, indicating an overlap in the process and the participating community. The SSPs are still in use today, taking perhaps a decade to fully supersede the SRES. The motivation for the SSPs were to both replace the SRES and constitute the background of a new parallel modelling framework designed to speed up the scenario generation process (Moss et al., 2010). The SSPs are a classic scenario quadrant, spanning the challenges to mitigation and the challenges to adaptation, with four storylines outlining the extremes of each quadrant, plus a middle-of-the-road scenario (Figure 16). The SSPs are not unique, with different axes and different storylines possible. In a climate context, The NGFS scenarios are also becoming more prominent, exploring the quadrants spanning transition risks and physical risks (NGFS, 2024).

The qualitative storylines or narratives are often translated into quantified pathways. For the SSPs, this required first quantifying scenario drivers such as population and GDP, which are then used as input into the different IAMs. These IAMs then encode the different SSP narratives and policies (Shared Policy Assumptions), to develop quantified scenarios (Keywan Riahi et al., 2017). Likewise, the NGFS scenarios have been quantified, using SSP2 as the baseline socioeconomic assumption(NGFS, 2024).

Further details on qualitative and quantitative scenarios can be found in Section 3.2.

Funded by
the European Union



Figure 16. The quadrants of the Shared Socioeconomic Pathways (SSPs) spanning axes of challenges to mitigation and challenges to adaptation, with a middle of the road at the origin.

In the DIAMOND project, our goal is to explore areas of the scenario space that are poorly explored (Figure 17). Model improvements will cover areas such as labour, finance, climate impacts, circular economy, human behaviour, and land use, with a co-design process determining how best to utilise these new model capabilities. These new scenarios will nevertheless need to be situated somewhere in the scenario space. The dimensionality and axes of that novel DIAMOND scenario space will be codesigned and described in D5.2. Hypothetically, one could think of uncertainties in climate, land, and socioeconomics (SSP) spanning axes of short-term and long-term climate action (Figure 15). Currently, of this three-dimensional representation of the scenario space, few of the boxes have been explored, representing gaps in the scenario space. Through a co-design process, we will design a novel scenario space that spans a plausible set of infilled solutions to a given sparse scenario description, reflecting the intrinsic uncertainties associated with harmonisation, infilling, land use assumptions, and climate responses.

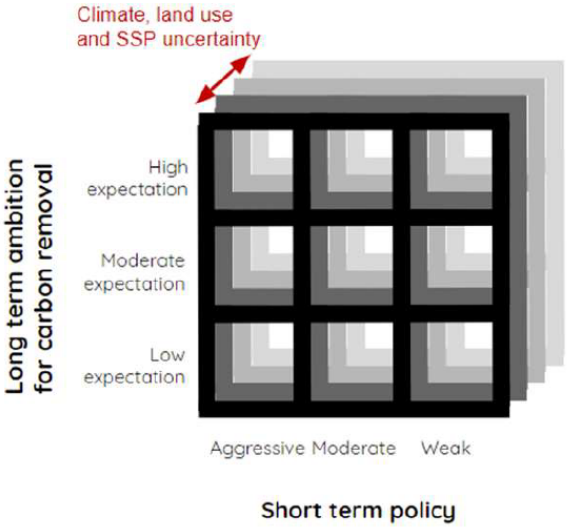


Figure 17. Hypothetical example of uncertainty in climate, land use, and socioeconomics spanning axes of short-term and long-term climate action

4.2 Co-design process in DIAMOND

The DIAMOND co-design process takes place in three distinct phases:

1. Mutual understanding: jointly framing project goals and requirements
2. Knowledge co-production: co-developing tools and knowledge
3. Model co-ownership: jointly validating and assessing quality

All three phases are in progress—with Phase 1 being at a more advanced implementation stage (see D2.1, 2.2 and 2.4)–while more details on the results of the co-creation process will be presented in upcoming WP2 deliverables and D5.2.

Insights from WP2 that are relevant to the definition of the scenario space presented in this report are mainly split into two categories: i) high-level scenario narratives drawing on the insights elicited in the foresight workshops during parts of Phase 2 (see Section 4.2.1); and ii) explicit research questions to guide the details of the scenario development during parts of Phase 1 (see Section 4.2.2). In the latter parts of Phase 2, these aspects will be synthesised to produce the finalised stakeholder insights for scenario development, with the final exchanges with WP5 expected to be reported in D5.2. Current developments along these axes are reported in the next sections.

4.2.1 Storylines based on foresight workshops

Three strategic foresight workshops feeding into the codesign process have been or will be held, with the two being of relevance to the development of the qualitative scenarios' narrative:

- 1st workshop: Requirements workshop intended to feed into the model development strategy in WP3 (see D3.2) (September 13, 2024)
- 2nd workshop: First foresight workshop (November 12, 2024)
- 3rd workshop: Second foresight workshop (early in 2025)

The first requirements workshop was conducted to co-define the requirements of improvements to the project's

IAMs, in order to address key stakeholder concerns early on in the development process. The inputs from this workshop were considered in updating the model development strategy in D3.2.

The next two foresight workshops focus on co-designing future scenario narratives with stakeholders. The first foresight workshop was conducted online on 12th Nov 2024, with participants guided through a structured process based on the Three Horizons methodology (Figure 18) to think about their desired futures without imposing any previous assumptions about economic and political feasibility. Following this framework, the transformation scenarios are set up in accordance with three conceptual pathways:

- H1: the undesirable features of the current system (a system we want to stray/transform from)
- H2: the necessary changes to shift from the current undesirable features to the desirable vision
- H3: the identification of a desirable vision of the future (a system we want to transform to).

