Studies in History and Philosophy of Science 117 (2026) 102149

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

Design principles for the IPCC emission scenario ensemble

Francesco Nappo

Dipartimento di Filosofia, Università Cattolica del Sacro Cuore, Largo Gemelli, 1, 20123, Milano, MI, Italy

ARTICLE INFO

Keywords:
Climate science
Scenarios
Multi-model ensembles
Uncertainty
Science-policy

ABSTRACT

Working Group III (WGIII) of the Intergovernmental Panel on Climate Change (IPCC) regularly assesses collections of emission scenarios submitted by the scientific community, linking long-term projections about socio-economic conditions and technology rollout with the changing climate. Current protocols for assembling and evaluating submitted scenarios have been shaped by the IPCC's institutional position at the science-policy interface, particularly its commitment to delivering a policy-neutral scientific assessment. While these protocols allow WGIII experts to address an important array of questions about climate change mitigation, they limit opportunities for drawing policy-relevant information from scenarios in other ways. In preparation to future IPCC reports, this paper considers alternative approaches to the construction and evaluation of WGIII's ensemble. A framework is proposed for clarifying and evaluating the available options, based on three dimensions: *ownership*, *logic*, and *purpose*.

1. Introduction

An important part of Working Group III's (WGIII) role in the assessments of scientific evidence cyclically undertaken by the Intergovernmental Panel on Climate Change (IPCC) is the collection and evaluation of long-term emission scenarios produced by research teams worldwide (IPCC, 2022:1841). Assessed scenarios are model-based simulations of the future human-climate system based on observed linkages between economic development, greenhouse gas (GHG) emissions, and global warming (Weyant, 2017). The integration of physical, technological, and socio-economic dimensions achieved by these tools aims to address crucial questions about the space of intervention for climate policies as well as the intensity of efforts needed to achieve globally agreed-upon mitigation goals (Kikstra et al., 2022). Are current legal frameworks and national pledges likely to limit global warming to 1.5 °C above pre-industrial levels by 2100, and in any case well below 2 °C, as set by the Paris Agreement? How would stricter mitigation policies affect not only energy supply and demand, but also land use, jobs, food security, and transportation? By contrast, how would climate inaction affect those same variables over the medium and long term?

Due to the complexity of the representational task and the long-term ambition of scenarios, the uncertainty that surrounds simulations from current models is severe: it includes not only uncertainty in many of the parameters used as input, but also in the representation of the initial conditions and of the key dynamics of the global system (Beck & Kruger, 2016). As in the physical climate sciences, a multi-model ensemble approach has become standard practice (Parker, 2010). However, WGIII's approach to assembling and reflecting on collected scenarios differs in important ways from the ensemble methods adopted by Working Group I (WGI) for physical climate simulations, which have received philosophical coverage (e.g., Jebeile & Crucifix, 2020; Katzav et al., 2012). WGIII's latest assessments adopt an *unstructured ensemble* approach, where scenarios from many different models are collected and assembled with no pre-determined design (Guivarch et al., 2021). Submissions from the scientific community are admitted to IPCC assessment based solely on their passing minimal *vetting*, whose criteria include basic informativeness (viz., covering key physical and economic indicators) and consistency with historical observations. Besides vetting, WGIII admits both submissions from individual models and from model inter-comparison projects and sets no upper limit on the number of scenarios from each model.

WGIII's ensemble methods are a worthy object of historical and philosophical interest for several reasons. From a theoretical viewpoint, they are part of a relatively unprecedented effort at synthesizing outputs of scientific models at a crucial juncture of the science-policy interface. They are, accordingly, a case study of high relevance for discussions on objectivity and science's role in policy (Havstad & Brown, 2017). On a more practical level, the evolving interests of policymakers paired with constraints on model-making generate the conditions for expectation mismatch and the need for critical evaluation (Hulme, 2022). In preparation to future ensemble exercises, it is important to build capacity to assist WGIII experts in their methodological evaluations. Echoing IPCC's

E-mail address: francesco.nappo@unicatt.it.

https://doi.org/10.1016/j.shpsa.2026.102149
Received 19 February 2025; Received in revised form 23 January 2026; Accepted 15 March 2026
Available online 20 March 2026

recent call for a "science of scenario assessment" (IPCC, 2023, p. 4), this paper addresses WGIII's choice of ensemble methods through philosophical tools. One of the aims is to provide a reconstruction of the practice – one which displays both how the aim of 'policy-neutrality' underlies the current rationale for the unstructured approach and the inherent benefits and limitations of the latter. In the way of providing this reconstruction, this paper will also extrapolate an evaluative framework to clarify the features of distinct ensemble approaches and indicate the dimensions that need to be addressed in order to robustly improve WGIII's future scenario assessments.

More in detail, the discussion is organized as follows. Section two offers a brief, non-technical overview of modeling frameworks underlying the emission scenarios that WGIII assesses. Section three moves on to describing WGIII's unstructured approach, tracing its historical evolution to a growing concern for respecting the IPCC's mandate to produce "policy-neutral" (IPCC, 2024) synthesis work. As highly influential climate policy evaluation reports, WGIII's assessments have received considerable critical attention (Pedersen et al., 2022). Unstructured protocols effectively dispense WGIII authors from making several value-laden choices in ensemble building, such as whose emission pathways to assess and which research questions are worth pursuing through scenarios. As section four shows, this minimal form of '*ownership'* of the ensemble's shape and content (to borrow a term from Parker, 2010) still allows WGIII to draw information of high policy relevance from submitted scenarios. Two main assessment methods used by WGIII will be distinguished, referred to as 'quantitative' and 'qualitative' methods, and illustrated by reference to WGIII's latest conclusions. This reconstruction will allow to characterize WGIII's unstructured approach not only in its ownership profile but also in the overall *logic* of its assessment.[1]

Having considered some of the advantages of WGIII's approach, section five will point out that unstructured protocols also limit opportunities for drawing policy-relevant information in distinctive ways. Two persisting issues are *missing variety*, viz., insufficient representation of the relevant space of uncertainty, and *missing linkages*, viz., insufficient elaboration of the regional and country-level impacts of global mitigation pathways. While it is commonplace to attribute these deficiencies to inherent limitations of climate economic models and underinvestment in modeling infrastructures (e.g., Beck & Mahony, 2019; Keysser & Lenzen, 2021; Kowarsch, 2016), these diagnoses are incomplete. Equally relevant are factors related to WGIII's ensemble construction workflow, as shaped by its unstructured design. Because unstructured protocols play a crucial role in securing WGIII's conception of policy-neutrality, what ultimately emerges is a tension between the stated aims of policy-neutrality and policy-relevance. To conceptualize this tension, section five makes the case for recognizing *purpose* as an additional dimension of evaluation of WGIII's ensemble methods alongside ownership and logic (Fig. 1). The examples provided will illustrate that the unstructured approach serves some policy purposes at the expense of others, implying that not all intended users of WGIII's assessments are equally well-served by them and that, to achieve greater balance, a reform of WGIII's unstructured approach may be necessary.

Section six will complete the discussion by reviewing proposals for reform of WGIII's approach, characterizing them along the same dimensions of *ownership*, *logic*, and *purpose* highlighted in the reconstruction. The options range from more radical proposals, which give an entirely different meaning to WGIII's notion of a 'policy-neutral' assessment, to more conservative proposals, which aim to address persisting issues while preserving the design choices that inspire current ensemble construction protocols. For convenience, the paper will focus on two specific proposals: the 'structured ensemble' (Peters et al., 2023) and the 'living database' approach (Pirani et al., 2024). While leaning on the side of reform, a point that this paper will insist on is that the choice between the options is not merely one of philosophical principles, but of facts. There are risks involved in abandoning WGIII's current unstructured approach, which any proposed reform for WGIII's ensemble assessment ought to consider. These risks derive specifically from the politically contested nature of climate economic modeling, on the one hand, and by IPCC's status as a distinctive actor at the science-policy interface, on the other. A comprehensive perspective is required, which takes into account both the epistemic and the institutional aspects of the problem of ensemble design.

2. The building blocks

The history of WGIII's assessments is intimately linked to that of Integrated Assessment Models (IAMs) that produce emission scenarios.[2] This branch of the climate sciences started out as a heterogeneous family of models, with origins in separate disciplines and modeling traditions (Cointe et al., 2021). Owing to their common ambition to inform climate policy, an intense coordinating effort was undertaken to render the models comparable (van Beek et al., 2020). The acronym 'IAM' came to refer to a variety of tools whose common feature is to incorporate a simplified climate model, where GHG emissions lead to higher temperatures, and a simplified model of socio-economic systems, which drives emissions and is affected by climate change (Weyant 2007). The latter typically consists of a core economic model to assess global production via a Cobb-Douglas function of labor, capital, and productivity, which is then linked to additional modules to capture relevant bio-physical (e.g., land use), social (e.g., energy demand) and technological detail (e.g., fuel reserves). The main empirical basis for IAMs consists in observed linkages between human activities, GHG emissions, and rising global temperature. Further support for IAM assumptions can be drawn from multiples sources: from consistency with near-term observations, to observed trends, expert judgment, and hindcasting exercises (i.e., simulations run backward in time; Wilson et al., 2021).

An IAM scenario exercise starts out from a qualitative narrative outlining a conceivable societal trend, which is then associated to quantitative inputs for the model. When provided with a consistent set of assumptions, IAMs compute detailed representations of time-evolutions of future GHG emissions and accompanying features of the target socio-economic system – an 'emission scenario' –, usually up to year 2100. These global, long-term scenarios are offered in a broader context of modeling exercises, from macro-economic outlooks that study short-term price fluctuations and income effects of global warming (OECD, 2024), to medium-term energy system and sectoral studies, that address decarbonization efforts from the bottom-up perspective of the material systems needed to implement it (Yue et al., 2018). The rationale for using IAMs for long-term projection is their accounting for feedback between the climatic, socio-economic and technical levels of the global system. Due to this specific feature, IAM scenarios were originally assigned an *explorative* use, outlining long-term emission projections for worlds that do not differ markedly in policy and socio-economic respects from observed trends. Increasingly, the same models started being used with a *normative* purpose, to draw credible and desirable societal pathways to sustainable futures, by simulating the

[1] There are broad affinities between the account developed in this paper and claims defended in Parker (2010). However, the overall project is different. Parker (2010) aims to single out *requirements* for the representation of uncertainty in ensemble assessment in the physical climate sciences. This paper is mostly interested in identifying *dimensions of evaluation* of ensemble methodology, without taking a stance over any specific requirements along these dimensions that ensemble assessments should fulfill. On the intended notion of 'ownership' used in this paper, see fn. 5.

