# Title:

A weighting framework to improve the use of emissions scenario ensembles of opportunity

Authors:   
Hamish Beath1, Chris Smith2,3, Jarmo Kikstra1,3,4, Mark Dekker5,6, Matthew J. Gidden3,7, Joeri Rogelj1,3,4

# Affiliations:

1Centre for Environmental Policy, Faculty of Natural Sciences, Imperial College, London, United Kingdom

2Department of Water and Climate, Vrije Universiteit Brussel, Brussels, Belgium

3International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

4The Grantham Institute for Climate Change and the Environment, Imperial College London, London, United Kingdom

5 PBL Netherlands Environmental Assessment Agency, The Hauge, Netherlands

6Copernicus Institute for Sustainable Development, Utrecht University, Utrecht, Netherlands

7Climate Analytics, Berlin, Germany

# Abstract:

Integrated assessment models (IAMs) produce large ensembles of socioeconomic scenarios that are used profusely in climate change research. The Intergovernmental Panel on Climate Change (IPCC), non-governmental organisations or national climate committees often rely on ensemble statistics to identify mitigation strategies and set climate targets. A limitation of such evidence is the opportunistic nature of scenario ensembles: they are an unstructured, serendipitous collection of evidence. Drawing on concepts from physical climate science and ensemble analysis, we present a novel approach for the flexible, multidimensional weighting of emission scenario data that accounts for relevance, quality, and diversity. Our illustrative application to the latest IPCC scenario database demonstrates a reduction in dominance of highly represented models and studies, and sees net-zero emission milestones differ to those originally reported. Our framework formalises decisions otherwise made in an ad hoc manner, providing a tool contributing to the broader challenge of assessing ensembles of opportunity.

# Main text:

Since the publication of the IPCC Special Report on Emissions Scenarios1 in 2000, integrated assessment models (IAM) have become central tools for exploring emissions and climate futures in climate research. Despite already being part of the IPCC Third2 and Fourth3 Assessments in 2001 and 2007, it was the IPCC’s Fifth Assessment4, the Special Report on Global Warming of 1.5°C5 (SR1.5) and the IPCC Sixth Assessment6 (AR6) that solidified the use of large scenario ensembles for the assessment of global mitigation strategies. With their expanded use also came an improved understanding and communication of their limitations7,8.

Scenarios that are collected as part of IPCC or other exercises9–11 represent ensembles of opportunity: a serendipitous collection of scenario data that is unstructured7 and in which the scenarios that are ultimately included vary in their purpose, design, comprehensiveness, coverage, quality and other characteristics. One key limitation of their use is that shortcomings or biases present in the collected ensemble can be propagated by subsequent secondary analysis7,8. Biases include dominance of specific models and intercomparison projects, which can represent a lack of diversity in organisational, or regional composition12–14. Unless corrected for, this could lead to spurious or biased results.

Typical shortcomings or biases relate to three main issues: scenario relevance, quality, and diversity. Relevance refers to whether a scenario, through the structural properties of the underlying model, its design, and outcome characteristics is relevant to the question that is being investigated by the secondary analysis. This can include the estimated level of global warming avoided by the scenario15, assessments of the feasibility of its described transitions16, or even subjective – but transparently communicated – preferences about technologies or strategies. Quality refers to whether the implementation and execution of the modelling lives up to pre-defined standards set out by the secondary analysis. These standards typically refer to technical modelling aspects such as the accuracy of historical data, time resolution, or plausibility of near-term trends and resource use. Diversity refers to the degree of additional information a scenario communicates compared to other members in the ensemble of opportunity. Not accounting for the latter can result in statistics across a scenario ensemble being too narrow or overconfident towards the results of a single model, modelling team or modelling exercise17.

In the past, such issues have been dealt with on an ad-hoc basis. For example, the IPCC SR1.55 checked for the completeness of variables available in scenarios, whether data is reported until 2100, or whether reported historical- or near-future data is consistent with observations9,18. For the assessment of global emission characteristics of pathways aligned with 1.5°C, SR1.5 also excluded 13 scenarios from a single modelling group19 because they included virtually no variation in emissions and their inclusion would have biased descriptive emission statistics. Criteria for including or excluding scenarios depend on context, which explains why these 13 scenarios were still included in the analysis of aspects other than the evolution of emissions in the SR1.5 report. Similar ad-hoc considerations were applied in the AR6 mitigation assessment of the IPCC6,8.