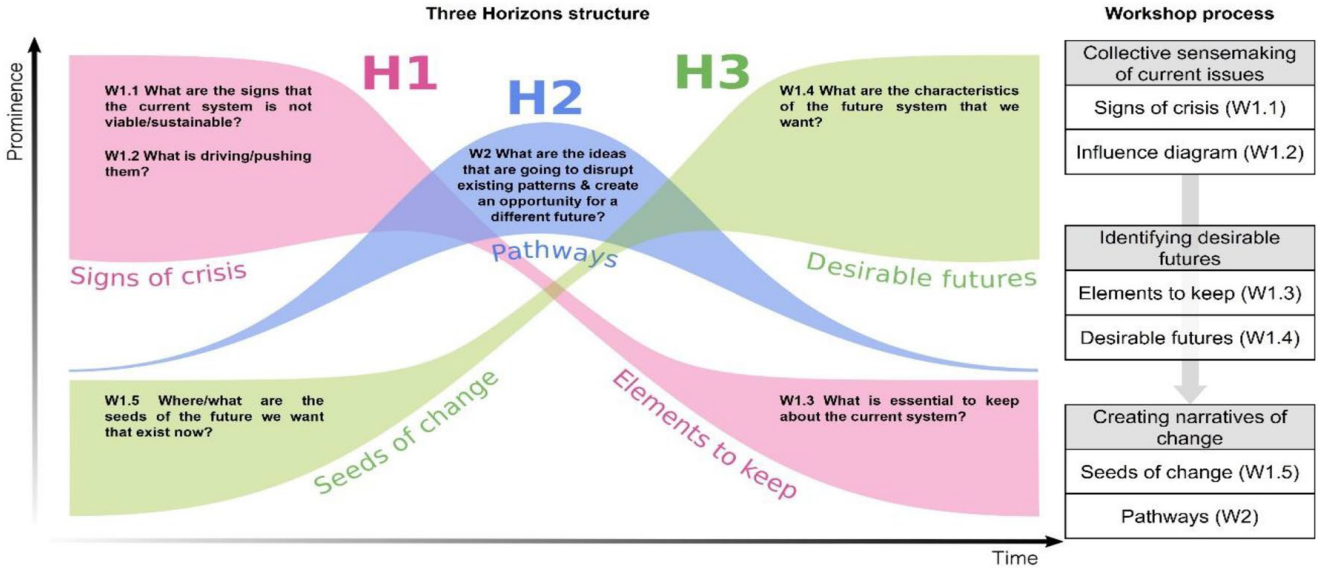


Figure 18. The Three Horizons conceptual framework used in the First Foresight Workshop. Figure Source: Adapted to the DIAMOND workflow based on Schaal, Mitchell, Scheele, Ryan, and Hanspach (2023).

In particular, within this workshop, stakeholders focused on two pathways: H1, which identifies undesirable aspects of the current system, and H3, which outlines a desirable future vision. Pathway H2, detailing the necessary changes to achieve this vision, will be discussed in the final foresight workshop. The discussions have so far been structured around several themes, including power and upstream, water, land use, biodiversity and food, residential and transport system, industry and circular economy, as well as human behaviour and household choices, with the potential to influence the scenario design process. Detailed findings from this process are reported in MS6.

4.2.2 Co-creating research questions for scenario development

To extend universally applied scenarios, such as the SSPs (van Ruijven et al., 2013), to more specific storylines that can be more useful in decision-making contexts, we are fostering a transdisciplinary approach to co-create a set of research questions that will guide scenario development. Along with the aim of creating scenarios that are more useful for decision-makers, a transdisciplinary approach also addresses potential research bias in global generic scenarios, which—despite intentions to make them policy-relevant—are primarily used by researchers and assessment communities (O'Neill et al., 2017).

To develop research questions that drive scenario development, the WP2 co-design process had two internal

stages as preparation for involving external actors in co-creating research questions.

Stage 1 focused on ensuring that the DIAMOND project team has reached a mutual understanding of goals, areas of interest, and model limitations. This stage is important to understand what is possible and achievable within the project boundaries, to avoid fostering expectations that will eventually end up unmet or mismatched. This stage was completed in June 2023 and reported in D2.4.

Stage 2 involved facilitating preparations for stakeholder engagement activities (on top of the activities for the organisation of the foresight workshops), where the DIAMOND teams went through several internal workshops, first focusing on different degrees of stakeholder engagement, different stakeholder types, and the various methods of engagement. The next series of internal workshops focused on refining the information we wanted to present to the stakeholders to set the stage for their collaboration on co-creating research questions. In a collaborative writing exercise and one in-person workshop, DIAMOND partners highlighted the topics of interest and indicated whether the IAMs in DIAMOND are able to deal with these topics.

The results were analysed using a thematic analysis approach (Braun & Clarke, 2012) and further refined by a core group in DIAMOND, collectively selected to assist in the stakeholder engagement within WP2.

The themes and related topics were divided into three groups, namely:

Group 1: Climate change

- Climate Impacts and Adaptation: Exploring the dimensions of climate impacts on human systems (e.g., equity, labour productivity) and earth systems (e.g., land, water)
- Land Use and Forestry: Nature-based solutions, bio-feedstock trade, and supply-demand balance

Group 2: Climate Action

- SDGs: Expanding the sustainable development goal (SDG) target space and linking climate action with broader sustainability objectives
- Disruptive Events: Systematic assessment of events disrupting systems and impacting progress to net-zero targets
- Circular Economy: Evaluating how circularity performance can enhance decarbonisation efforts

Group 3: Novel scenario space fuelled with user-driven scenarios

- Financing the transition
- Labour: making a just transition happen
- Representing household heterogeneity and human behaviour

The final area of preparation for the core team was to find relevant actors and invite them to participate in a series of dedicated workshops towards co-creating research questions in Stage 3. Ensuring a well-balanced representation of stakeholders was a key goal and required reviewing and iterating the process of identifying and inviting/recruiting the right stakeholders.

Stage 3 consists of a kick-off webinar to introduce stakeholders to the DIAMOND project, the concept of our novel scenario space, and the structure of the upcoming workshops. The webinar took place on January 14, 2025, with approximately 53 stakeholders representing industry, civil society and policymakers, and 12 DIAMOND partners attending. Following this, each of the three Groups will hold parallel multistakeholder discussion groups to co-create research questions. These discussion groups will bring together relevant actors with expertise in a specific topic and, following a facilitated discussion process, will aim to co-create research questions—or adapt

existing research plans—in line with the perspectives of an inclusive group of stakeholders (Fry, 2021). The multi-stakeholder discussion groups will put all participants on an equal footing and aims to build trust and ownership of the activity while allowing them to share and understand different perspectives, interests, and tacit knowledge. These workshops will take place for each of the three groups between February and March 2025.