[2] There is some inevitable vagueness over what counts as an 'IAM' (cf. Weyant, 2017). Some modeling exercises in the broad traditions of macro-economic outlooks and energy systems modeling are also capable of producing long-term global emission scenarios.

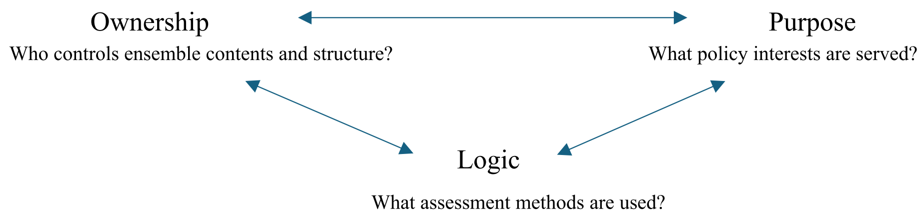

Fig. 1. Ownership, logic, and purpose as dimensions of ensemble evaluation. The arrows indicate relations of dependence among choices along each dimension (see section six).

effect of climate policies over a no-policy baseline projection (Fig. 2; Braunreiter et al., 2021). In this further role, IAMs must strive to include endogenous preferences and the heterogeneity of societal actors involved in energy transitions.

Compared to physical climate models, philosophical reception of IAMs has been mild. A body of literature takes issue with *cost-benefit* IAMs with *macroeconomic* coverage (e.g., Frisch, 2013). These models aim to compute economically optimal scenarios of future carbon emissions based on projections of macroeconomic trends together with a choice about the discount rate to apply to enjoyment of future goods (Mintz-Woo, 2021; Nordhaus, 2008). Besides often missing relevant detail at the regional and sectoral scales, however, cost-benefit IAMs suffer from large uncertainties surrounding the so-called 'damage function' that links different levels of temperature increase to climate-related damages (Diaz & Moore, 2017). Given the unprecedented rapidity of global warming, the shape of this function is highly underdetermined by historical observations (Piontek et al., 2021). However, a different class of IAMs currently informs WGIII's assessments: these are *cost-effective* models that lean squarely towards *detail-process* representation – 'detail-process IAMs' in short (Weyant, 2017). Compared to cost-benefit, these IAMs limit themselves to computing minimal-cost pathways when a policy or climate target is given, thereby sidestepping the problem of fully specifying the damages; they also offer higher resolution of biophysical and technological detail for each world region to better capture the dynamics of energy production, land use, technology rollout, and their relation to the emission and absorption cycle of GHGs.

Despite having greater accuracy than simple cost-benefit models, the uncertainties affecting projections from detail-process IAMs remain high (Tavoni & Valente, 2022). On the one hand, an emission scenario aims to serve as an archetype of possible time-evolutions of the human-climate system: not so much a future prospect but an average (so to speak) over possible evolutions. Thus, a pre-2020 scenario of growing emissions in the EU may well have matched the statistics of the evolution observed over the past six years even though the actual course of history saw a drop in emissions due to a pandemic, followed by delayed replacement of fossil fuels due to emerging regional conflict. In this respect, a key strength of scenarios is that data about key emission drivers such as population growth, urbanization and gross domestic product (GDP) are available and methodologies for projection are well-established (KC & Lutz, 2017). On the other hand, detail-process IAMs are used to simulate what would happen to the global human-climate system assuming that additional climate policies are in place (Fig. 2). Modeling the fine-grained dynamics of global energy production, technology rollout, and the emission and absorption cycle of GHGs under climate policies requires approximations and estimates regarding a variety of long-term socio-economic and technological aspects, for which any model calibration proves difficult. This results in a variety of approaches being used across 'IAM families', i.e., across types of detail-process IAMs developed by distinct research facilities worldwide.

Two dimensions standardly used for classifying current detail-process IAMs are *coverage* and *dynamics* (Harmsen et al., 2021). The former concerns the level of detail about the system's initial conditions that is included as input data and the choice of endogenous and

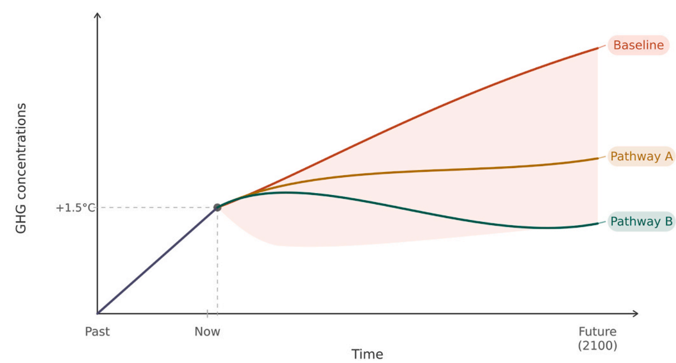

Fig. 2. IAM-based mitigation scenarios (Pathways A and B) as the policy-induced curbing of a no-policy emission baseline. Source: Carbonbrief.org.

exogenous components for the model. With regards to the economy, IAMs are typically either *partial* or *general equilibrium* models, depending on whether they work with specific sectors of the global economy and take other parts as exogenous, or rather attempt a representation of the full economic system. While in the former case climate policies are implemented as costs for the represented sectors of the economy, in the latter they are represented as global consumption or GDP losses (Harmsen et al., 2021). IAMs also differ significantly in their representation of utility, with some models assuming a single global decision-maker that implements climate policies while others countenancing a handful of decision-makers representing the utility of distinct world regions. Finally, IAMs' coverage can differ with regards to their representation of existing data about energy supply, land use, and the specific types of greenhouse gases considered. For instance, projections about technology or land use are provided exogenously in some IAMs, while in others they are endogenous, meaning that they are allowed to vary based on interventions to other features of the global system.

Dynamics relates to IAMs' representation of the evolution of the economic system in light of decision-making processes and implementation. One distinction is between *inter-temporal optimization* models, where decision-makers are viewed as minimizing costs for climate targets at each point in time, knowing the long-term consequences of their decisions, and *recursive-dynamic* models, in which strategies for reaching climate targets are updated at regular time periods (Kriegler et al., 2015). A related design choice is the level of market imperfection that is assumed in IAMs, as measured by the delay in bringing to efficacy of climate policies simulated in each scenario. Recursive-dynamic models with market imperfections tend to yield higher overall costs of emission reductions than inter-temporal optimization models (Kriegler et al., 2015), especially when the latter do not incorporate market imperfections. Differences in realism can be compensated by other means, for instance, by assuming stronger constraints on the maximum speed of technology diffusion, higher discount rates for future goods, or lower levels of cooperation among world regions in achieving climate targets. Calibrating choices of model structure and parameters requires expert judgment, which accrues partly from comparison with other models. In view of the complexity of the task, IAMs are regularly curated and updated to strike an ever-shifting balance between representing every relevant aspect and computational treatability.

A particularly challenging aspect of designing detail-process IAMs is fine-tuning the constraints imposed to variables within a model, in a context of great variability and absence of long-term records (Lunt et al., 2024). Assumptions about the cost, availability, and upper bounds for deployment of mitigation technologies play an important role in determining scenario outcomes and have traditionally been given much attention by modelers (Bosetti et al., 2015). Despite many developments, the uncertainties regarding maximum speed of the energy transition remain large. Recent analyses, for instance, disagree about the possible extent and rate of deployment of carbon removal technologies (Way et al., 2022). Increasingly, the IAM community is recognizing that aspects related to economic preferences, social acceptability of policies and technologies, and severe climate impacts deserve just as much critical attention as technology projections (Peng et al., 2021). In general, the sensitivity of the human-climate system to disruptive events – climatic tipping points, so-called 'black swan' events, political turnovers, etc. – is high. Factors ranging from severe climate impacts to radical behavior change can spur unprecedented innovation or, alternatively, generate economic or technological lock-ins in a way that is hard to capture via models that frame sociotechnical change in an incremental way (McKay et al., 2015).

Evidently, as approximations about the initial conditions of the global system compound with uncertainties about future trends, so grows the risk of emission scenarios losing grip over what may realistically occur over the medium and long term. In this respect, it is open to debate whether IAM scenarios are individually more trustworthy than anticipation and foresight exercises produced by experts (Saltelli et al., 2024). At the same time, the comparability of IAMs to one another constitutes a key strength of their outputs. Through quantification, it is possible to accurately compare both design assumptions and scenario outputs in a way that is not possible for qualitative foresight exercises. Moreover, in the way of comparing IAMs many otherwise opaque assumptions come to the fore, enhancing their overall inspectability (Lenhard & Winsberg, 2010). These considerations lead IPCC's WGIII to the use of ensemble methods for scenario assessment, where uncertainty is recognized alongside scientific consensus and rendered itself a subject of study. The next section will elaborate historically on WGIII's approach to building its emission scenario database.

3. Constructing the ensemble

Collective reflection on virtues and limitations of IAMs has significantly shaped the evolution of WGIII's ensemble approach. To understand what makes WGIII's current ensemble methods distinctive, a comparison with WGI's assessment of simulations of future Earth's climate is useful. Multi-model ensemble exercises in WGI are structured around a stable body of established experimental protocols – the Coupled Model Intercomparison Project (CMIP) for historical simulations and the Diagnostic, Evaluation and Characterization of Klima (DECK) for climate sensitivity – together with a variable number of Model Intercomparison Projects (MIPs). The latter must follow strict submission rules for their simulations to be admissible in WGI's database (Eyring et al., 2016). MIPs can only address research questions that the CMIP has antecedently endorsed and consider a fixed type of simulations from each climate model to mitigate the problem of model over-representation (Katzav et al., 2012). These ensemble protocols can be interpreted as WGI's proprietary way of adhering to IPCC's mandate to synthesize "policy-relevant but policy-neutral" (IPCC, 2024) information, by constructing an ensemble that aims to be as much as possible free from biases in model representation. By contrast, WGIII's recent history displays a clear trend towards a more minimal form of control over the shape and content of its ensemble.

Early WGIII assessment exercises saw an active role of WGIII experts in commissioning emission scenarios (Nakicenovic & Swart, 2000). A particularly influential ensemble exercise was the IPCC's Special Report on Emission Scenarios (SRES) in 2000, where forty IAM-based scenarios were commissioned to the modeling community. The SRES exercise featured a *structured ensemble* now common in IAM intercomparison projects, in which scenario outputs are produced with a pre-determined design (Guivarch et al., 2021). Specifically, a stated aim of the SRES exercise was to cover a wide variety of long-term emission trajectories consistent with observed physical, socio-economic, and technological conditions. The SRES also marked an important step in the development of techniques for IAM inter-comparison (Pedersen et al., 2022). Following the SRES assessment, four representative trajectories combined with economic and population growth estimates were selected from the database. Their role was to be used as a 'no-policy' emission baseline to compare against scenarios with climate policies implemented (Fig. 2). In addition to their utility in evaluating climate policy, the use of these 'SRES baselines' quickly expanded beyond the IAM community and became standard reference also in the physical climate science and the impact, adaptation, and vulnerability (IAV) communities (van Beek et al., 2020).