Other climate research communities have also grappled with similar issues. The Earth System Modelling community has established methods to down-weight models based on their similarity to other models20–22, as well as for model quality measured as performance relative to historical observations22,23. In emission scenario ensemble analysis, issues of scenario similarity have been considered24, but as of yet not systematically addressed.

Here, we use efforts from various communities20–24 as a starting point to develop and present a scenario-weighting method, applicable to scenarios from IAMs or energy-economic models. Our method addresses issues of relevance, quality and diversity (Fig. 1), and provide a systematic approach that unifies and expands on previous ad-hoc scenario assessment decisions1–6. Further, it presents an alternative to other approaches that reweight or present summary statistics accounting for common model or intercomparison study14,17. We apply it to the latest available IPCC AR6 scenario ensemble10 to illustrate its use in practice. We discuss the strengths and weaknesses of such an approach, with a view to understand the limits to its current applicability.

## A generalised scenario weighting framework

Our starting point takes a generalised view that each scenario in an ensemble of opportunity must be assigned a weight for subsequent analysis. In the past, re-analysis of scenario ensembles would decide to include or exclude scenarios from their analysis, effectively assigning weights of 1 or 0. Here we formalise this step. Each scenario *i* contributing to the assessment and member of the ensemble of opportunity *E* is assigned a generalized weight *gwi*:

where *0 ≤ R(i) ≤ 1* is a measure of how relevant a scenario is to answer a specific question (relevance weighting), *0 ≤ Q(i) ≤ 1* is a metric of scenario quality (quality weighting) and *D(i) >= 1* is a metric of scenario uniqueness (diversity weighting). This ensures *gwi* has a weight between 0 and 1. Scenarios with lower weights are less relevant, of lower quality, or very similar to others, and hence contribute less towards the final statistics of the ensemble. A detailed explanation of each part of the weighting framework is provided in the Methods.

A diagram of weighting and diversity

AI-generated content may be incorrect.

**Figure 1 | A scenario weighting framework for the analysis of scenario ensembles of opportunity.** Schematic showing how the unstructured, serendipitous collection of evidence in a scenario ensemble of opportunity can be translated in a weighted ensemble, accounting for a scenario’s relevance, quality and diversity.

## Application to IPCC AR6 database

To illustrate the usefulness of our weighting framework, we apply it to the scenarios included in the IPCC AR6 Scenario Database10 and look at the influence on key scenario assessment outcomes of the AR66. We focus this illustrative application on scenarios that limit warming to 1.5°C with no or limited overshoot (C1), the subset of C1 scenarios also reaching net zero greenhouse gas (GHG) emissions over the course of the 21st century25 (C1a), and scenarios that return warming to 1.5°C after a high overshoot (C2). We apply binary question-specific relevance weights R(i) to select the desired scenarios, although we demonstrate an alternative continuous relevance weighting approach in Supplementary Results 1. We apply a continuous quality weighting, , based on the IPCC AR6 vetting procedure26 (see Methods). Diversity weighting D(i) considers the similarities between scenarios in four key dimensions (emissions, economy, mitigation strategy, and energy) and a total of 15 variables (Methods Table 1). Due to correlations between variables, we use 8 representative variables for the main results (see Methods). We primarily focus on diversity weighting, the most complex dimension, considered separately from continuous quality weighting for interpretability. The outcomes represent one set of weighting inputs; further examples are provided in the Supplementary Results \_.

## reweighting sensitivity of benchmarks

Applying question-specific relevance, quality, and diversity weighting to the IPCC AR6 database reveals a range of final weights across scenarios (Fig. 2a). Higher-emission scenarios in the IPCC AR6 scenario database (IPCC categories C3 and higher) show less relative diversity than mitigation scenarios in the C1 and C2 categories and hence see a larger number of scenarios with lower weights. This feature can be understood when considering that the scenario compilation by IPCC AR6 aimed to explore diverse mitigation futures. Quality weights (Fig. 2d) have a spikey distribution, as scenarios from specific intercomparison projects and models have harmonised input data.