The co-created research questions will be finalised in April 2025 and reported in D2.5 ('Stakeholder-informed research questions (updated)'), in May 2025.

5 Bibliography

Al Khourdajie, A., Skea, J., & Green, R. (2024). Climate ambition, background scenario or the model? Attribution of the variance of energy-related indicators in global scenarios. *Energy and Climate Change, 5*, 100126. doi:10.1016/j.egycc.2024.100126

Alcamo, J. (2008). Chapter Six The SAS Approach: Combining Qualitative and Quantitative Knowledge in Environmental Scenarios. In J. Alcamo (Ed.), *Developments in Integrated Environmental Assessment* (Vol. 2, pp. 123-150): Elsevier.

Allen, M. R., Frame, D. J., Friedlingstein, P., Gillett, N. P., Grassi, G., Gregory, J. M., . . . Zickfeld, K. (2024). Geological Net Zero and the need for disaggregated accounting for carbon sinks. *Nature*. doi:10.1038/s41586-024-08326-8

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures, 46*, 23-40. doi:10.1016/j.futures.2012.10.003

Anderson, K., Buck, H. J., Fuhr, L., Geden, O., Peters, G. P., & Tamme, E. (2023). Controversies of carbon dioxide removal. *Nature Reviews Earth & Environment, 4*(12), 808-814. doi:10.1038/s43017-023-00493-y

Anderson, K., & Peters, G. P. (2016). The trouble with negative emissions | Science. *Science, 354*(6309), 182-183.

Andrew, R. M. (2020). A comparison of estimates of global carbon dioxide emissions from fossil carbon sources. *Earth System Science Data, 12*(2), 1437-1465. doi:10.5194/essd-12-1437-2020

Bauer, N., Bertram, C., Schultes, A., Klein, D., Luderer, G., Kriegler, E., . . . Edenhofer, O. (2020). Quantification of an efficiency–sovereignty trade-off in climate policy. *Nature, 588*(7837), 261-266. doi:10.1038/s41586-020-2982-5

Beck, S., & Mahony, M. (2017). The IPCC and the politics of anticipation. *Nature Climate Change, 7*(5), 311-313. doi:10.1038/nclimate3264

Beck, S., & Mahony, M. (2018). The politics of anticipation: the IPCC and the negative emissions technologies experience. *Global Sustainability, 1*, e8. doi:10.1017/sus.2018.7

Bernthal, F., Dowdeswell, E., Luo, J., Attard, D., Vellinga, P., & Karimanzira, R. (1990). *Climate Change: The IPCC Response Strategies (1990)*: World Meteorological Organization, United Nations Environment Program.

Bertram, C., Brutschin, E., Drouet, L., Luderer, G., van Ruijven, B., Aleluia Reis, L., . . . Riahi, K. (2024). Feasibility of peak temperature targets in light of institutional constraints. *Nature Climate Change, 14*(9), 954-960. doi:10.1038/s41558-024-02073-4

Böhringer, C., Peterson, S., Rutherford, T. F., Schneider, J., & Winkler, M. (2021). Climate policies after Paris: Pledge, Trade and Recycle. *Energy Economics, 103*. doi:10.1016/j.eneco.2021.105471

Bornemann, B., & Christen, M. (2020). Navigating between Complexity and Control in Transdisciplinary Problem Framing: Meaning Making as an Approach to Reflexive Integration. *Social Epistemology, 34*(4), 357-369. doi:10.1080/02691728.2019.1706120

Böttcher, M., Adolphsen, O., Geden, O., Könneke, J., Schenuit, F., & Thielges, S. (2024). Foresight: Multilevel Climate Policy in 2030. In: Stiftung Wissenschaft und Politik (SWP).

Braun, V., & Clarke, V. (2012). Thematic analysis. In Harris M. Cooper, Paul Marc Camic, & D. L. Long (Eds.), *APA handbook of research methods in psychology*. (pp. 57-71). Washington, DC: American Psychological Association

Braunreiter, L., van Beek, L., Hajer, M., & van Vuuren, D. (2021). Transformative pathways – Using integrated assessment models more effectively to open up plausible and desirable low-carbon futures. *Energy Research & Social Science, 80*, 102220. doi:10.1016/j.erss.2021.102220

Briers, S., & Vienni-Baptista, B. (2024). D2.4 – Stakeholder-informed research questions. In: DIAMOND.

Briers, S., & Vienni Baptista, B. (Forthcoming). Translating and contextualising key transdisciplinary principles: The case of an integrated assessment modelling project.

Calcaterra, M., Aleluia Reis, L., Fragkos, P., Briera, T., de Boer, H. S., Egli, F., . . . Tavoni, M. (2024). Reducing the cost of capital to finance the energy transition in developing countries. *Nat Energy, 9*(10), 1241-1251. doi:10.1038/s41560-024-01606-7

Clarke, L., Jiang, K., Akimoto, K., Babiker, M., Blanford, G., Fisher-Vanden, K., . . . van Vuuren, D. P. (2014). Assessing Transformation Pathways. In O. Edenhofer, T. Zwickel, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, V. von

Stechow, & J. C. Minx (Eds.), *IPCC, 2014: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York, NY, USA: Cambridge University Press.