As the work of IAM modelers in the IPCC became more prominent, it became subject to growing attention and criticism (Girod et al., 2009; IPCC, 2005a; Pedersen et al., 2022). Earlier controversies over economic evaluations carried out by WGIII to estimate climate-related damages found a new target in the opacity and policy-partiality of IAMs' design assumptions (Asayama et al., 2022). Critical analyses performed on IAMs gradually evidenced the relevance of assumed constraints about technology, including upper bounds for deployment of renewable energy and carbon removal tools (Krey et al., 2019; Way et al., 2022). In addition to parametric assumptions that are inherently difficult to validate (Bosetti et al., 2015), structural aspects of IAMs also raised concerns. The optimization methods used by IAMs for identifying pathways that achieve given temperature goals, based on identifying cost-effective solutions while neglecting responsibility for GHG emissions, were increasingly perceived as partial to technological solutions to the climate problem and at odds with the recognition of justice as a key element in the debate over responses (Zebrowski et al., 2022).

In view of the emerging limitations of IAMs in the policy context, a discussion ensued over the shape that the WGIII scenario assessments should take to preserve their stated policy-neutrality. The 2005 IPCC workshop report on new emission scenarios (IPCC, 2005b) contains a detailed account of both experts' and country representatives' stances concerning the IPCC assessment of mitigation options and the role that WGIII authors should assume in the construction of future scenario databases. The objection to the IPCC's commissioning scenarios as in previous reports was voiced in terms of "a potential conflict of interest – i.e., IPCC both creating and assessing scenarios" (IPCC, 2005b, p. 12), accentuated by the large involvement of IAM modelers within WGIII itself.[3] The decision that eventually emerged was to reduce WGIII's role to one of assessment (IPCC, 2008). In practice, WGIII would limit itself to evaluating scenario exercises voluntarily submitted by the scientific community (Pedersen et al., 2022). Accordingly, for its Fourth Assessment report, IPCC WGIII opened calls for scenarios while retaining ownership of post-SRES baselines, to assist researchers in the submission process and ensure comparability (IPCC, 2007). Starting from the Fifth Assessment Report (AR5), the IPCC gave up on ownership of baselines as well (IPCC, 2014). The separation of the IPCC from scenarios collected by WGIII is recognized explicitly in the latest Assessment Report, AR6: "the IPCC is neutral with regards to the assumptions underlying the scenarios …, which do not cover all possible futures" (IPCC, 2022:21)

With the passage to an unstructured approach, the role of coordination of modeling efforts was deferred to the scientific community. In preparation for AR5, a community-led research effort brought about the

[3] On procedures of expert selection in the IPCC, see Jebeile (2020).

Representative Concentration Pathways (RCPs), detailing alternative GHG concentration baselines for up to the end of the century. RCPs span a range of changes in atmospheric composition measured in terms of net energy transfer through the atmosphere, from a low-emission scenario of 1.9 W/m^2 by 2100 to higher-emission scenarios of 4.5, 6.0 and 8.5 W/m^2 (van Vuuren et al., 2011). The latest addition for scenario baselines are the so-called Shared Socioeconomic Pathways (SSPs), which integrate RCPs with multi-IAM projections of key emission drivers – population growth, GDP, and urbanization. SSPs are construed as quantitative expressions of five distinct narratives of how human society may evolve over this century (Riahi et al., 2017), classified by level of challenge to mitigation and adaptation. SSP1 ('sustainability') and SSP3 ('regional rivalry') are at the two extremes, with SSP1 presenting low challenges for mitigation and adaptation due to global cooperation, and SSP3 presenting high challenges on both dimensions due to severe regional conflict. SSP4 ('inequality') presents a world with low challenges in mitigation and high challenges in adaptation, while the opposite holds for SSP5 ('fossil-fuel development'). Finally, SSP2 was intentionally designed to represent a central, 'middle-of-the-road' scenario, featuring uneven climate action and moderate-to-high challenges for both adaptation and mitigation.

WGIII's ensemble construction workflow in IPCC AR6 reflects the hybrid role currently occupied by the IAM community, as both formally separated from the IPCC and informally an indispensable source of know-how for the ensemble construction and curation process. To invite submissions, the IPCC opens a public call for scenarios. Submitted scenarios articulate the divergences, due to climate policies, from a chosen emission baseline. Modeling teams can use RCP/SSP baselines for their submissions, but do not need to. The main requirement for passing the vetting phase is that scenarios incorporate values for a set of indicators in a range consistent with historical observations. Current indicators include specific ranges for all major GHGs emissions as well as global energy production from major sources, including fossil fuels, nuclear, solar, and wind (IPCC, 2022:1883). While WGIII has welcomed scenarios beyond IAMs, a significant burden is placed upon non-IAM modeling teams to render their outputs as comprehensive in required detail (Cointe 2024). For all scenarios that pass the vetting phase, WGIII experts curate the calibration of their inputs and boundary assumptions with the database format ('harmonization'), insert any missing physical dimensions of interest ('infilling'), and classify scenario outputs by temperature category. The latter classification has been achieved by means of climate model emulators devised by WGI, which assign likely temperature outcomes to emission projections (Kikstra et al., 2022). The full scenario database has been disclosed only at the end of WGIII's assessment.

Regarding the variety of scenarios represented in the ensemble, more than 2000 global emission scenarios have been submitted to AR6 by the scientific community and nearly 1700 have passed the vetting phase (IPCC, 2022:1884). These are higher figures than the number of simulations collectively assessed in WGI by CMIP, DECK and CMIP-endorsed MIPs in the latest report (Eyring et al., 2016). WGIII classifies submissions by originating modeling framework, where sporadically two or more frameworks belong to the same model family (IPCC, 2022:1888). Vetted scenarios come from about two dozen IAMs, with high disparities in submissions per participating model family (Fig. 3) and, at the same time, a fair variety of temperature categories represented in the ensemble (Fig. 4). In addition to submissions from single models, scenarios resulting from multi-model comparison exercises are also present. Finally, the AR6 database contains about 1000 scenarios of regional, national, or sectoral coverage, which are produced by a greater variety of modeling frameworks (mostly energy system models), and are assessed separately from global scenarios. This evolution is noteworthy considering the historic prevalence of IAMs in WGIII's past assessments.

In summary, this section has highlighted how controversy over IAM scenarios has not only shaped the evolution of WGIII's assessment methods but led to a squarely different ensemble construction approach from that in use in the physical climate sciences.[4] Inherent in WGIII's unstructured protocols is a different conception of a "policy-neutral" (IPCC, 2024) assessment than what is assumed in WGI. Whereas the latter sets up a highly structured ensemble partly to control the number and type of simulations of future climate considered in it, WGIII has favored minimizing decisions as to whose emission scenarios ought to be assessed in its exercise (besides minimal informativeness and consistency with historical record) and which research questions should be pursued through scenarios.[5] To borrow an expression from Parker's (2010) work on multi-model studies, WGIII's approach exemplifies a minimal form of "*ownership"* of a scenario ensemble. Specifically, while recognizing a role for scenarios in the assessment of climate policy options, WGIII does not assume control over the baselines and scenarios assessed in its ensemble exercise, thereby maintaining a neutral position with regards to whether the scenarios explored in the assessment are representative of possible future outcomes, or what policy questions the modeling community should explore via scenarios to assist WGIII's evaluation.[6] As the next two sections will explore, this distinctive feature of WGIII's latest ensemble assessments has important implications for the type of policy-relevant information that can be extrapolated from the scenario databases and the type of policy interests that can be served by them.

4. Assessing the ensemble

What can be learned from an unstructured ensemble of emission scenarios? Any discussion of this issue should start from acknowledging the practical value of a curated scenario database. Peer-reviewed scientific publications and grey literature reports employing emission scenarios often present IAM modeling inputs and outputs with different definitions and boundary assumptions, or report some relevant quantities but not others, which makes it difficult to compare scenario studies to one another. Disclosing a scenario database with fully standardized outputs and harmonized assumptions therefore contributes to the transparency of scientific assessments of climate policy options and the inspectability of scenarios by other experts (IPCC, 2022:1876). Moreover, the database curation process itself often represents a learning opportunity. For instance, the abovementioned process of categorizing emission scenarios by their likely temperature outcome has been the result of a coordinated effort across WGI and WGIII (IPCC, 2022:1889). This and similar coordinated activities across IPCC Working Groups have been, in the past, an important occasion for scientific growth.

With this clarification, we can proceed to asking what type of policy-relevant information can be drawn from a large unstructured ensemble of emission scenarios. This question has not received the attention it deserves in the philosophical literature; without addressing it, we cannot evaluate virtues and limits of WGIII's approach. For the physical climate sciences, Jebeile and Crucifix (2020) articulate two ideals informing multi-model ensemble studies: one analyzable formally in

[4] The evolution of the unstructured approach is an instance of the phenomenon of the 'proceduralization' of the IPCC in response to controversy, as highlighted in the work of De Pryck (2021; cf. Hulme, 2022).

[5] This is not to say that, as a matter of fact, WGI's ensemble are free from bias. See, e.g., Parker (2011), Jebeile and Crucifix (2020) on the 'ensemble of opportunity' problem.

[6] In Parker's language, "ownership" concerns the extent to which estimates of uncertainty resulting from the ensemble are ones "that the scientists are willing to claim as their own" (993). This conception relates to the specific problem addressed in her paper, viz., whether outputs of multi-model ensemble analyses can be given a probabilistic interpretation. As an anonymous referee has helpfully pointed out, the relevant notion of 'ownership' in this paper is different, since it concerns instead the aspect of *control* over scenario inputs. The main questions are: whose emission pathways are being assessed in WGIII's ensemble exercise? And who determines which questions ought to be addressed via scenarios?

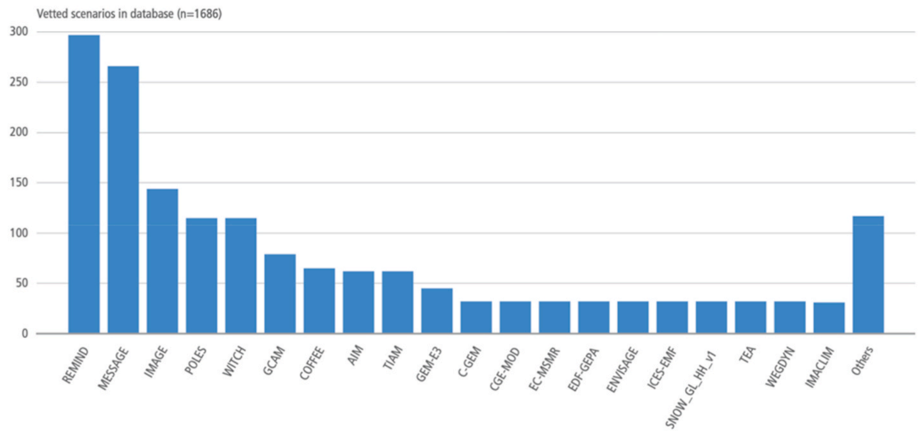

Fig. 3. Number of vetted scenario submissions in the IPCC AR6 database per model family (identified by their acronym). Source: IPCC (2022:306).