Visually, 2050 CO2 emission and peak warming ranges of C1 and C2 show little change (Figs. 2b, c e & f). Under diversity weighting, the median 2050 CO2 emissions reduces slightly for C1 and C1a, while all categories show wider IQRs and 5th-95th ranges (Fig. 2b). Median peak temperatures barely change for C1 and C1a but fall for C2; all categories have wider interquartile ranges. For quality weighting (Figs 2e & f), the impact on medians varies by categories.

A group of graphs showing different types of weight

AI-generated content may be incorrect.

**Figure 2 | Diversity weighting of scenarios available in the IPCC AR6 scenario database. a.** Diversity weights *D(i)* for scenarios per IPCC scenario category; **b.** unweighted and diversity weighted distributions for 2050 CO2; **c.** unweighted and diversity weighted distributions for median peak temperature (MAGICCC); **d** Quality weights *Q(i)* for scenarios per IPCC scenario category; **e.** unweighted and diversity weighted distributions for 2050 CO2  **f.** unweighted and quality weighted distributions for median peak temperature (MAGICCC). For panels b, c e and f, the darker coloured distributions for each temperature represented weighted ones. The wides shaded ranges represent the 5th and 95th percentiles, with the narrow ones the interquartile range. Weighted and unweighted quantiles are computed using the same non-interpolating approach (see Methods).

For other outcomes of our diversity weighting, including net-zero years (Fig. 3a) and the trajectory of certain variables (Fig. 3b-l, Supplementary Figures 2 & 3), some metrics show noteworthy changes. For C1 scenarios, the median year of reaching net-zero GHG emissions is advanced by a decade (from 2098 to 2088); but C1a scenarios, it is advanced by much less (2071 to 2068). For C2 scenarios the median net-zero GHG date moves earlier, but by less than a year, while the upper quartile reduces by 2 years (from 2084 to 2082). Median net-zero CO₂ years remain virtually unchanged (C1/C1a <1 year earlier; C2 unchanged).

For many of the 15 variables reported by all scenarios, changes to timeseries are negligible (Supplementary Figures 2 & 3). However, some mitigation-relevant variables show visible changes, highlighting where the ensemble is sensitive to diversity weighting. For example, the median and upper quartile values for carbon capture and storage (CCS) are higher after 2050 for C1 and C2. In 2070, medians are 13% higher for C1 and 7% for C2.

The data suggest that reweighting has greater impact where outcomes are widely spread and unevenly distributed. However, this is not systematic. For example, Primary Energy from Gas for C1 scenarios (Fig. 2d) exhibits a wide interquartile range, with the 75th percentile diverging from the median after 2050, but little change in the reweighted ensemble. This highlights that the effect also depends on similarity within and across temperature categories in the wider ensemble, and the parameters used for determining similarity (Supplementary Results 2).

Although only illustrative, these observations indicate that a more balanced consideration of the diversity of the scenario evidence suggests small but measurable changes to climate action benchmarks. Given our observed change in net zero GHG years, the IPCC AR6 scenario assessment would in this instance be understood as conservative regarding mitigation benchmarks  compatible with the Paris Agreement*.*

We explore further diversity weighting procedures and their impact in the Supplementary Results 2 and 3. These further explorations highlight which changes broadly persist across inputs (e.g., reduction in net zero GHG years, higher CCS), and which vary depending on the choice of weighting inputs (e.g., Primary energy from Gas). Further, some variables such as CCS and Primary Energy from Nuclear appear more sensitive to changes in sigma values (Supplementary Figs 4 & 8). The impact of continuous quality weighting is shown in Supplementary Figs. 5–6.

A collage of graphs and diagrams

AI-generated content may be incorrect.