Cointe, B. (2024). The AR6 Scenario Explorer and the history of IPCC Scenarios Databases: evolutions and challenges for transparency, pluralism and policy-relevance. *npj Climate Action, 3*(1). doi:10.1038/s44168-023-00075-0

Cointe, B., Cassen, C., & Nadaï, A. (2019). Organising Policy-Relevant Knowledge for Climate Action: Integrated Assessment Modelling, the IPCC, and the Emergence of a Collective Expertise on Socioeconomic Emission Scenarios. *Science & Technology Studies, 32*(4), 36-57. doi:10.23987/sts.65031

Dekker, M. M., Daioglou, V., Pietzcker, R., Rodrigues, R., de Boer, H.-S., Dalla Longa, F., . . . van Vuuren, D. (2023). Identifying energy model fingerprints in mitigation scenarios. *Nature Energy, 8*(12), 1395-1404. doi:10.1038/s41560-023-01399-1

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development, 9*(5), 1937-1958. doi:10.5194/gmd-9-1937-2016

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., . . . Zheng, B. (2023). Global Carbon Budget 2023. *15*(12), 5301-5369. doi:10.5194/essd-15-5301-2023

Fry, P. (2021). Multi-stakeholder discussion group. Retrieved from https://naturalsciences.ch/co-producing-knowledge-explained/methods/td-net_toolbox/multi_stakeholder_discussion_group doi:10.5281/zenodo.4627090

Fuhrman, J., Speizer, S., O'Rourke, P., Peters, G. P., McJeon, H., Monteith, S., . . . Wang, F. M. (2024). Ambitious efforts on residual emissions can reduce CO2 removal and lower peak temperatures in a net-zero future. *Environmental Research Letters, 19*(6). doi:10.1088/1748-9326/ad456d

Fuss, S., Canadell, J. G., Peters, G. P., Tavoni, M., Andrew, R. M., Ciais, P., . . . Yamagata, Y. (2014). Betting on negative emissions. *Nature Climate Change, 4*(10), 850-853. doi:10.1038/nclimate2392

Galende-Sánchez, E., & Sorman, A. H. (2021). From consultation toward co-production in science and policy: A critical systematic review of participatory climate and energy initiatives. *Energy Research & Social Science, 73*, 101907. doi:10.1016/j.erss.2020.101907

Gambhir, A., Ganguly, G., & Mittal, S. (2022). Climate change mitigation scenario databases should incorporate more non-IAM pathways. *Joule, 6*(12), 2663-2667. doi:10.1016/j.joule.2022.11.007

Giarola, S., Chiani, L., Drouet, L., Marangoni, G., Nappo, F., Muttarak, R., & Tavoni, M. (2024). Underestimating demographic uncertainties in the synthesis process of the IPCC. *npj Climate Action, 3*(1), 1-20. doi:10.1038/s44168-024-00152-y

Giarola, S., Mittal, S., Vielle, M., Perdana, S., Campagnolo, L., Delpiazzo, E., . . . van de Ven, D. J. (2021). Challenges in the harmonisation of global integrated assessment models: A comprehensive methodology to reduce model response heterogeneity. *Sci Total Environ, 783*, 146861. doi:10.1016/j.scitotenv.2021.146861

Gidden, M. J., Brutschin, E., Ganti, G., Unlu, G., Zakeri, B., Fricko, O., . . . Riahi, K. (2023). Fairness and feasibility in deep mitigation pathways with novel carbon dioxide removal considering institutional capacity to mitigate. *Environmental Research Letters, 18*(7). doi:10.1088/1748-9326/acd8d5

Gidden, M. J., Fujimori, S., van den Berg, M., Klein, D., Smith, S. J., van Vuuren, D. P., & Riahi, K. (2018). A methodology and implementation of automated emissions harmonization for use in Integrated Assessment Models. *Environmental Modelling & Software, 105*, 187-200. doi:10.1016/j.envsoft.2018.04.002

Gidden, M. J., Gasser, T., Grassi, G., Forsell, N., Janssens, I., Lamb, W. F., . . . Riahi, K. (2023). Aligning climate scenarios to emissions inventories shifts global benchmarks. *Nature, 624*(7990), 102-108. doi:10.1038/s41586-023-06724-y

Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., . . . Takahashi, K. (2019). Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific Model Development, 12*(4), 1443-1475. doi:10.5194/gmd-12-1443-2019

Girod, B., Wiek, A., Mieg, H., & Hulme, M. (2009). The evolution of the IPCC's emissions scenarios. *Environmental Science & Policy, 12*(2), 103-118. doi:10.1016/j.envsci.2008.12.006

Grant, N., Hawkes, A., Mittal, S., & Gambhir, A. (2021). The policy implications of an uncertain carbon dioxide