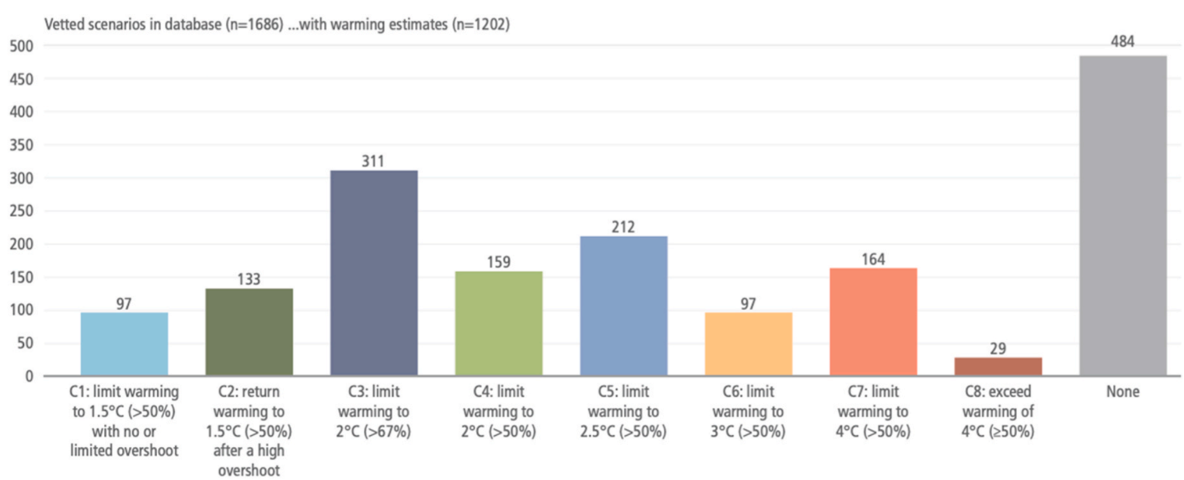

Fig. 4. Number of vetted scenario submissions in IPCC AR6 per temperature category. Source: IPCC (2022:307)

terms of sampling procedures, the other materially in terms of assessing collections of expert judgment. The former, statistical interpretation is clearly inapplicable to WGIII's ensemble for reasons that have emerged previously. First, as it happens with physical climate models, IAMs used to produce scenarios have not been developed independently from one another: while some structural design choices are independent, others have a shared origin, resulting in a classic instance of the 'ensemble of opportunities' problem (Parker, 2010). In addition, WGIII's unstructured protocols are distinguished by their placing no constraint on the number of submissions allowed from any participating model family. The inclusion of single-scenario studies, multi-scenario studies from a single model, and multi-model inter-comparison studies in WGIII's databases results in collection of scenarios with considerable under- and over-representation of model families (Fig. 3). It follows that no probabilistic significance whatsoever can be attached to elements such as the 'average scenario' or the 'interquartile range' in WGIII's latest ensembles.[7]

Unsurprisingly, the sheer size of WGIII's latest ensembles also pose an obstacle to a qualitative ensemble assessment, of a type more in line with the recognition of assessed scenarios as agglomerates or extensions of expert judgment (Jebeile & Crucifix, 2020). Whereas in a probabilistic assessment scenarios are evaluated by their divergence from an average output, in a qualitative assessment they are evaluated individually and collectively in the respect of their contribution to a given research question, mostly through like-for-like comparisons (Guivarch et al., 2021). The purpose of such an assessment – common in many IAM intercomparison studies – is to gain insights from observing how models that are similar in some respects and different in others respond to a given input. An important obstacle to the adoption of a comparative methodology is that the latest scenario databases contain several hundred submissions, addressing a large variety of research questions: from the feasibility of low-demand pathways to the carbon impacts of investments in transportation and education. The problem of finite reviewing capacity is only partially mitigated by ensemble construction protocols that require the provision of rich meta-data alongside scenarios (Minx et al., 2017).

The actual methodology of WGIII's latest ensemble assessments is a hybrid one, combining quantitative and qualitative components. To avoid adding layers of complexity, the discussion below will be limited to describing the main feature of each method, without delving into the hard questions of justification. The quantitative component consists in classifying scenarios by their respective temperature outcome and describing statistics concerning each category. This approach is only superficially related to the statistical assessment mentioned earlier, where submissions are treated as sampling the space of possibility. In this case, the ensemble is viewed as representative of scenario exercises

[7] There is disagreement in the philosophical literature as to whether statistical tools are appropriate for multi-model ensembles of physical climate simulations (Dethier, 2022; Parker, 2011). Insofar as WGIII's unstructured ensembles represent a special case, the claims made in this paper are compatible with Dethier's (2022) defense of the informativeness of statistical analyses in some multi-model ensemble studies – such as those performed by WGI.

available in the scientific literature and the aim is typically to identify one or more 'emerging constraints' concerning the achievement of long-term temperature targets (Winsberg, 2018). An example is AR6's reporting the median peak year for GHG emissions in vetted scenarios reaching the 1.5° Paris Agreement target with various degrees of probability, accompanied by the 5th-95th percentile in square brackets (IPCC, 2022:330-2). The fact that no emission scenario in AR6's ensemble reaches the +1.5 °C temperature goal with a peak emission year later than 2025 is then reported as a key finding in the *Summary for Policymakers* ("Global GHG emissions are projected to peak between 2020 and at the latest before 2025 in global modelled pathways that limit warming to 1.5 °C (>50%)"; IPCC, 2022:17).

Not every descriptive statistic in the assessment identifies an emerging constraint in the sense above. In the case of emission peak years, support for the claim that rapid and deep climate action is required derives from considering that submitted scenarios reflect the effort of distinct modeling teams to explore a wide range of realistic mitigation pathways – including ones where mitigation action is significantly delayed. WGIII's conclusion is therefore based on a convergence argument, of a type familiar to the climate sciences (Harris & Frigg, 2023). Here, two key considerations are the high number of assessed mitigation scenarios and, at the same time, the fact that they must be minimally distinct in their make, due to the contingent modeling choices often involved in producing them. In other cases, descriptive statistics from the database are relevant despite failing to identify an emerging constraint. This is true, for instance, for statistics concerning the contribution of different types of energy sources in assessed scenarios. While vetted scenarios meeting the most ambitious temperature goals (categories C1-C3 in the assessment) show a deep and, in most cases, immediate reduction of fossil fuel sources, the specific energy mix adopted by scenarios in these categories is otherwise under-constrained: some scenarios meet temperature goals mostly through heavy use of renewables, whereas others count on behavior change or biomass use with carbon capture and storage. The resulting lesson of 'policy freedom' for achieving mitigation is given emphasis in the *Summary* (IPCC, 2022:18).

WGIII's practice of detailing descriptive statistics is accompanied by a more qualitative assessment, via a process of selection of 'illustrative' emission scenarios. Here the collective voice of WGIII's authors becomes most noticeable, as selected pathways purport to "illustrate key mitigation-strategy themes that flow through several chapters in this report" (2022:77). To streamline the selection process, a stricter vetting procedure is applied to identify candidates, requiring greater accuracy in accounting for the historical record (IPCC, 2022:1884). The choice within the restricted set of scenarios is based mainly on qualitative criteria, including the perceived relevance to other parts of the assessment and the consistency between the mitigation strategies adopted, the chosen socio-economic narrative (i.e., the underlying SSP), and the projected effects of policy implementations for different world regions (IPCC, 2022:77). Through such a process, AR6 identifies five *Illustrative Mitigation Pathways* (IMP), standing for alternative policy routes to achieving similar temperature stabilization goals, as well as two *Illustrative Reference Pathways*, namely ModAct ('moderate action') and CurPol ('current policies'), outlining scenarios of increasingly higher climate risk, as global temperature increase rises steeply above +2 °C.

While one of the aims of IMPs is to catalyze democratic discourse about policy options (Edenhofer & Kowarsch, 2015), IMPs are also used as a source of insight at several points in WGIII's latest reports. This is achieved through two types of comparative exercises. Sometimes, IMPs are compared against descriptive statistics from whole-ensemble analysis. For example, AR6's Chapter 3 presents the median global warming of vetted scenarios divided by temperature category (IPCC, 2022:317) and pairs it with data on global warming across selected IMPs. In other occasions, a comparison is proposed directly among IMPs, to draw insights about mitigation strategies. For example, Chapter 4 in AR6 compares two IMP scenarios, 1.5-SP and 1.5-Ren (IPCC, 2022:453), which happen to be generated by the same IAM. The former, 1.5-SP, is a rapid transition pathway based on SSP1 assumptions – that is, a world that "shifts gradually, but pervasively, toward a more sustainable path" (Riahi et al., 2017, p. 154). The latter, 1.5-Ren, is a scenario featuring heavy use of renewables but under SSP2 assumptions – that is, in a world characterized by merely partial global climate action. The report points out that 1.5-SP achieves its long-term 1.5 °C target with considerably less costs than 1.5-Ren (IPCC, 2022:453). Because the two IMPs are based on the same underlying IAM, the comparison of the respective scenario outputs makes salient the importance of international cooperation and early action in facilitating climate ambition.

The quantitative and qualitative assessment methods just reviewed differ in the level of involvement of expert judgment. The selection of IMPs is an especially delicate stage for WGIII insofar as it assumes that not all scenarios are qualitatively on a par. However, the difference in the role of expert judgment should not be overstated. First, both quantitative and qualitative ensemble methods require that IAMs are considered fit enough for purpose. This implies a judgment about the extent to which model inputs and outputs can be regarded as credible when outlining scenarios (IPCC, 2022:304). Secondly, the task of selecting relevant descriptive statistics from the ensemble is an inherently interpretative one. To illustrate, WGIII AR6 reports that most emission scenarios reaching Paris goals require carbon capture and storage technology (IPCC, 2022:24), indicating the need for further research and development in this area. Conversely, statistics concerning the representation of policy tools adopted in the same scenarios are not emphasized. Despite carbon pricing and carbon budget limits being most frequently used in IAMs for ease of simulation, they are usually interpreted as placeholders for whichever portfolio of policy tools is most effective. AR6 explicitly states that diverse policy portfolios that include regulations, subsidies, and incentives tend to be more effective than single policy instruments, such as carbon pricing (2022:125).

