nature climate change

Article

https://doi.org/10.1038/s41558-026-02565-5

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

Received: 10 December 2024

Accepted: 15 January 2026

Published online: 24 February 2026

Check for updates

Hamish Beath[1] ✉, Chris Smith[2], Jarmo S. Kikstra[1,2,3], Mark M. Dekker[4,5], Matthew J. Gidden[2,6] & Joeri Rogelj[1,2,3] ✉

Integrated assessment models produce large ensembles of socioeconomic scenarios that are used profusely in climate change research. The Intergovernmental Panel on Climate Change (IPCC), non-governmental organizations 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 an 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 formalizes decisions otherwise made in an ad hoc manner, providing a tool contributing to the broader challenge of assessing ensembles of opportunity.

Since the publication of the Intergovernmental Panel on Climate Change (IPCC) Special Report on Emissions Scenarios[1] 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 Third[2] and Fourth[3] Assessments in 2001 and 2007, it was the IPCC Fifth Assessment[4], the Special Report on Global Warming of 1.5 °C (ref. 5) (SR1.5) and the IPCC Sixth Assessment[6] (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 limitations[7,8].

Scenarios that are collected as part of IPCC or other exercises[9–11] represent ensembles of opportunity: a serendipitous collection of scenario data that is unstructured[7] 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 analysis[7,8]. Biases include dominance of specific models and intercomparison projects, which can represent a lack of diversity in organizational or regional composition[12–14]. 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

[1]Centre for Environmental Policy, Faculty of Natural Sciences, Imperial College, London, UK. [2]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. [3]The Grantham Institute for Climate Change and the Environment, Imperial College London, London, UK. [4]PBL Netherlands Environmental Assessment Agency, The Hague, the Netherlands. [5]Copernicus Institute for Sustainable Development, Utrecht University, Utrecht, the Netherlands. [6]Center for Global Sustainability, University of Maryland, College Park, MD, USA. ✉e-mail: hamish.beath16@imperial.ac.uk; j.rogelj@imperial.ac.uk

https://doi.org/10.1038/s41558-026-02565-5

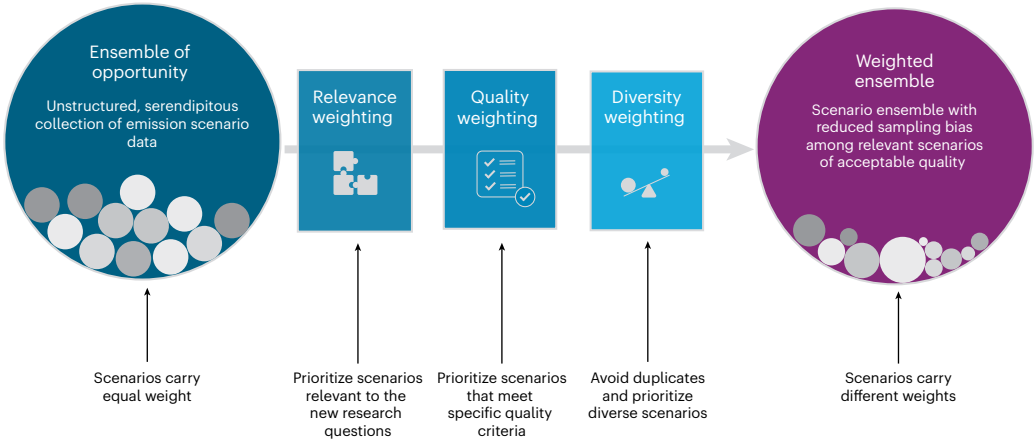


Fig. 1 | A scenario-weighting framework for the analysis of scenario ensembles of opportunity. Schematic showing how the unstructured, serendipitous collection of evidence in a scenario ensemble of opportunity can be translated in a weighted ensemble, accounting for the relevance, quality and diversity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

estimated level of global warming avoided by the scenario[15], assessments of the feasibility of its described transitions[16] 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 predefined 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 with 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 exercise[17].

In the past, such issues have been dealt with on an ad hoc basis. For example, the IPCC SR1.5[5] checked for the completeness of variables available in scenarios, whether data are reported until 2100 or whether reported historical- or near-future data are consistent with observations[9,18]. For the assessment of global emission characteristics of pathways aligned with 1.5 °C, SR1.5 also excluded 13 scenarios from a single modelling group[19] 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 IPCC[6,8].

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 models[20–22], as well as for model quality measured as performance relative to historical observations[22,23]. In emission scenario ensemble analysis, issues of scenario similarity have been considered[24], but as of yet not systematically addressed.

Here we use efforts from various communities[20–24] 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 provides a systematic approach that unifies and expands on previous ad hoc scenario assessment decisions[1–6]. Further, it presents an alternative to other approaches that reweight or present summary statistics accounting for common model or intercomparison study[14,17]. We apply it to the latest available IPCC AR6 scenario ensemble[10] 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.