**Figure 3 | Key assessment quantities in diversity weighted and unweighted scenario category ranges from the IPCC AR6 scenario database. a.** Net zero CO2 and greenhouse gas (GHG) years; **b.** Carbon sequestration from carbon capture and storage for C1; **c.** Carbon sequestration from carbon capture and storage for C2 **d. & e,** Primary Energy from Gas timeseries for C1 and C2 respectively; **e & h**, Primary Energy from Nuclear timeseries for C1 and C2 respectively; **f & I,** Carbon Price timeseries for C1 and C2 respectively. Our plots show data from three scenario categories: scenarios limiting warming to 1.5°C with no or limited overshoot (IPCC AR6 category C1) n=97, the subset of these scenarios also reaching net zero GHG emissions in the second half of the century (IPCC AR6 category C1a) n=50, and scenarios returning warming to below 1.5°C in 2100 after a high (0.1–0.3°C) overshoot (IPCC AR6 category C2) n=133. Individual scenarios are shown as dots. Violins show unweighted and weighted distributions, respectively, which are distinguished through lighter and darker areas. Dashed and dotted lines in the plots indicate key distribution statistics as defined by the legend in panel **a**. For timeseries plots, the coloured shaded areas represent the interquartile range of the reweighted distribution, with grey shaded areas representing the unweighted interquartile range. The coloured dashed lines represent weighted medians, whilst black dashed lines represent unweighted medians, as shown in the legend in panel h. Weighted and unweighted quantiles are computed using the same non-interpolating approach (see Methods).

## Exploring robustness

It is well-established that IAMs behaviour is to a large degree determined by their framework logic, structure and input assumptions27,28. Equally, model intercomparison studies use harmonised scenario design assumptions, dominating key scenario insights29. Understanding whether insights are robust across model typographies, structures and scenario design assumptions is therefore important. Jack-knife resampling allows the estimation of bias of a test statistic and has been previously applied to scenario ensembles1 17. Such a test calculates how outcomes *Vd* vary across the ensemble, *E* when iteratively removing scenarios according to each instance *d* in resampling dimension *D* (e.g., removing specific modelling frameworks).

We perform two jack-knife resampling tests to see how outcomes are affected by the diversity reweighting of scenarios in the ensemble. We first test the sensitivity to unbalanced contributions by model frameworks, and second, the sensitivity to contributions by projects (Fig. 4, and Supplementary Tables S1 and S2). Project refers to model intercomparison studies and other coordinated modelling efforts.

In some instances, diversity weighting alters ensemble sensitivity to removal of projects or model frameworks, including narrower ranges under our weighted examples. For example, 5th and 25th percentiles for net-zero GHG years for C1 scenarios (Fig 4a). This indicates that these earlier shifts from diversity weighting are robust. Some statistics show greater variability post-weighting. For example, the C1 median net zero GHG for project and model, and the 95th percentile of net zero CO2 year for project show wider ranges than their unweighted counterparts (Fig 4a, b). This indicates greater sensitivity to model or project composition, and less stability in the weighted ensemble. However, there are nuances when interpreting wider ranges. For C1 GHG medians, the lower bound of the project-based jack-knife values occurs when removing the ENGAGE project. ENGAGE is the most prevalent C1 project, with its share reduced in the weighted ensemble (Fig 5b). In this instance, the reduced lower bound value (and thus wider range) for the weighted jack-knife median net-zero GHG year reflects higher diversity weights of scenarios from less prevalent projects.

**A diagram of different colored lines

AI-generated content may be incorrect.**

**Figure 4 | Interplay between weighting of scenario ensembles and jack-knife resampling outcomes. a.** Net zero GHG years; **b.** Net zero GHG years CO2. Statistics are documented in the figure legend, which should be used for interpreting both panels a and b. For each statistic of the unweighted and the weighted ensemble, the jackknife resampling illustrates the potential variation in the estimators. The jack-knife resampling based on model framework iteratively excludes each respective larger modelling framework group as reported in the IPCC AR6 database. The resampling based on project iteratively excludes each respective contributing project or individual study. See Supplementary Tables S1 and S2 for a list. Both weighted and unweighted quantiles are computed using the same non-interpolating approach (see Methods).