A further methodological development of AR6 cannot be discussed in detail but is worth mentioning. After the IPCC Special Report on 1.5°, WGIII's latest assessment includes an evaluation of scenarios compatible with Paris Agreement goals based on their *feasibility* (Brutschin et al., 2021). Considering mostly analogies with historically observed technological and social transitions, the feasibility framework adopted by WGIII associates each mitigation scenario with a score of (roughly) how unprecedented is the speed of societal and technological changes that it demands for reaching its temperature outcome (IPCC, 2022 III:380).[8] By classifying the economic, societal, and technological requirements of low-emission scenarios based on their feasibility, WGIII contributes to correcting the well-documented issue of selection bias in scenario ensembles, where submissions reaching the most ambitious temperature goals incorporate, on average, particularly 'optimistic' assumptions, relative to scenarios that do not meet those goals (Tavoni & Tol, 2010). In addition to a debiasing role, feasibility considerations also serve to identify knowledge gaps. One of the results of AR6's feasibility analysis highlighted in the report is that socio-economic and institutional aspects of the energy transition in scenarios compatible with Paris Agreement goals impose greater feasibility challenges than geo-physical or technological aspects (IPCC, 2022:146), indicating the need for further research on those dimensions.

5. Consequences of the unstructured approach

By allowing any scenario that meets a threshold of informativeness and consistency with historical records, the unstructured approach significantly reduces WGIII's involvement in ensemble construction. At the same time, WGIII's quantitative and qualitative methods still allow to address an important variety of policy-relevant issues. The question

[8] For a critical uptake of feasibility analyses in IPCC assessments, see Hollnaicher (2025).

whether the methods used make WGIII's assessments 'policy-neutral' depends on the stipulated meaning of the latter expression and will not be a central concern here (cf. Elliott, 2025; Havstad & Brown, 2017; Jebeile, 2020). A less ambiguous question that can be addressed is how well current unstructured protocols serve the purpose of offering information about mitigation options. Historically, information concerning the capacity of modelled pathways to reach ambitious temperature goals has been a key input to the design of international agreements, including the Paris Agreement (Kowarsch et al., 2017). However, the users of WGIII ensemble information have expanded since the earlier reports. Today, 'policymakers' addressed by the IPCC are users with increasingly distant policy interests (Peng et al., 2021). Compared to diplomats and designers of international agreements, for instance, government representatives must plan for country-level mitigation and adaptation strategies with a view to an increasing number of climate-induced risks, exacerbated by lack of cooperation and global inaction; their interests in ensemble information may therefore be different as well.

This section aims to show that, by the nature of its methods, WGIII's unstructured approach currently addresses some policy-relevant questions at the expense of others. The limitations discussed below are independent of the problem of 'ensembles of opportunity' often raised in the literature. This critique is common among 'post-growth' and other non-mainstream economists, when they claim that WGIII's ensemble displays partiality towards structural modeling assumptions that are typical of neo-classical growth economics (Hickel et al., 2021; Kallis et al., 2018; Keysser & Lenzen, 2021). While important, the problem of limited diversity of model structures will not be discussed below for two reasons. First, it is by no means specific to the unstructured approach, as it affects virtually all multi-model ensemble exercises in the climate sciences (including the physical climate sciences) regardless of design (Parker, 2011). Second, addressing it requires *capacity building*, viz., the construction of modeling frameworks with alternative structures and supported by alternative theories. The velocity with which capacity building occurs in the domain of climate economic modeling depends on the resources allocated by funding agencies and the availability of theories and experts; it is not attributable to WGIII's unstructured approach – which, on the contrary, was originally designed to be open to a wide variety of scenario exercises. What the following subsections will try to highlight is that WGIII's choice of an unstructured approach has distinctive costs of its own, which ought to be recognized as such in our evaluations.

5.1. Variety

Both challenges that will be discussed in what follows have a common cause: the lack of coordination in the current process of scenario elicitation. The first problem, *missing variety*, is a clear instance. An important fraction of the uncertainty space which is missing or underrepresented in WGIII's latest scenario ensembles is of a relatively avoidable kind. For example, as IPCC AR6 recognizes, the wide majority (around 90%) of scenarios assessed in AR6 employ a SSP2 ('middle-of-the-road') baseline (IPCC, 2022:309). By contrast, the representation of scenarios assuming other SSP baselines - particularly the gloomier outlooks of SSP3 and SSP5 – is relatively sparse in the AR6 ensemble (Dekker et al., 2023).[9] The disparity in representation is even clearer when considering specific SSP inputs. For population, the 25-75 percentile range of vetted scenarios in AR6 centers narrowly around SSP2's projection of (slightly above) 9 billion people by 2100 (Giarola et al., 2024). WGIII authors themselves note the discrepancy with the demographic projections produced by the United Nations, which deem nearly 11 billion people as the median (UN 2019), and relate the lower estimates in AR6 to more optimistic assumptions about rising education levels (IPCC, 2022:313). However, SSP2's demographic projections are conservative even by alternative estimations, such as probabilistic methods. Moreover, AR6 is no exception, since a comparable under-sampling of population is present in AR5 (Giarola et al., 2024).

The missing representation of baseline assumptions in WGIII's latest databases is the consequence of an uncoordinated collection of scenarios inherent in the unstructured approach. In this specific instance, the over-representation of SSP2 in the AR6 ensemble can be partly attributed to an implicit judgment that 'middle-of-the-road' is the most plausible SSP baseline to assume given observed trends (cf. Riahi et al., 2017). Consequently, the modeling community may have prioritized mitigation scenarios that address policy requirements under this specific baseline. Furthermore, a selection bias is likely to be present: compared to baselines such as regional rivalry (SSP3) or fossil-fuel development (SSP5), IAMs running SSP2 projections stand a greater chance of finding solutions with relatively low temperature outcomes, making it more likely that such scenarios will appear in the ensemble. Notwithstanding this over-representation, the robustness of the most significant conclusions that WGIII draws from the AR6 ensemble remains unaffected. In a 'no-policy' SSP2 baseline, global temperature increase is projected to range from 2.1 °C to 3.9 °C (Riahi et al., 2017) – far above Paris Agreement targets. Consequently, statistics about emission peak years and primary energy sources for SSP2-based mitigation scenarios reaching below 2 °C of global warming are still relevant for highlighting constraints on effective climate action.

At the same time, missing SSP variety has repercussions on other aspects of WGIII's ensemble assessments. First, it appears to reduce the *level of confidence* with which WGIII authors express claims concerning the high-emission end of the uncertainty space. An illustration can be found in statement C.1 in WGIII AR6's *Summary for Policymakers.* While the first sentence states with "*high confidence*" that "Global GHG emissions are projected to peak between 2020 and at the latest before 2025 in global modelled pathways that limit warming to 1.5 °C (>50%)", WGIII authors continue by noting that "Modelled pathways that are consistent with NDCs announced prior to COP26 until 2030 and assume no increase in ambition thereafter have higher emissions, leading to a median global warming of 2.8 [2.1–3.4] °C by 2100 (*medium confidence*)" (IPCC, 2022:17). Although WGIII authors do not elaborate on the reason why ranges in emission peak years are reported with higher confidence than ranges in temperature, an important difference between the two cases is that emission scenarios consistent with current NDCs and less optimistic SSP assumptions are represented very sparsely in the AR6 ensemble, compared to SSP2-based scenarios. This difference in representation could result in lower confidence when making claims about modelled pathways with higher emission trajectories, compared to mitigation pathways that seek to keep temperature outcome ranges closer to Paris Agreement targets.

Second, and most importantly, insufficient variety of SSP assumptions also affects the *type of information* about mitigation strategies that one can draw from the database. A collection of scenarios with a sufficient distribution of temperature outcomes can be useful to provide insights about the constraints that global climate action is subject to if the world is to achieve Paris Agreement targets. However, it is at the very least not obvious that the most cost-effective mitigation strategies to reach those goals would remain effective under a broader variety of future circumstances, including more pessimistic outlooks. For example, mitigation strategies that require high upfront investment and strong international cooperation (e.g., nuclear energy) may generate technological lock-ins in worlds marked by severe regional conflict. Including a broader set of SSPs could help provide meaningful statistics for a broader variety of interests and users of WGIII's ensemble assessments. For instance, national governments increasingly need to plan long-term transition investments under conditions that may diverge sharply from 'middle-of-the-road' assumptions, including futures shaped by inequality, regional conflict, and slow adaptation to technological

[9] Another clear instance of missing variety is reported in Sognnaes and Peters (2025).

change. For their policy needs, expanding the database to countenance other SSP baselines could prove useful.

Recent applications of whole-ensemble methods to large scenario databases corroborate the above observations. Dekker et al. (2023) adopt variable decomposition tools to obtain more informative statistics from large scenario ensembles. In short, scenario outputs in the ensemble are decomposed in terms of the contributing role of model used, temperature target achieved, and underlying SSP assumption. This disaggregation is made possible by the large number of scenarios and the choice to disregard models with few scenario submissions (Dekker et al., 2023). Consistent with the above remarks, their analysis proves somewhat rewarding only when studying the variation of policies under alternative temperature targets. In a follow-up study, Al Khourdajie et al. (2024) obtain stronger disaggregation of climate policies by both temperature outcome and SSP assumptions from a structured ensemble. Their decomposition not only gives indications about which mitigation strategies for a given temperature outcome depend on the model, but also - and perhaps most importantly – about which mitigation strategies tend to be adopted in similar models when varying SSP baseline. An implication of the study is that greater power to discriminate which policies tend to be stably represented under a variety of alternative futures is possible when the ensemble is less biased in its representation of SSPs. Evidently, an unstructured approach such as WGIII's cannot guarantee that scenarios in the database are well-distributed along both temperature category and SSP assumptions: to possess the relevant features for decomposition, WGIII would have to impose further constraints at the stage of scenario elicitation.

5.2. Linkages

Connecting the information drawn from global mitigation pathways with the regional and local level where political decisions are made and felt is another persisting source of difficulty for WGIII's assessments. This problem – *missing linkages* – had already emerged in criticisms to AR5 (Beck & Mahony, 2019; Kowarsch, 2016), where a misalignment was noted with the Paris Agreement's recognition of the "imperatives of a just transition" (G7 Leaders Declaration, June 2015) as an enabler of climate action. Despite many developments, missing linkages persist as a problem for AR6. An illustration is given by the abovementioned treatment of IMPs. Comparing the scenarios 1.5-SP and 1.5-Ren that are produced by the same IAM, WGIII authors note that the former achieves its ambitious mitigation goals with considerably less societal costs than the latter (IPCC, 2022:304). However, this observation reports a merely aggregative fact about the costs represented in each scenario and is not directly associated with a fairer distribution of burdens across and within countries. Information about the SSP instantiated in each IMP – SSP1 for 1.5-SP and SSP2 for 1.5-Ren, respectively – is useful but insufficient. While SSP indicators include aspects such as aggregate economic inequality, average income, and levels of educational attainment and gender disparity, they are not informed by differential responsibility for GHG emissions or by consideration of region-specific vulnerabilities (Zimm et al., 2024).