removal potential. *Joule, 5*(10), 2593-2605. doi:10.1016/j.joule.2021.09.004
Grassi, G., House, J., Dentener, F., Federici, S., den Elzen, M., & Penman, J. (2017). The key role of forests in meeting climate targets requires science for credible mitigation. *Nature Climate Change, 7*(3), 220-226. doi:10.1038/nclimate3227
Grassi, G., House, J., Kurz, W. A., Cescatti, A., Houghton, R. A., Peters, G. P., . . . Zaehle, S. (2018). Reconciling global-model estimates and country reporting of anthropogenic forest CO2 sinks. *Nature Climate Change, 8*(10), 914-920. doi:10.1038/s41558-018-0283-x
Green, F., Bois von Kursk, O., Muttitt, G., & Pye, S. (2024). No new fossil fuel projects: The norm we need. *Science, 384*(6699), 954-957. doi:10.1126/science.adn6533
Grubler, A., Wilson, C., Bento, N., Boza-Kiss, B., Krey, V., McCollum, D. L., . . . Valin, H. (2018). A low energy demand scenario for meeting the 1.5 °C target and sustainable development goals without negative emission technologies. *Nature Energy, 3*(6), 515-527. doi:10.1038/s41560-018-0172-6
Guivarch, C., Le Gallic, T., Bauer, N., Fragkos, P., Huppmann, D., Jaxa-Rozen, M., . . . Wagner, F. (2022). Using large ensembles of climate change mitigation scenarios for robust insights. *Nature Climate Change, 12*(5), 428-435. doi:10.1038/s41558-022-01349-x
Hahnel, U. J. J., Chatelain, G., Conte, B., Piana, V., & Brosch, T. (2020). Mental accounting mechanisms in energy decision-making and behaviour. *Nature Energy, 5*(12), 952-958. doi:10.1038/s41560-020-00704-6
Hausfather, Z., & Peters, G. P. (2020). Emissions – the 'business as usual' story is misleading. *Nature, 577*, 618-620.
Hedelin, B., Gray, S., Woehlke, S., BenDor, T. K., Singer, A., Jordan, R., . . . Sterling, E. (2021). What's left before participatory modeling can fully support real-world environmental planning processes: A case study review. *Environmental Modelling & Software, 143*. doi:10.1016/j.envsoft.2021.105073
Herbig, V. (2024). *From stakeholder engagement to inclusivity: Advancing participatory modeling for net-zero sustainable development: A case study of the DIAMOND Project.* (Master), Swiss Federal Institute of Technology (ETH) Zurich.
Herbig, V., Briers, S., & Vienni-Baptista, B. (forthcoming). From Consultation to Inclusive Engagement: Advancing Participatory Modeling for Net-Zero Sustainable Development.
Hermansen, E. A. T., Boasson, E. L., & Peters, G. P. (2023). Climate action post-Paris: how can the IPCC stay relevant? *npj Climate Action, 2*(1), 1-8. doi:10.1038/s44168-023-00058-1
Hickel, J., Brockway, P., Kallis, G., Keyßer, L., Lenzen, M., Slameršak, A., . . . Ürge-Vorsatz, D. (2021). Urgent need for post-growth climate mitigation scenarios. *Nature Energy, 6*(8), 766-768. doi:10.1038/s41560-021-00884-9
Hoffmann, S., Deutsch, L., Klein, J. T., & O'Rourke, M. (2022). Integrate the integrators! A call for establishing academic careers for integration experts. *Humanities and Social Sciences Communications, 9*(1). doi:10.1057/s41599-022-01138-z
Hong, H., Karolyi, G. A., & Scheinkman, J. A. (2020). Climate Finance. *The Review of Financial Studies, 33*(3), 1011-1023. doi:10.1093/rfs/hhz146
Houghton, J. T., Callander, B. A., & Varney, S. K. (1992). *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment*: Cambridge University Press, Cambridge, Great Britain, New York, NY, USA, and Victoria , Australia.
Huntington, H. G., Weyant, J. P., & Sweeney, J. L. (1982). Modeling for insights, not numbers: the experiences of the energy modeling forum. *Omega, 10*(5), 449-462. doi:10.1016/0305-0483(82)90002-0
Huppmann, D., Kriegler, E., Krey, V., Riahi, K., Rogelj, J., Rose, S. K., . . . Zhang, R. (2018). IAMC 1.5°C Scenario Explorer and Data hosted by IIASA (Publication no. 10.22022/SR15/08-2018.15429). from Integrated Assessment Modeling Consortium & International Institute for Applied Systems Analysis https://data.ene.iiasa.ac.at/iamc-sr15-explorer
Huppmann, D., Rogelj, J., Kriegler, E., Krey, V., & Riahi, K. (2018). A new scenario resource for integrated 1.5 °C research. *Nature Climate Change, 8*(12), 1027-1030. doi:10.1038/s41558-018-0317-4
IEA. (2024). *World Energy Outlook 2024*: Inernational Energy Agency.
IPCC. (1994). *Climate Change 1994: Radiative Forcing of Climate Change and An Evaluation of the IPCC IS92 Emission Scenarios*. Cambridge: Cambridge University Press.
IPCC. (2000). *Special Report on Emission Scenarios (SRES)*. Cambridge: Cambridge University Press.
IPCC. (2022a). Annex III: Scenarios and Modelling Methods. In *Climate Change 2022 - Mitigation of Climate Change*

(pp. 1841-1908).

Ipcc. (2022b). *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (P. R. Shukla, J. Skea, R. Slade, A. A. Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, & J. Malley Eds.). Cambridge, UK and New York, NY, USA: Cambridge University Press.

Jahn, T., Bergmann, M., & Keil, F. (2012). Transdisciplinarity: Between mainstreaming and marginalization. *Ecological Economics, 79*, 1-10. doi:10.1016/j.ecolecon.2012.04.017

Jaxa-Rozen, M., & Trutnevyte, E. (2021). Sources of uncertainty in long-term global scenarios of solar photovoltaic technology. *Nature Climate Change, 11*(3), 266-273. doi:10.1038/s41558-021-00998-8

Kanitkar, T., Mythri, A., & Jayaraman, T. (2024). Equity assessment of global mitigation pathways in the IPCC Sixth Assessment Report. *Climate Policy*.

Kazlou, T., Cherp, A., & Jewell, J. (2024). Feasible deployment of carbon capture and storage and the requirements of climate targets. *Nature Climate Change, 14*(10), 1047-1055. doi:10.1038/s41558-024-02104-0

Keepin, B., & Wynne, B. (1984). Technical analysis of IIASA energy scenarios. *Nature, 312*(5996), 691-695. doi:10.1038/312691a0

Kikstra, J. S., Li, M., Brockway, P. E., Hickel, J., Keysser, L., Malik, A., . . . Lenzen, M. (2024). Downscaling down under: towards degrowth in integrated assessment models. *Economic Systems Research, 36*(4), 576-606. doi:10.1080/09535314.2023.2301443

Kikstra, J. S., Nicholls, Z. R. J., Smith, C. J., Lewis, J., Lamboll, R. D., Byers, E., . . . Riahi, K. (2022). The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways: from emissions to global temperatures. *Geoscientific Model Development, 15*(24), 9075-9109. doi:10.5194/gmd-15-9075-2022

Koasidis, K., Koutsellis, T., Xexakis, G., Nikas, A., & Doukas, H. (2023). Understanding expectations from and capabilities of climate-economy models for measuring the impact of crises on sustainability. *Journal of Cleaner Production, 414*. doi:10.1016/j.jclepro.2023.137585

Koasidis, K., Nikas, A., & Doukas, H. (2023). Why integrated assessment models alone are insufficient to navigate us through the polycrisis. *One Earth, 6*(3), 205-209. doi:10.1016/j.oneear.2023.02.009

Korsbakken, J. I., & Peters, G. (2023). *IAM_COMPACT_D4.3_Broad_Scenario_Logic*. Retrieved from https://www.iam-compact.eu/publications/deliverables

Krey, V., Guo, F., Kolp, P., Zhou, W., Schaeffer, R., Awasthy, A., . . . van Vuuren, D. P. (2019). Looking under the hood: A comparison of techno-economic assumptions across national and global integrated assessment models. *Energy, 172*, 1254-1267. doi:10.1016/j.energy.2018.12.131

Kriegler, E., Edmonds, J., Hallegatte, S., Ebi, K. L., Kram, T., Riahi, K., . . . van Vuuren, D. P. (2014). A new scenario framework for climate change research: the concept of shared climate policy assumptions. *Climatic Change, 122*(3), 401-414. doi:10.1007/s10584-013-0971-5

Lahn, B., Sognnaes, I., & Hermansen, E. A. T. (2025). Analysing the use of emissions scenarios in practice. *npj Climate Action, (in review)*.