## Impact on model and project dominance

Accepting that scenario similarity is due in part to commonality of model or intercomparison study, one would expect weighting to even out relative contributions to summary statistics. Our illustrative diversity weighting indeed shows small but consistent adjustments in the relative contributions of projects and models (Figs 5b, d). We apply the Herfindahl-Hirschman Index (HHI)31, an indicator used for measuring concentration of dataset components. A lower HHI in the diversity weighted summary statistics than in the unweighted ones means less dominance of large components (see Methods). For model framework and model type, when compared to the unweighted ensemble, in the weighted ensemble there is an 8-9% reduction in HHI in C1 and C2 categories, and for project, an 11% reduction in the HHI for C1, and a 6% reduction for C2 (Supplementary Figure 11). Sensitivity tests show these general trends are robust to a range of alternative inputs to the diversity weighting scheme (see Supplementary Results 4). Although a reduced HHI equals a more balanced ensemble, this does not mean that the result is guaranteed to be unbiased, as weighting does not resolve gaps in coverage.

Focussing on the impact on specific scenarios, the prevalence of models and projects can explain which scenarios achieve high or low weights. The top-10 diversity weighted scenarios for both C1 and C2 originate from a model framework and/or project with low prevalence (Figs 5a&c). Likewise, the bottom-10 scenarios originate from either a project or model with high prevalence, or in many instances both. The bottom five scenarios (Fig. 5c) for C2 originate from MESSAGE (second most prevalent model for C2) and are variants with 50 Gt CO2 differences (400-650) in full-century carbon budgets (Fig. 5). In our C2 top 10, there are three scenarios which are also carbon budget variants; however, these come from the AIM model framework (3rd least prevalent) and have wider carbon budget differences (100Gt CO2).

A screenshot of a diagram

AI-generated content may be incorrect.

**Figure 5 | Insights into how scenarios diversity weighted according to model frameworks and projects. a.** connectedrank plot showing C1 scenarios grouped by model (left ranking), diversity weight (central ranking) and project (right ranking). **b** the unweighted and weighted proportions of each project, model type and model framework for C1 scenarios; **c,** equivalent connected rank plot from panel **a**, but for C2 scenarios; d, stacked bars showingthe unweighted and weighted proportions of each project, model type and model framework for C2 scenarios. For the rank plots (panels **a** & **c**), in the left and right ranks, scenarios are grouped with the largest model/project at the bottom. Thicker more visible connection lines are shown the for scenarios in the highest or lowest 10 scenarios by diversity weighting. For the stacked bars (panels **b** and **d)** the lighter shaded bars are unweighted, and the darker shaded bars are diversity weighted. Only the 8 largest shares shown, with other grouped together as “Other”, as visible on the key.

## Reweighting for improved scenario assessment

This weighting framework presented here is designed to be simple and transparent. Whilst our illustrative application to AR6 scenarios is by no means exclusive, it does indicate that reweighting for diversity may lead to a strengthening of certain climate action milestones. Additionally, sensitivity analysis indicates that some changes, such as earlier net zero GHG years for C1 scenarios, are maintained across a range of diversity weighting inputs. The diversity weighting procedure reduces the dominance of prevalent models and intercomparison studies; with this finding being robust to a range of sensitivities (Supplementary Results 4).

The framework could be applied to explore further dimensions. Scenario quality or relevance weighting could be used to integrate questions around sustainable development within pathway assessments, for example, with limits for biomass or carbon-dioxide removal as part of the broader United Nations Sustainable Development Agenda5,32,33. Alternatively, it could be used to consider questions regarding feasibility of scenario pathways across multiple dimensions16,34. Depending on the needs of the user, quality weighting criteria may extend to structural model characteristics. For questions relating to use of a specific technology for example, representation within models may influence quality weights. Practically this could extend to considering factors such as temporal resolution or adoption heterogeneity. Finally, reweighting of scenario ensembles might also be useful to improve methods that derive relationships from scenario ensembles, such as for the completion of missing species of GHG emissions35.

Beyond global IAM emission scenarios, the framework could also be applied to ensembles of national or sectoral pathways36. However, the context for such pathways is quite different, as there is rather a dearth than an excess of national pathways for any specific country or for any specific sector. Weighting procedures could therefore rather focus on relevance and quality characteristics, especially given inherent heterogeneity in scenarios representing different countries or regions.