The fact that ensemble assessments do not yet speak to the policy interests of many regional and country-level actors is standardly attributed to one aspect of IAMs: their tendency to solve for cost-effective pathways to temperature outcomes via decisions that take into account the full portfolio of market options, without regard to who ultimately bears the cost of the transition (Zebrowski et al., 2022). Recognizing this limitation, AR6 adopts an integrative approach, marking an important step beyond exclusive attention to IAM-based scenarios. The task of addressing the regional- and country-level scale is distributed across various AR6 chapters. In particular, Chapter 4 considers a dataset comprising roughly 500 scenarios from national development models, providing short- and medium-term outlooks on economic development and mitigation efforts for several countries (IPCC, 2022:1891). Submitted by research teams under the sole condition of being peer-reviewed, national scenarios aim to reflect the diverse starting points and development priorities, to allow for a more granular understanding of mitigation potential and obstacles. Country-level information is complemented by a multi-chapter synthesis of sectoral scenarios – spanning from buildings to transportation to food systems –, detailing how emission reductions in each sector contribute to the United Nations' Sustainable Development Goals (SDGs) agenda.

While IPCC AR6's effort at drawing missing linkages between global scenarios and regional information marks an important progress relative to previous WGIII reports, some important limitations persist, which can again be attributed to unstructured protocols. First, as WGIII authors recognize, the national scenario database in AR6 shows consistent gaps in representation of some countries' perspectives, particularly for low-income regions due to limited data availability and missing modelling infrastructure (IPCC, 2022:1892). Second, associating SDGs' achievement with greater perceived fairness and political acceptability of mitigation and adaptation options is not uniformly supported. For instance, various counterexamples are discussed in recent literature on energy justice (Carley & Konisky, 2020). Third, it deserves stressing that the assessment of national and sectoral perspectives in AR6 does not culminate in an enhanced appraisal of the social and political feasibility of global mitigation scenarios – not even the IMPs previously discussed. To obtain the sought-for integration, modeling exercises should have been made available and used for comparison in a more systematic way for each world region and sector, which is not consistent with current assessment workflows. Consequently, AR6's integration between global and regional perspectives on mitigation options remains partial, constraining the degree to which scenario assessments can be considered relevant to informing regional climate policy.

Examples of enhanced modeling integration to identify region- and country-specific solutions compatible with global targets are not missing in the literature. For instance, a recent study by Koberle et al. (2022) analyzes and compares the verdicts of global and regional modeling exercises on Brazil's prospects for biofuel use as a mitigation strategy. As the authors note, Brazil's biofuel production is linked to an established chain of legacy economic activities for the country, which involve both agriculture (sugarcane production) and industry (such as mills and distilleries), with both positive social impacts (e.g., jobs) and negative externalities. Koberle et al. argue that global IAMs do not adequately consider the specificity of biofuel production in the context of Brazil's economy, instead yielding recommendations for a cost-effective energy mix to reach long-term mitigation goals that are mostly in line with those provided for many other world regions (2022:24). At the same time, their study suggests that region-specific models designed to offer greater detail on Brazil's biofuel production, such as the regional model COFFEE, tend to miss opportunities for transition cost reduction deriving from consideration of the global context in which Brazil's policy choices are made (2022:25). The methodological lesson that Koberle et al. highlight is therefore that global and regional-level modeling frameworks are best conceived as partial and in need of reciprocal integration when informing proper mitigation action.

The methodological ideal of deep integration prospected by Koberle et al. may be viewed as a further step towards enhancing WGIII's assessment of mitigation options, furthering the work that has already been accomplished in AR6 to connect global with regional scenarios. In practice, however, it is hard to see how such a prospect could be achieved by WGIII while retaining its current unstructured approach. One way for WGIII to bridge the gap between the global and the regional- or country-level scales would be to review and assess selected combinations of global emission scenarios, regional and country-level modeling exercises, and sectoral studies for key regions and sectors. Yet again, such coordination cannot be merely expected to result from current protocols. Some form of anticipation seems to be necessary: some structure that can bring about the desired integration of models non-accidently. Missing linkages therefore provide a further reason, in addition to the missing variety of scenario assumptions, for seeking greater coordination across participating modeling exercises than what WGIII's unstructured protocols currently allow for.

6. Beyond the unstructured approach?

Should WGIII reform the unstructured approach? To answer this question, two pieces of theory would be needed. The former is an epistemological model, grounded in ideal theory, that addresses what can in principle be learned from a collection of model-based scenarios such as WGIII's. The latter is a piece of non-ideal theory that addresses the material realization of a scenario ensemble and places WGIII's work within the network of relevant scientific and policy actors, particularly in relation to the IPCC's mandate for “policy-relevant but policy-neutral” (IPCC, 2024) work. This paper has provided neither of these accounts. The previous sections have had the sole objective to introduce the epistemological and institutional dimensions of WGIII's choice of ensemble approach, singling out the main considerations that any adequate answer to the reform question would need to take into account. One of the main points of the previous discussion is that, in addition to the benefits of the unstructured approach already discussed, there are costs associated with this choice, which any comprehensive appraisal should recognize. These costs can be expressed in terms of a difference in service that WGIII's ensemble assessments provide to users with different policy interests. Even so, there remains an open question whether it is sensible to reform the unstructured approach to remedy these disparities, or whether other measures should be prioritized to enhance the informativeness of WGIII's future ensemble assessments.

One of the key elements that has emerged from the previous reconstruction is that three dimensions are particularly relevant to characterizing ensemble methodologies in the context of WGIII's scenario assessments. These are: who controls which scenarios and associated uncertainties are assessed by WGIII experts (*ownership*), which methods are used to extrapolate relevant ensemble information (*logic*), and what type of climate policy information is ultimately delivered to the report's intended users (*purpose*). The previous discussion has further attempted to clarify in what sense these dimensions are interconnected in WGIII's choice of unstructured protocols, meaning that we cannot attribute WGIII's choice of unstructured approach solely to considerations alongside a single dimension, but rather to a convergence of considerations that pertain to all three dimensions simultaneously (Fig. 1). In effect, without recognizing how the unstructured approach achieves a balance between these different dimensions, we are unable to fully grasp the role that it plays in WGIII's positioning in the complex landscape of global climate policy evaluation. Based on the above results, this section briefly reviews two reforms of ensemble methods recently proposed by past WGIII authors: a ‘structured’ (Peters et al., 2023) and a ‘living database’ (Pirani et al., 2024) reform. Rather than offering an argument for either prospect, the main contribution of the following subsections is the clarification and elaboration of each proposal through the framework developed above – based on *ownership*, *logic*, and *purpose.*

6.1. Structured ensemble

A first reform prospect for WGIII's assessment is a return to a structured ensemble. The SRES precedent is unlikely to be a fitting model given the low number of scenarios assessed in that exercise and the evolution of IAM-based research over the past decades. A more natural suggestion is to borrow the structure behind WGI's assessments of physical climate simulations, which features a central overseeing body and an accompanying series of inter-comparison projects (MIPs). Reflecting on experiences in past WGIII assessments, Peters et al. (2023) defend a reform proposal of this kind. They envision a process of scenario database curation in WGIII AR7 that is structured around several MIPs, with peer-review processes overseeing “research questions, scenario protocols, and expected outcomes” (2023:4) and determined before submission to each MIP is allowed. This is intended to ensure that potential insights from comparison of scenarios that address similar research questions do not get lost in the way of assessing a scenario database featuring thousands of submissions. Furthermore, in analogy to WGI, Peters et al. call for a new “overarching entity that oversees and governs these activities [viz., the MIPs]” (2023:4) modelled after WGI's CMIP, possibly accompanied by an analogue of WGI's DECK that oversees “standardized experiments and model diagnostics” (4). The result is a radical reform of the institutional structure that IPCC WGIII currently employs for the construction and assessment of its scenario database.

Embracing a structured approach has strong potential to address the assessment limitations discussed in the previous section. Among other things, the prospective overarching entity could impose requirements to individual MIPs to ensure that a variety of model families, SSP assumptions, and climate targets are represented in the ensemble. The problem of missing linkages could also be mitigated by WGIII's adopting tailored MIPs for different world regions or, as suggested by Peters et al. (2023), through a dedicated “Equity MIP” (4) that addresses the local and regional impacts of global mitigation scenarios. In this sense, a structured ensemble promises significant improvements along the dimensions of logic and purpose. At the same time, ownership is an obstacle. As discussed in section two, being precise about *whose* mitigation pathways are assessed by WGIII has been, historically, a critical element in WGIII's conception of a policy-neutral assessment. While the institutional reform prospected by Peters et al. does not countenance WGIII's producing their own scenarios, it counteracts a clear trend towards minimal ownership of the scenario ensemble – one whose reasons we can only appreciate if we adopt a historical perspective. Due to the value-ladenness of climate economics as a social science and the uncertainties surrounding the design of IAM-based scenarios, an overarching entity in charge of deciding what are the relevant research questions for MIPs to address and what are the inclusion criteria for each assessment cycle is likely to raise new concerns of policy-partiality for WGIII.

Anticipating the risks associated with greater ownership of the ensemble, Peters et al. (2023) do not go all the way in replacing WGIII's unstructured protocols with structured ones. Recognizing that social science MIPs are unlikely to cover all possible questions and modeling approaches of potential relevance, they countenance the “submission of individual studies” (4) to future WGIII databases in addition to those submitted through MIPs. This implies that the resulting full scenario database will still not be “a random statistical sample of all hypothetical scenarios” (4), but will be likely subject to some degree of over- or under-representation of model families and scenario assumptions just as WGIII's latest ensembles. As a result, Peters et al.’s (2023) considered view turns out to be a hybrid between a structured and an unstructured approach. Even so, it must be stressed that the idea of structuring a significant part of the multi-model scenario assessment would require WGIII to take much greater control over its ensemble than it currently does. The disagreement about whether the prospected state of affairs is desirable is philosophical and runs deep (Hermansen et al., 2023): for some, a policy-neutral assessment is an impossible goal to achieve and should never have been part of the IPCC's defining mission; for others, the IPCC's taking up greater ownership for its scenario ensemble may be a strategic mistake in the long run.