Lamboll, R. D., Nicholls, Z. R. J., Kikstra, J. S., Meinshausen, M., & Rogelj, J. (2020). Silicone v1.0.0: an open-source Python package for inferring missing emissions data for climate change research. *Geoscientific Model Development, 13*(11), 5259-5275. doi:10.5194/gmd-13-5259-2020

Lang, D. J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., . . . Thomas, C. J. (2012). Transdisciplinary research in sustainability science: Practice, principles, and challenges. *Sustainability Science, 7*(SUPPL. 1), 25-43. doi:10.1007/s11625-011-0149-x

Lawrence, M. G., Williams, S., Nanz, P., & Renn, O. (2022). Characteristics, potentials, and challenges of transdisciplinary research. *One Earth, 5*(1), 44-61. doi:10.1016/j.oneear.2021.12.010

McCollum, D., Nagai, Y., Riahi, K., Marangoni, G., Calvin, K., Pietzcker, R., . . . VAN DER ZWAAN, B. (2013). Energy Investments under climate policy: A comparison of Glovbal Models. *Climate Change Economics, 04*(04), 1340010. doi:10.1142/s2010007813400101

McCollum, D. L., Gambhir, A., Rogelj, J., & Wilson, C. (2020). Energy modellers should explore extremes more systematically in scenarios. *Nature Energy, 5*(2), 104-107. doi:10.1038/s41560-020-0555-3

McGookin, C., Süsser, D., Xexakis, G., Trutnevyte, E., McDowall, W., Nikas, A., . . . Ó Gallachóir, B. (2024). Advancing participatory energy systems modelling. *Energy Strategy Reviews, 52*. doi:10.1016/j.esr.2024.101319

Meadows, D. (1981). A Critique of the IIASA Energy Models. *The Energy Journal, 2*(3), 17-28.

Moallemi, E. A., de Haan, F. J., Hadjikakou, M., Khatami, S., Malekpour, S., Smajgl, A., . . . Bryan, B. A. (2021). Evaluating Participatory Modeling Methods for Co-creating Pathways to Sustainability. *Earth's Future, 9*(3). doi:10.1029/2020ef001843

Morris, J., Hone, D., Haigh, M., Sokolov, A., & Paltsev, S. (2023). Future energy: in search of a scenario reflecting current and future pressures and trends. *Environmental Economics and Policy Studies, 25*(1), 31-61. doi:10.1007/s10018-021-00339-1

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., . . . Wilbanks, T. J. (2010). The next generation of scenarios for climate change research and assessment. *Nature, 463*(7282), 747-756. doi:10.1038/nature08823

Muttitt, G., Price, J., Pye, S., & Welsby, D. (2023). Socio-political feasibility of coal power phase-out and its role in mitigation pathways. *Nature Climate Change, 13*(2), 140-147. doi:10.1038/s41558-022-01576-2

NGFS. (2024). NGFS long-term scenarios for central banks and supervisors. In: Network for Greening the Financial System.

Nikas, A., Elia, A., Boitier, B., Koasidis, K., Doukas, H., Cassetti, G., . . . Chiodi, A. (2021). Where is the EU headed given its current climate policy? A stakeholder-driven model inter-comparison. *Sci Total Environ, 793*, 148549. doi:10.1016/j.scitotenv.2021.148549

Nikas, A., Frilingou, N., Al Khourdajie, A., & Gambhir, A. (2024). Science must up its game to support climate finance negotiations. *Nature, 636*(8041), 45-45. doi:10.1038/d41586-024-03945-7

Nikas, A., Lieu, J., Sorman, A., Gambhir, A., Turhan, E., Vienni-Baptista, B., & Doukas, H. (2020). The desirability of transitions in demand: Incorporating behavioural and societal transformations into energy modelling. *Energy Research & Social Science, 70*, 101780. doi:10.1016/j.erss.2020.101780

Nikas, A., Xexakis, G., Koasidis, K., Acosta-Fernández, J., Arto, I., Calzadilla, A., . . . Doukas, H. (2022). Coupling circularity performance and climate action: From disciplinary silos to transdisciplinary modelling science. *Sustainable Production and Consumption, 30*, 269-277. doi:10.1016/j.spc.2021.12.011

Norström, A. V., Cvitanovic, C., Löf, M. F., West, S., Wyborn, C., Balvanera, P., . . . Österblom, H. (2020). Principles for knowledge co-production in sustainability research. *Nature Sustainability, 3*(3), 182-190. doi:10.1038/s41893-019-0448-2

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., . . . Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development, 9*(9), 3461-3482. doi:10.5194/gmd-9-3461-2016

O'Neill, B. C., Carter, T. R., Ebi, K., Harrison, P. A., Kemp-Benedict, E., Kok, K., . . . Pichs-Madruga, R. (2020). Achievements and needs for the climate change scenario framework. *Nature Climate Change, 10*(12), 1074-1084. doi:10.1038/s41558-020-00952-0

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., . . . Solecki, W. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change, 42*, 169-180. doi:10.1016/j.gloenvcha.2015.01.004

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., . . . van Vuuren, D. P. (2013). A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change, 122*(3), 387-400. doi:10.1007/s10584-013-0905-2

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

Pedersen, J. S. T., Duarte Santos, F., van Vuuren, D., Gupta, J., Encarnação Coelho, R., Aparício, B. A., & Swart, R. (2021). An assessment of the performance of scenarios against historical global emissions for IPCC reports. *Global Environmental Change, 66*, 102199. doi:10.1016/j.gloenvcha.2020.102199