Our weighting framework, offers promise, but no silver bullet. Its flexibility allows it to be applied to specific research questions. However, expert judgment exists in the identification and choice of the weighting approaches, distance metrics or relevant indicators, and in interpretation of weighted ensembles. In this paper, we demonstrate how some expert judgments might influence outcomes. Although flexibility of our framework is a strength, proliferation of weighting approaches could lead to reduced accessibility and transparency. It is therefore vital that there is clear and consistent reporting of applied weightings and justification of their underlying expert judgments. Acknowledging this, the structured nature of the framework can help to ensure the necessary expert judgment is well-guided and transparent.

Moreover, transparency is essential: scenario data have a range of applications, some of which may incentivise selective presentation. Without a clear framework and best-practice guidance, re-weighting scenario ensembles could enable parties to present ensemble statistics to suit a preferred narrative.

Scenario ensembles are commonly used to calculate headline statistics, rather than to present the intricacies of their composition8. Our reweighting framework aims to improve the robustness of ensemble statistics, allowing the user to remove redundancy, and apply tailored quality and relevance criteria. However, even after reweighting, scenario ensembles are incomplete, and there is thus a danger that reweighting may produce a different kind of overconfidence in ensemble statistics12. Indeed, reweighting does not identify gaps in the explored solution space, and caution must therefore accompany its use.

As similar scenarios are down-weighted by diversity weighting, a balance must be struck between accounting for repeated instances of scenarios that are fundamentally similar, whilst retaining useful information from areas in which they differ. Near-identical scenarios but from a different model framework may not represent redundancy. Indeed, near-identical Scenarios that push the boundaries of conventional thinking and are less similar to more traditional modelled pathways, including emerging efforts to implement circular economy and degrowth storylines or those that focus on low energy demand37,38 are likely to be weighted higher and contribute more to category statistics under a weighted assessment. The drawback of this is that potential novel and emerging scenario approaches might over-emphasise scenarios with technologies at the boundaries of realism. Indeed, diversity weighting in isolation could incentivise the creation of such scenarios in unintended ways. There is a need to explore extremes and novel approaches39, but not without practical consideration of feasibility.

All three dimensions of the weighting framework that jointly cover aspects of relevance, quality and diversity should be carefully considered. At the same time, scenario ensemble assessments also need to reflect on the meaning and usefulness of headline statistics that this framework improves. In that sense, we here present a first approach that can contribute to a larger toolbox of improved methods and approaches for scenario assessment, but the role of experts will remain as important as ever.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## Acknowledgments

H.B. acknowledges funding from the Hitachi-Imperial Centre for Decarbonisation and Natural Climate Solutions. C.J.S. was supported by a NERC/IIASA collaborative research fellowship (NE/T009381/1). J.S.K. was supported by the Natural Environment Research Council under grant agreement NE/S007415/1. M.J.G. was supported by funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (GENIE – grant agreement No 951542). J.R. and H.B. were supported by the European Union’s Horizon 2020 research and innovation programme (ESM2025 – grant agreement No 101003536) and also acknowledge funding from the Bezos Earth Fund through the Scenario Compass Initiative (Grant G-2023-201305841)

## Author Contributions

C.S., J.K, M.G, and J.R. conceived the original study, with additional study elements conceptualised by H.B. and M.D. H.B. and C.S. performed the experiments, analysed the data and produced visualisations. Code was conceptualised by C.S. with code used for the present study written by H.B. J.R. and H.B. led the writing of the study with contributions from M.D., M.G. and J.K. H.B., C.S., J.K, M.D, M.G and J.R. reviewed the manuscript before submission.

## Competing interests Statement

The authors declare no competing interests.

## References

1. Nakicenovic, N. & Swart, R. *IPCC Special Report on Emissions Scenarios*. (Cambridge University Press, Cambridge, United Kingdom, 2000).

2. Morita, T. *et al.* Greenhouse Gas Emission Mitigation Scenarios and Implications. in *Climate Change 2001: Mitigation. A Report of Working Group III of the Intergovernmental Panel on Climate Change* (eds Chadwick, M. & Parikh, J. K.) 115–166 (Cambridge University Press, Cambridge, UK, 2001).