6.2. Living database

An alternative to structuring the ensemble could be moving towards a ‘living database’ approach. Rather than adding a governing body, this approach removes the constraint whereby the scenario database is accessible to the scientific community only after the assessment by WGIII experts. Instead, WGIII would provide the scientific community with access to the database while it is assembled. In their call for a smoother workflow between IPCC Working Groups, Pirani et al. (2024) defend a proposal of this kind. Their argument is based mainly on the pressure that the current timeline for ensemble construction and assessment imposes on WGIII experts. According to Pirani et al., the time passing from the initial drafting to the report's final approval is relatively short; since a functioning database requires harmonizing and infilling submitted scenarios, the amount of time that WGIII experts can spend for proper analysis and assessment of scenarios is even more limited. To address these and other issues with coordination across Working Groups that have emerged in the latest assessment, Pirani et al. propose an open WGIII database with live community-led vetting and harmonizing of scenarios. In deferring vetting and harmonizing to the scientific community, the 'living database' reform effectively radicalizes the minimal ownership approach that is distinctive of WGIII's latest ensemble assessments.

The question that concerns us is whether a living database would address the limitations with unstructured protocols discussed above. Evidently, if the approach is to address the problems of missing variety and missing linkages, it must appeal to the self-coordinating capacity of modeling teams. This would require the modeling community to not only assume a more active role in curating the ensemble, but also in filling in the missing dimensions of uncertainty and missing linkages with regional- and country-level scenarios. Yet, it is far from clear that merely giving access to the IPCC scenario database would result in such an outcome. Pirani et al.'s (2024) discussion appears rather ambiguous in this regard, presumably being torn between a stronger and a weaker proposal for greater community engagement. In effect, a first challenge for a living database solution would be to specify the affordances and incentives that would make it likely for the modeling community to contribute productively to ensemble construction. A second and even more crucial question has to do with the dimension of logic: specifically, whether an ensemble that has been so heavily curated in its construction would warrant the same conclusions, and with the same level of confidence, as an unstructured ensemble featuring 'blind' scenario submissions.

Regarding the affordances, encouraging signals for a living database approach come from the field of statistics. Recent literature explores the use of emulators based on machine learning techniques for the detection of missing structural and parametric uncertainties in scenario ensembles (Wang et al., 2019; Doury et al., 2023). While imperfect and fallible in many ways, such tools could potentially allow WGIII authors to perform multiple quick analyses on missing dimensions of uncertainty in the database. Based on them, WGIII could then elicit further scenarios from the modeling community before moving on to assessment. A source of greater concern remains whether the resulting ensemble would support at least the same amount of learning from scenarios as an unstructured one. While a database that is inaccessible before the assessment inevitably contains some noise, it guarantees that collected scenarios will reflect independent efforts of modeling teams as much as possible. In this sense, the convergence of 'blind' submissions from independent modeling teams can be regarded as fulfilling an evidential function. Conversely, a living database could allow for strategic alignment in scenario submissions. If the approach is to be regarded as a serious candidate for improving on unstructured protocols, then, its defenders must take up the burden of explaining how the risk of artificially inflated results can be mitigated, so that the credibility of the ensemble assessments provided by WGIII experts is not affected. Further research, beyond what can be presently achieved, would be needed to address this question.

7. Conclusion

WGIII's methods for assembling and evaluating emission scenarios for climate policy evaluation are philosophically interesting as a case study on the science-policy interface and yet underexplored. This paper partly remedies the gap by combining epistemological analysis and historical perspective. Any evaluation of the practice cannot neglect how WGIII's ensemble assessments yield figures about the timing of climate action, the benefits of a concerted effort, and the opportunities for mitigation that continue to inform the political and legal discourse. At the same time, some limitations of these assessments must also be noted. As section three has discussed, WGIII's assessments suffer from a similar problem of 'ensembles of opportunity' as the one that affects most other ensemble exercises in climate science: models underlying emission scenarios are not entirely independent from one another in their design (Parker, 2011). In addition, this paper has argued that an unstructured ensemble approach carries further limitations of its own, including persisting issues of missing variety in scenario assumptions and missing linkages with regional information. Even though these deficiencies are often discussed together with 'ensembles of opportunity', it is important to stress that the causes are distinct. Insofar as WGIII's assessments are situated in an evolving policy context, this observation allows us to consider the reform of unstructured protocols for scenario elicitation as separate from the challenge of diversifying modeling frameworks.

From the viewpoint of philosophical contribution, the previous reconstruction has identified relevant elements for evaluating ensemble approaches in the climate sciences. As discussed in section two, current protocols for WGIII's ensemble construction have been shaped by the IPCC's institutional position at the science-policy interface, particularly its commitment to delivering a 'policy-neutral' assessment. This has resulted in a specific configuration of construction protocols and an accompanying methodology for assessment. Alongside the dimensions of *ownership* of scenarios and *logic* of assessment, this paper has stressed the need to include *purpose* as a further key dimension (Fig. 1). The rationale for this additional layer of analysis is that it allows to evaluate the performance of an ensemble assessment in relation to the evolving interests of policy users. We are therefore able to translate questions that resist univocal interpretation, such as 'is WGIII's assessment of policy options *objective*?' or 'Is it *policy-neutral*?', with less ambiguous and potentially testable questions: 'is this specific configuration of ownership and logic instantiated by current ensemble methods serving the purposes of intended users?' And *which* of these purposes in particular?'.[10] By showing how to replace more ambiguous terms with less ambiguous ones, the present analysis contributes to progress in the methodological arena.

The substantive question as to whether WGIII should reform its unstructured approach is one that this discussion leaves open. Although persistent missing variety and missing linkages speak in favor of reform, one of the messages of the previous discussion is that finding a replacement for the unstructured approach is by no means straightforward. The latter has so far allowed for an assessment with high potential for delivering policy-relevant information, while deferring ownership of ensemble contents to the modeling community. The result is an impressive, even if imperfect, balance between the dimensions of ownership, logic, and purpose. An implication is that, just as we cannot reduce the reasons for the unstructured approach to solely institutional or epistemic aspects, but rather to a combination of them, so we should not approach the reform question with a single dimension of evaluation in mind. Accordingly, while it is important to point out the opportunities for further learning that come from reform, these broadly logical aspects should be accompanied by a consideration of the institutional aspects, including the consequences of abandoning minimal ownership, as well as the purposes that would be prioritized by the reform. Similarly, institutional considerations cannot be fully addressed without also anticipating the consequences of reform on the logical dimension of assessment. In this sense, the question whether there is a better option can only be answered through a comprehensive outlook.

[10] This conclusion should be welcomed by defenders of 'adequacy-for-purpose' accounts of model evaluation, e.g., Parker (2020).

Declarations of interest

None.

Acknowledgments

This paper has greatly benefitted from discussions with Alaa Al Khourdajie, Kevin Elliott, Rawad El Skaf, Mathias Frisch, Sara Giarola, Alessandro Giordani, Margherita Harris, Leah Henderson, Tarja Knuuttila, Vincent Lam, Giacomo Marangoni, Luis Mirales-Flores, Massimo Tavoni, Henrik Thoren, and Giovanni Valente. Part of the writing of this paper took place during a productive visit at the SOCRATES center in Hannover, Germany.

Data availability

No data was used for the research described in the article.

References

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*, Article 100126.

Asayama, S., De Pryck, K., & Hulme, M. (2022). Controversies. In K. De Pryck, & M. Hulme (Eds.), *A critical assessment of the intergovernmental panel on climate change*. Cambridge: Cambridge University Press.

Beck, M., & Kruger, T. (2016). The epistemic, ethical, and political dimensions of uncertainty in integrated assessment modeling. *Wiley Interdisciplinary Reviews Climate Change, 7*(5), 627–645.

Beck, S., & Mahony, M. (2019). The IPCC and the politics of anticipation. *Nature Climate Change, 7*, 311–313.

Bosetti, V., Marangoni, G., Borgonovo, E., Diaz Anadon, L., Barron, R., McJeon, H., Politis, S., & Friley, P. (2015). Sensitivity to energy technology costs: A multi-model comparison analysis. *Energy Policy, 80*, 244–263. May 2015.

Braunreiter, L., van Beek, L., Hajer, N., & van Vuureen, D. (2021). Transformative pathways – Using integrated assessment models more effectively to open up plausible and desirable low-carbon futures. *Energy Research & Social Science, 80*, Article 102220, 2021.

Brutschin, E., Pianta, S., Tavoni, M., et al. (2021). A multidimensional feasibility evaluation of low-carbon scenarios. *Environmental Research Letters, 16*, Article 064069.

Carley, S., & Konisky, D. (2020). The justice and equity implications of the clean energy transition. *Nature Energy, 5*, 569–577.

Cointe, B., Cassen, C., & Nadai, A. (2021). Organizing policy-relevant knowledge for climate action: Integrated assessment modeling, the IPCC, and the emergence of a collective expertise on socioeconomic emission scenarios. *Science and Technology Studies, 32*(4), 36–56.

De Pryck, K. (2021). Controversial practices: Tracing the proceduralization of the IPCC in time and space. *Global Policy, 12*(7), 80–89.

Dekker, M., Andries, H., van den Berg, M., Daioglou, V., et al. (2023). Spread in climate policy scenarios unravelled. *Nature, 624*, 309–316.

Dethier, C. (2022). When is an ensemble like a sample? "Model-based" inferences in climate modeling. *Synthese, 200*, 52.

Diaz, D., & Moore, F. (2017). Quantifying the economic risks of climate change. *Nature Climate Change, 7*, 774–782.

Doury, A., Somot, S., Gadat, S., et al. (2023). Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach. *Climate Dynamics, 60*, 1751–1779.

IPCC. (2014). Climate change 2014. Mitigation of climate change. In O. Edenhofer (Ed.), *Contribution of working group III to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge, UK: Cambridge University Press.

Edenhofer, O., & Kowarsch, M. (2015). Cartography of pathways: A new model for environmental policy assessments. *Environmental Science & Policy, 51*, 56–64.

Elliott, K. (2025). Characterizing the value-free ideal: From a dichotomy to a multiplicity. In Tsou, et al. (Eds.), *Values, pluralism, and pragmatism: Themes from the work of Matthew J. Brown*. Cham: Springer.

Eyring, V., Bony, S., Meehl, G., et al. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development, 9*, 1937–1958.

Frisch, M. (2013). Modeling climate policies: A critical look at integrated assessment models. *Philosophy & Technology, 26*(2), 117–137.

Giarola, S., Chiani, L., Drouet, L., Marangoni, G., Nappo, F., Muttarak, R., & Tavoni, M. (2024). Underestimating demographic uncertainties in the IPCC synthesis process. *Npj Climate Action, 3*, 71.