Pedersen, J. S. T., van Vuuren, D., Gupta, J., Santos, F. D., Edmonds, J., & Swart, R. (2022). IPCC emission scenarios: How did critiques affect their quality and relevance 1990–2022? *Global Environmental Change, 75*, 102538. doi:10.1016/j.gloenvcha.2022.102538

Peng, W., Iyer, G., Bosetti, V., Chaturvedi, V., Edmonds, J., Fawcett, A. A., . . . Weyant, J. (2021). Climate policy models need to get real about people — here's how. *Nature, 594*(7862), 174-176. doi:10.1038/d41586-021-01500-2

Peters, G. P., Al Khourdajie, A., Sognnaes, I., & Sanderson, B. M. (2023a). AR6 scenarios database: an assessment of current practices and future recommendations. *npj Climate Action, 2*(1), 1-6. doi:10.1038/s44168-023-

00050-9
Peters, G. P., Al Khourdajie, A., Sognnaes, I., & Sanderson, B. M. (2023b). AR6 scenarios database: an assessment of current practices and future recommendations. *npj Climate Action, 2*(1). doi:10.1038/s44168-023-00050-9
Peters, G. P., Andrew, R. M., Boden, T., Canadell, J. G., Ciais, P., Le Quéré, C., . . . Wilson, C. (2013). The challenge to keep global warming below 2 °C. *Nature Climate Change, 3*(1), 4-6. doi:10.1038/nclimate1783
Pirani, A., Fuglestvedt, J. S., Byers, E., O'Neill, B., Riahi, K., Lee, J.-Y., . . . Tebaldi, C. (2024). Scenarios in IPCC assessments: lessons from AR6 and opportunities for AR7. *npj Climate Action, 3*(1). doi:10.1038/s44168-023-00082-1
Pye, S., Broad, O., Bataille, C., Brockway, P., Daly, H. E., Freeman, R., . . . Watson, J. (2021). Modelling net-zero emissions energy systems requires a change in approach. *Climate Policy, 21*(2), 222-231. doi:10.1080/14693062.2020.1824891
Ram, M., Osorio-Aravena, J. C., Aghahosseini, A., Bogdanov, D., & Breyer, C. (2022). Job creation during a climate compliant global energy transition across the power, heat, transport, and desalination sectors by 2050. *Energy, 238*, 121690. doi:10.1016/j.energy.2021.121690
Ranjan, A., Kanitkar, T., & Jayaraman, T. A new scenario framework for equitable and climate-compatible futures. *Climate and Development*, 1-13. doi:10.1080/17565529.2024.2365939
Riahi, K., Bertram, C., Huppmann, D., Rogelj, J., Bosetti, V., Cabardos, A.-M., . . . Zakeri, B. (2021). Cost and attainability of meeting stringent climate targets without overshoot. *Nature Climate Change, 11*(12), 1063-1069. doi:10.1038/s41558-021-01215-2
Riahi, K., Schaeffer, R., Arango, J., Calvin, K., Guivarch, C., Hasegawa, T., . . . Van Vuuren, D. P. (2022). Mitigation pathways compatible with long-term goals. In. J. S. P.R. Shukla, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley (Series Ed.), In IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK and New York, NY, USA: Cambridge University Press, . doi:10.1017/9781009157926.005
Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., . . . Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change, 42*, 153-168. doi:10.1016/j.gloenvcha.2016.05.009
Ritchie, J., & Dowlatabadi, H. (2017). Why do climate change scenarios return to coal? *Energy, 140*, 1276-1291. doi:10.1016/j.energy.2017.08.083
Roe, S., Streck, C., Obersteiner, M., Frank, S., Griscom, B., Drouet, L., . . . Lawrence, D. (2019). Contribution of the land sector to a 1.5 °C world. *Nature Climate Change, 9*(11), 817-828. doi:10.1038/s41558-019-0591-9
Roelfsema, M., van Soest, H. L., Harmsen, M., van Vuuren, D. P., Bertram, C., den Elzen, M., . . . Vishwanathan, S. S. (2020). Taking stock of national climate policies to evaluate implementation of the Paris Agreement. *Nature Communications, 11*(1), 2096. doi:10.1038/s41467-020-15414-6
Rogelj, J., Popp, A., Calvin, K. V., Luderer, G., Emmerling, J., Gernaat, D., . . . Tavoni, M. (2018). Scenarios towards limiting global mean temperature increase below 1.5 °C. *Nature Climate Change, 8*(4), 325-332. doi:10.1038/s41558-018-0091-3
Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., . . . Vilariño, M. V. (2018). Mitigation pathways compatible with 1.5°C in the context of sustainable development. In *Special Report on the impacts of global warming of 1.5 °C*. Geneva: Intergovernmental Panel on Climate Change.
Sanderson, B. M., Brovkin, V., Fisher, R., Hohn, D., Ilyina, T., Jones, C., . . . Ziehn, T. doi:10.5194/egusphere-2024-3356
Sandstad, M., Aamaas, B., Johansen, A. N., Lund, M. T., Peters, G. P., Samset, B. H., . . . Skeie, R. B. (2024). CICERO Simple Climate Model (CICERO-SCM v1.1.1) – an improved simple climate model with a parameter calibration tool. *Geoscientific Model Development, 17*(17), 6589-6625. doi:10.5194/gmd-17-6589-2024
Schaal, T., Mitchell, M., Scheele, B. C., Ryan, P., & Hanspach, J. (2023). Using the three horizons approach to explore pathways towards positive futures for agricultural landscapes with rich biodiversity. *Sustainability Science, 18*(3), 1271-1289. doi:10.1007/s11625-022-01275-z
Schaeffer, M., Gohar, L., Kriegler, E., Lowe, J., Riahi, K., & van Vuuren, D. (2015). Mid- and long-term climate projections for fragmented and delayed-action scenarios. *Technological Forecasting and Social Change,*

90, 257-268. doi:10.1016/j.techfore.2013.09.013

Schäfer, M., & Kröger, M. (2016). Joint problem framing in sustainable land use research: Experience with Constellation Analysis as a method for inter- and transdisciplinary knowledge integration. *Land Use Policy, 57*, 526-539. doi:10.1016/j.landusepol.2016.06.013