3. Fisher, B. *et al.* Issues related to mitigation in the long term context. in *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Inter-governmental Panel on Climate Change* (eds Metz, B., Davidson, O. R., Bosch, P. R., Dave, R. & Meyer, L. A.) 169–250 (Cambridge University Press, Cambridge, UK, 2007).

4. Clarke, L. *et al.* Assessing Transformation Pathways. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) 413–510 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014).

5. Rogelj, J. *et al.* Mitigation pathways compatible with 1.5°C in the context of sustainable development. in *Global Warming of 1.5 °C: an IPCC special report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* (eds Flato, G., Fuglestvedt, J., Mrabet, R. & Schaeffer, R.) 93–174 (IPCC/WMO, Geneva, Switzerland, 2018).

6. Riahi, K. *et al.* Mitigation pathways compatible with long-term goals. in *IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Shukla, P. R. et al.) (Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022). doi:10.1017/9781009157926.005.

7. Huppmann, D., Rogelj, J., Kriegler, E., Krey, V. & Riahi, K. A new scenario resource for integrated 1.5 °C research. *Nat. Clim. Change* **8**, 1027–1030 (2018).

8. Guivarch, C. *et al.* Using large ensembles of climate change mitigation scenarios for robust insights. *Nat. Clim. Change* **12**, 428–435 (2022).

9. Huppmann, D. *et al.* *IAMC 1.5°C Scenario Explorer and Data Hosted by IIASA*. (Integrated Assessment Modeling Consortium & International Institute for Applied Systems Analysis, 2018). doi:10.22022/SR15/08-2018.15429.

10. Byers, E. *et al.* AR6 Scenarios Database. Zenodo https://doi.org/10.5281/zenodo.5886912 (2022).

11. Nakicenovic, N., Victor, N. & Morita, T. Emissions Scenarios Database and Review of Scenarios. *Mitig. Adapt. Strateg. Glob. Change* **3**, 95–120 (1998).

12. Peters, G. P., Al Khourdajie, A., Sognnaes, I. & Sanderson, B. M. AR6 scenarios database: an assessment of current practices and future recommendations. *Npj Clim. Action* **2**, 31 (2023).

13. Jaxa-Rozen, M. & Trutnevyte, E. Sources of uncertainty in long-term global scenarios of solar photovoltaic technology. *Nat. Clim. Change* **11**, 266–273 (2021).

14. Van De Ven, D. J. *et al.* Energy and socioeconomic system transformation through a decade of IPCC-assessed scenarios. *Nat. Clim. Change* **15**, 218–226 (2025).

15. Rogelj, J. *et al.* Emission pathways consistent with a 2°C global temperature limit. *Nat. Clim Change* **1**, 413–418 (2011).

16. Brutschin, E. *et al.* A multidimensional feasibility evaluation of low-carbon scenarios. *Environ. Res. Lett.* **16**, 064069 (2021).

17. Sognnaes, I. & Peters, G. P. Influence of individual models and studies on quantitative mitigation findings in the IPCC Sixth Assessment Report. *Nat. Commun.* **16**, 8343 (2025).

18. Forster, P. *et al.* Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development Supplementary Material. in *Global Warming of 1.5 °C: an IPCC special report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* (World Meteorological Organisation, Geneva, Switzerland, 2018).

19. Liu, J.-Y. *et al.* Socioeconomic factors and future challenges of the goal of limiting the increase in global average temperature to 1.5°C. *Carbon Manag.* 1–11 (2018) doi:10.1080/17583004.2018.1477374.

20. Knutti, R. *et al.* A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett.* **44**, 1909–1918 (2017).

21. Sanderson, B. M., Knutti, R. & Caldwell, P. A Representative Democracy to Reduce Interdependency in a Multimodel Ensemble. *J. Clim.* **28**, 5171–5194 (2015).

22. Brunner, L. *et al.* Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth Syst. Dyn. Discuss.* 1–23 (2020) doi:https://doi.org/10.5194/esd-2020-23.

23. Tokarska, K. B. *et al.* Past warming trend constrains future warming in CMIP6 models. *Sci. Adv.* **6**, eaaz9549 (2020).

24. Gütschow, J., Jeffery, M. L., Schaeffer, M. & Hare, B. Extending Near-Term Emissions Scenarios to Assess Warming Implications of Paris Agreement NDCs. *Earths Future* **6**, 1242–1259 (2018).