Girod, B., Wiek, A., Mieg, H., & Hulme, M. (2009). The evolution of the IPCC's emission scenarios. *Environmental Science & Policy, 12*(2), 103–118.

Guivarch, C., Le Gallic, T., Bauer, N., Fragkos, P., Huppmann, D., et al. (2021). Using large ensembles of climate change mitigation scenarios for robust insights. *Nature Climate Change, 12*, 428–435.

Harmsen, M., Kriegler, E., van Vuureen, D., van der Wijst, K., Luderer, G., et al. (2021). Integrated assessment model diagnostic: Key indicators and model evolution. *Environmental Research Letters, 16*(5), Article 054046.

Harris, M., & Frigg, R. (2023). Climate models and robustness analysis – Part I: Core concepts and premises. In G. Pellegrino, & M. Di Paola (Eds.), *Handbook of philosophy of climate change*. Cham: Springer.

Havstad, J., & Brown, M. (2017). Neutrality, relevance, prescription, and the IPCC. *Public Affairs Quarterly, 31*(4), 303–324.

Hermansen, E., Boasson, E., & Peters, G. (2023). Climate action post-Paris: How can the IPCC stay relevant? *Npj Climate Action, 2*, 30.

Hickel, J., Brockway, P., Kallis, G., et al. (2021). Urgent need for post-growth climate mitigation scenarios. *Nature Energy, 6*, 766–768.

Hollnaicher, S. (2025). *Assessing feasibility with value-laden models*. Cham: Springer.

Hulme, M. (2022). Scientific consensus-seeking. In K. De Pryck, & M. Hulme (Eds.), *A critical assessment of the intergovernmental panel on climate change*. Cambridge: Cambridge University Press.

IPCC. (2005a). *IPCC expert meeting on emission scenarios*. Washington, DC. 12-14 January.

IPCC. (2005b). *Workshop on new emission scenarios*. Laxenburg, Austria. 29 Jun-1 July.

IPCC. (2008). Towards new scenarios for analysis of emissions, climate change, impacts and response strategies. In *IPCC expert meeting report*. Geneva: Intergovernmental Panel on Climate Change.

IPCC. (2023). *IPCC workshop on the use of scenarios in the sixth assessment report and subsequent assessments*. Bangkok, Thailand. 25-27 April.

IPCC. (2024). The intergovernmental panel on climate change. Retrieved from the IPCC website: https://www.ipcc.ch.

Jebeile, J. (2020). Values and objectivity in the intergovernmental panel on climate change. *Social Epistemology, 34*(5), 453–468.

Jebeile, J., & Crucifix, M. (2020). Multi-model ensembles in climate science: Mathematical structures and expert judgments. *Studies in History and Philosophy of Science, 83*, 44–52.

Kallis, G., Kostakis, V., Lange, S., Muraca, B., Paulson, S., & Schmelzer, M. (2018). Research on degrowth. *Annual Review of Environment and Resources, 43*, 291–316.

Katzav, J., Dijkstra, H., & de Laat, A. (2012). Assessing climate model projections: State of the art and philosophical reflections. *Studies in History and Philosophy of Modern Physics, 43*(4), 258–276.

Kc, S., & Lutz, W. (2017). The human core of the shared socio-economic pathways: Population scenarios by age, sex, and level of education for all countries to 2100. *Global Environmental Change, 42*, 181–192.

Keysser, L., & Lenzen, M. (2021). 1.5°C degrowth scenarios suggest the need for new mitigation pathways. *Nature Communications, 12*, 2676.

Kikstra, J., Nicholls, Z., Smith, C., Lewis, J., Lamboll, R., et al. (2022). The IPCC sixth assessment report WGIII climate assessment of mitigation pathways: From emissions to global temperatures. *Geoscientific Model Development, 15*, 9075–9109.

Koberle, A., Daioglou, V., Rochedo, P., Lucena, A., Szklo, A., Fujimori, S., et al. (2022). Can global models provide insights into regional mitigation strategies? A diagnostic model comparison study of bioenergy in Brazil. *Climatic Change, 170*, 2.

Kowarsch, M. (2016). *A pragmatist orientation for the social sciences in climate policy*. Cham: Springer (Boston Studies in History and Philosophy of Science).

Kowarsch, M., Garard, J., Riousset, P., Lenzi, D., Dorsch, M., Knopf, B., Harrs, J., & Edenhofer, O. (2017). Scientific assessments to facilitate deliberative policy learning. *Palgrave Communications, 2*, Article 16092.

Krey, V., Guo, F., Kolp, P., Zhou, W., Schaeffer, R., Awasthy, A., Bertram, C., et al. (2019). Looking under the hood: A comparison of techno-economic assumptions across national and global integrated assessment models. *Energy, 172*, 1254–1267.

Kriegler, E., Peterman, N., Krey, V., Schwanitz, V., Ludered, G., Ashina, S., Bosetti, V., et al. (2015). Diagnostic indicators for integrated assessment models of climate policy. *Technological Forecasting and Social Change, 90*(A), 45–61.

Lenhard, J., & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Modern Physics, 41*, 253–262.

Lunt, D. J., Otto-Bliesner, B. L., Brierley, C., et al. (2024). Paleoclimate data provide constraints on climate models' large-scale response to past CO2 changes. *Communications Earth & Environment, 5*, 419.

Mckay, D., Staal, A., Abrams, J., Winkelmann, R., et al. (2015). Exceeding 1.5°C global warming could trigger multiple climate tipping points. *Science., 377*, 6611.

Mintz-Woo, K. (2021). The ethics of measuring climate impacts. In T. Letcher (Ed.), *The impacts of climate change* (pp. 521–535). Elsevier.

Minx, J., Callaghan, M., Lamb, W., Garard, J., & Edenhofer, O. (2017). Learning about climate change solutions in the IPCC and beyond. *Environmental Science & Policy, 77*, 252–259.

Nakicenovic, N., & Swart, R. (2000). *Special report on emission scenarios*. Cambridge, UK: Cambridge University Press.

Nordhaus, W. (2008). *A question of balance*. New Haven: Yale University Press.

OECD. (2024). Economic outlook. *ume 2024*(Issue 2). Available at: oecd.org.

Parker, W. (2010). Whose probabilities? Predicting climate change with ensembles of models. *Philosophy of Science, 77*, 985–997.

Parker, W. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science, 78*(4), 579–600.

Parker, W. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science, 87*, 457–477.

Pedersen, J., van Vuureen, D., Gupta, J., Santos, F., et al. (2022). IPCC emission scenarios: How did critiques affect their quality and relevance? 1990-2022. *Global Environmental Change, 72*, Article 102538.

Peng, W., Iver, G., Bosetti, V., Chaturvedi, V., et al. (2021). Climate policy models need to get real about people – Here's how. *Nature, 594*, 174–176.

Peters, G., Al Khourdaje, A., Sognnaes, I., & Sanderson, B. (2023). AR6 scenarios database: An assessment of current practices and future recommendations. *Npj Climate Action, 2*, 31.
IPCC. (2007). Climate change 2007: Synthesis report. In Peuchari (Ed.), *Contributions of working groups I, II, and III to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge, UK: Cambridge University Press.
Piontek, F., Drouet, L., Emmerling, J., Kompas, T., et al. (2021). Integrated perspectives on translating biophysical to economic impacts of climate change. *Nature Climate Change, 11*, 563–572.
Pirani, A., Fuglestvedt, J., Byers, E., O'Neill, B., Riahi, K., Lee, J., Marotzke, J., Rose, S., Schaeffer, R., & Tebaldi, C. (2024). Scenarios in IPCC assessments: Lessons from AR6 and opportunities for AR7. *Npj Climate Action, 3*, 1.
Riahi, K., van Vuuren, D., Kriegler, E., Edmonds, J., O'Neill, B., et al.M. (2017). The shared socioeconomic pathways and their energy, land use, and greenhouse gas emission implications: An overview. *Global Environmental Change, 42*, 153–168.
Saltelli, A., Puy, A., & Di Fiore, M. (2024). Mathematical models: A state of exception. *International Review of Applied Economics, 39*, 862–879.
IPCC. (2022). Climate change 2022: Mitigation of climate change. In P. R. Shukla (Ed.), *Contribution of working group III to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge, UK: Cambridge University Press.
Sognnaes, I., & Peters, G. (2025). Influence of individual models and studies on quantitative mitigation findings in the IPCC sixth assessment report. *Nature Communications, 16*, 8343.
Tavoni, M., & Tol, R. (2010). Counting only the hits? The risk of underestimating the costs of stringent climate policy. *Climatic Change, 100*, 768–778.
Tavoni, M., & Valente, G. (2022). Uncertainty in integrated assessment models of climate change. *Perspectives on Science, 30*(2), 321–351.
Van Beek, L., Hajer, M., Pelzer, P., van Vuuren, D., & Cassen, C. (2020). Anticipating futures through models: The rise of integrated assessment modeling in the climate science-policy interface since 1970. *Global Environmental Change, 65*, Article 102191.
Van Vuuren, D., Edmonds, J., Kainuma, M., Riahi, K., et al. (2011). The representative concentration pathways: An overview. *Climatic Change, 109*, 5.
Wang, H., Lei, Z., Zhang, X., Zhow, B., & Peng, J. (2019). A review of deep learning for renewable energy forecasting. *Energy Conversion and Management, 198*, Article 111799.
Way, R., Ives, R., Mealy, P., & Doyne Farmer, J. (2022). Empirically grounded technology forecasts and the energy transition. *Joule, 6*, 2057–2082.
Weyant, J. (2017). Some contributions of integrated assessment models of global climate change. *Review of Environmental Economics and Policy, 11*(1), 115–137.
Wilson, C., Guivarch, C., Kriegler, E., Ruijven, B., van Vuuren, D., Krey, V., et al. (2021). Evaluating process-based integrated assessment models of climate change mitigation. *Climatic Change, 166*, 3.
Winsberg, E. (2018). What does robustness teach us in climate science: A re-appraisal. *Synthese, 198*, 5099–5122.
Yue, X., Pye, S., DeCarolis, J., Li, F., Rogan, F., & O'Gallachoir, B. (2018). A review of approaches to uncertainty assessment in energy system optimization models. *Energy Strategy Reviews, 21*, 204–217.
Zebrowski, P., Dieckmann, U., Brännström, Å., Franklin, O., & Rovenskaya, E. (2022). Sharing the burdens of climate mitigation and adaptation: Incorporating fairness perspectives into policy optimization models. *Sustainability, 14*, 3737.
Zimm, C., Mintz-Woo, K., Brutschin, E., Hanger-Kopp, S., et al. (2024). Justice considerations in climate change research. *Nature Climate Change, 14*, 22–30.