Schweizer, V. J., & O'Neill, B. C. (2013). Systematic construction of global socioeconomic pathways using internally consistent element combinations. *Climatic Change, 122*(3), 431-445. doi:10.1007/s10584-013-0908-z

Schweizer, V. J., & O'Neill, B. C. (2014). Systematic construction of global socioeconomic pathways using internally consistent element combinations. *Climatic Change, 122*(3), 431-445. doi:10.1007/s10584-013-0908-z

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

Skea, J., van Diemen, R., Portugal-Pereira, J., & Khourdajie, A. A. (2021). Outlooks, explorations and normative scenarios: Approaches to global energy futures compared. *Technological Forecasting and Social Change, 168*. doi:10.1016/j.techfore.2021.120736

Skeie, R. B., Peters, G. P., Fuglestvedt, J., & Andrew, R. (2021). A future perspective of historical contributions to climate change. *Climatic Change, 164*(1), 24. doi:10.1007/s10584-021-02982-9

Sognnaes, I., Gambhir, A., van de Ven, D.-J., Nikas, A., Anger-Kraavi, A., Bui, H., . . . Peters, G. P. (2021). A multi-model analysis of long-term emissions and warming implications of current mitigation efforts. *Nature Climate Change, 11*(12), 1055-1062. doi:10.1038/s41558-021-01206-3

Sognnaes, I., & Peters, G. (2025). Influence of individual models and studies on quantitative mitigation findings in the IPCC Sixth Assessment Report. Retrieved from https://www.researchsquare.com/article/rs-5698716/v1 doi:https://doi.org/10.21203/rs.3.rs-5698716/v1

Sovacool, B. K., Ali, S. H., Bazilian, M., Radley, B., Nemery, B., Okatz, J., & Mulvaney, D. (2020). Sustainable minerals and metals for a low-carbon future. *Science, 367*(6473), 30-33. doi:10.1126/science.aaz6003

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

van Beek, L., Oomen, J., Hajer, M., Pelzer, P., & van Vuuren, D. (2022). Navigating the political: An analysis of political calibration of integrated assessment modelling in light of the 1.5 °C goal. *Environmental Science & Policy, 133*, 193-202. doi:10.1016/j.envsci.2022.03.024

van de Ven, D.-J., Mittal, S., Gambhir, A., Lamboll, R. D., Doukas, H., Giarola, S., . . . Nikas, A. (2023). A multimodel analysis of post-Glasgow climate targets and feasibility challenges. *Nature Climate Change, 13*(6), 570-578. doi:10.1038/s41558-023-01661-0

van de Ven, D. J., Mittal, S., Nikas, A., Xexakis, G., Gambhir, A., Hermwille, L., . . . Peters, G. P. (2025). Energy and socioeconomic system transformation through a decade of IPCC-assessed scenarios. *Nature Climate Change*. doi:10.1038/s41558-024-02198-6

van Maanen, N., Lissner, T., Harmsen, M., Piontek, F., Andrijevic, M., & van Vuuren, D. P. (2023). Representation of adaptation in quantitative climate assessments. *Nature Climate Change, 13*(4), 309-311. doi:10.1038/s41558-023-01644-1

van Ruijven, B. J. (2016). Mind the gap – the case for medium level emission scenarios. *Climatic Change, 138*(3), 361-367. doi:10.1007/s10584-016-1744-8

van Ruijven, B. J., Levy, M. A., Agrawal, A., Biermann, F., Birkmann, J., Carter, T. R., . . . Schweizer, V. J. (2013). Enhancing the relevance of Shared Socioeconomic Pathways for climate change impacts, adaptation and vulnerability research. *Climatic Change, 122*(3), 481-494. doi:10.1007/s10584-013-0931-0

van Soest, H. L., Aleluia Reis, L., Baptista, L. B., Bertram, C., Després, J., Drouet, L., . . . van Vuuren, D. P. (2021). Global roll-out of comprehensive policy measures may aid in bridging emissions gap. *Nature Communications, 12*(1), 6419. doi:10.1038/s41467-021-26595-z

van Vuuren, D. P., & Carter, T. R. (2013). Climate and socio-economic scenarios for climate change research and assessment: reconciling the new with the old. *Climatic Change, 122*(3), 415-429. doi:10.1007/s10584-013-0974-2

van Vuuren, D. P., Hof, A. F., van Sluisveld, M. A. E., & Riahi, K. (2017). Open discussion of negative emissions is urgently needed. *Nature Energy, 2*(12), 902-904. doi:10.1038/s41560-017-0055-2

van Vuuren, D. P., Stehfest, E., Gernaat, D. E. H. J., van den Berg, M., Bijl, D. L., de Boer, H. S., . . . van Sluisveld, M. A. E. (2018). Alternative pathways to the 1.5 °C target reduce the need for negative emission technologies. *Nature Climate Change, 8*(5), 391-397. doi:10.1038/s41558-018-0119-8

Vandyck, T., Keramidas, K., Saveyn, B., Kitous, A., & Vrontisi, Z. (2016). A global stocktake of the Paris pledges: Implications for energy systems and economy. *Global Environmental Change, 41*, 46-63. doi:10.1016/j.gloenvcha.2016.08.006

Vienni-Baptista, B., Fletcher, I., Lyall, C., & Pohl, C. (2022). Embracing heterogeneity: Why plural understandings strengthen interdisciplinarity and transdisciplinarity. *Science and Public Policy, 49*(6), 865-877. doi:10.1093/scipol/scac034

Walker Wood, T., Richter, K., & Atkins, E. (2024). Modelling beyond growth perspectives for sustainable climate futures: The case for rethinking Shared Socioeconomic Pathways. *Energy Research & Social Science, 117*. doi:10.1016/j.erss.2024.103705

Wilbanks, T. J., & Ebi, K. L. (2014). SSPs from an impact and adaptation perspective. *Climatic Change, 122*(3), 473-479. doi:10.1007/s10584-013-0903-4