25. Schleussner, C.-F., Ganti, G., Rogelj, J. & Gidden, M. J. An emission pathway classification reflecting the Paris Agreement climate objectives. *Commun. Earth Environ.* **3**, 1–11 (2022).

26. Guivarch, C. *et al.* IPCC, 2022: Annex III: Scenarios and modelling methods. in *IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022).

27. Kriegler, E. *et al.* Diagnostic indicators for integrated assessment models of climate policy. *Technol. Forecast. Soc. Change* **90, Part A**, 45–61 (2015).

28. Harmsen, M. *et al.* Integrated assessment model diagnostics: key indicators and model evolution. *Environ. Res. Lett.* **16**, 054046 (2021).

29. Rogelj, J. *et al.* A new scenario logic for the Paris Agreement long-term temperature goal. *Nature* **573**, 357–363 (2019).

30. Rogelj, J. *et al.* Zero emission targets as long-term global goals for climate protection. *Environ. Res. Lett.* **10**, 105007 (2015).

31. Rosenbluth, G. Measures of Concentration. in *Business Concentration and Price Policy* 57–99 (National Bureau of Economic Research, Inc, 1955).

32. Roy, J. *et al.* Sustainable Development, Poverty Eradication and Reducing Inequalities. in *Global Warming of 1.5 °C: an IPCC special report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* (eds Krakovska, S., Madruga, R. P. & Sanchez, R.) (World Meteorological Organisation, Geneva, Switzerland, 2018).

33. Deprez, A. *et al.* Sustainability limits needed for CO2 removal. *Science* **383**, 484–486 (2024).

34. Warszawski, L. *et al.* All options, not silver bullets, needed to limit global warming to 1.5 °C: a scenario appraisal. *Environ. Res. Lett.* **16**, 064037 (2021).

35. Lamboll, R. D., Nicholls, Z. R. J., Kikstra, J. S., Meinshausen, M. & Rogelj, J. Silicone v1.0.0: an open-source Python package for inferring missing emissions data for climate change research. *Geosci. Model Dev.* **13**, 5259–5275 (2020).

36. IAMC. Scientific Working Group on National Scenarios. *IAMC: National Scenarios* https://www.iamconsortium.org/scientific-working-groups/national-scenarios/ (2025).

37. Grubler, A. *et al.* A low energy demand scenario for meeting the 1.5 °C target and sustainable development goals without negative emission technologies. *Nat. Energy* **3**, 515–527 (2018).

38. Millward-Hopkins, J. Inequality can double the energy required to secure universal decent living. *Nat. Commun.* **13**, 5028 (2022).

39. McCollum, D. L., Gambhir, A., Rogelj, J. & Wilson, C. Energy modellers should explore extremes more systematically in scenarios. *Nat. Energy* **5**, 104–107 (2020).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## Methods References

40. Kikstra, J. S. *et al.* The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways: from emissions to global temperatures. *Geosci. Model Dev.* **15**, 9075–9109 (2022).

41. Nicholls, Z. *et al.* Cross-Chapter Box 7.1: Physical emulation of Earth System Models for scenario classification and knowledge integration in AR6. in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Masson-Delmotte, V. et al.) (Cambridge University Press, 2021).

42. Beath, H., Mittal, S., Lamboll, R. & Rogelj, J. An exploration and evaluation framework for climate change mitigation scenarios with varying feasibility and desirability. *Environ. Res. Lett.* **20**, 074020 (2025).

43. Colson, A. R. & Cooke, R. M. Expert Elicitation: Using the Classical Model to Validate Experts’ Judgments. *Rev. Environ. Econ. Policy* **12**, 113–132 (2018).

44. Dekker, M. M. *et al.* Identifying energy model fingerprints in mitigation scenarios. *Nat. Energy* **8**, 1395–1404 (2023).

45. Dekker, M. M. *et al.* Spread in climate policy scenarios unravelled. *Nature* **624**, 309–316 (2023).

46. IPCC. Summary for Policymakers. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* 3−32 (2021) doi:10.1017/9781009157896.001.