Table 1 | Variable for estimation of scenario similarity or diversity

Variable	b_n expert	b_n correlation-adjusted	$\sigma_{S,n}$	Group
CO_2 emissions	1/8	1/8	2.75 $GtCO_2\,yr^{-1}$	Emissions
CH_4 emissions	1/24	0	31.2 $MtCH_4\,yr^{-1}$	Emissions
N_2O emissions	1/24	1/8	0.898 $MtN_2O\,yr^{-1}$	Emissions
Sulfur emissions	1/24	0	8.48 $MtSO_2\,yr^{-1}$	Emissions
Consumption	1/8	1/8	US$41,664 billion (2010)	Economy
GDP (PPP)	1/8	1/8	US$3,983 billion (2010)	Economy
Carbon capture and storage	1/8	1/8	2,811 $MtCO_2\,yr^{-1}$	Mitigation
Carbon price	1/8	1/8	US$19.9 (2010)	Mitigation
Primary energy: oil	1/48	0	35.7 $EJ\,yr^{-1}$	Energy
Primary energy: gas	1/48	0	36.7 $EJ\,yr^{-1}$	Energy
Primary energy: coal	1/48	0	28.5 $EJ\,yr^{-1}$	Energy
Primary energy: nuclear	1/48	1/8	9.74 $EJ\,yr^{-1}$	Energy
Primary energy: biomass	1/48	0	18.1 $EJ\,yr^{-1}$	Energy
Primary energy: non-biomass renewables	1/48	1/8	23.7 $EJ\,yr^{-1}$	Energy
Final energy	1/8	0	40.7 $EJ\,yr^{-1}$	Energy

The 15 variables from the IPCC AR6 database[10] are used to determine the scenario similarity. For each variable the importance weighting b_n assigned to each is shown (equation (6)) along with the radius of scenario similarity $\sigma_{S,n}$. The group that each variable applies to is shown, with each group having a total weight of 1/4.

A generalized scenario-weighting framework

Our starting point takes a generalized 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 formalize this step. Each scenario i contributing to the assessment and member of the ensemble of opportunity E is assigned a generalized weight gw_i:

https://doi.org/10.1038/s41558-026-02565-5

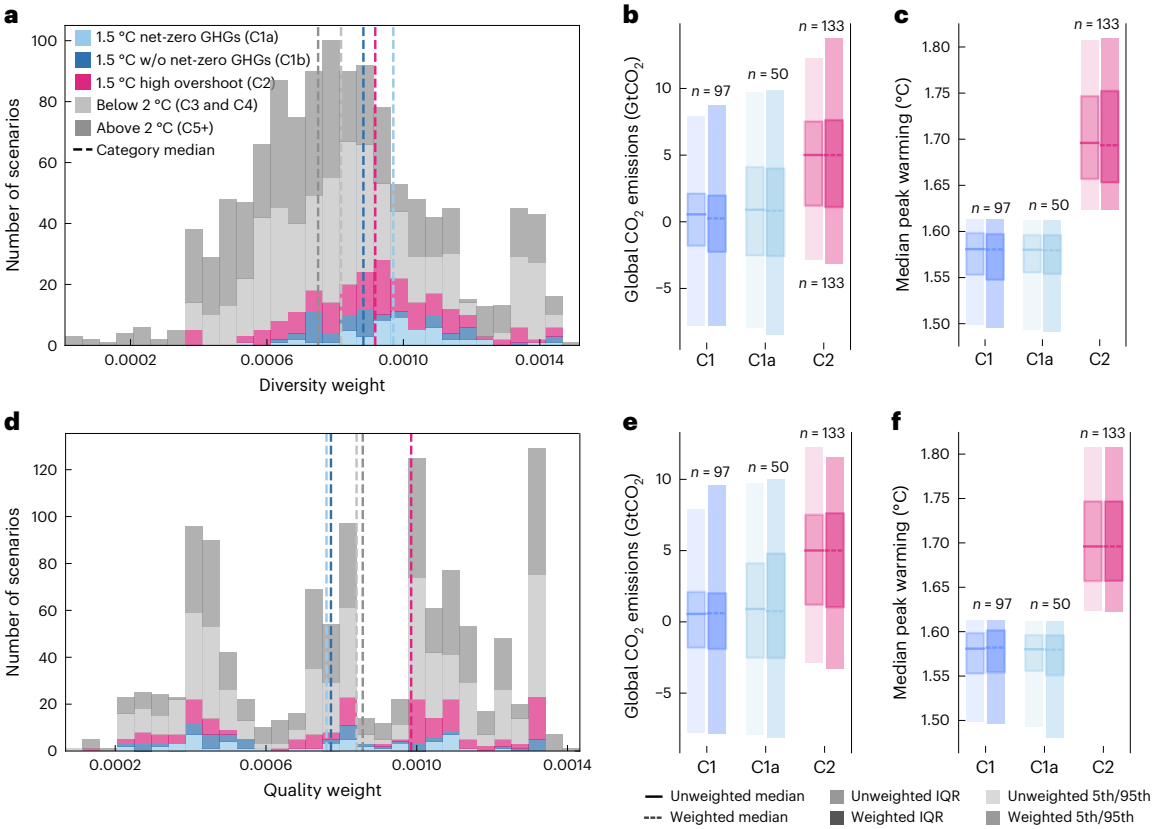


Fig. 2 | Diversity weighting of scenarios available in the IPCC AR6 scenario database. a, Diversity weights $D(i)$ for scenarios per IPCC scenario category. **b**, Unweighted and diversity-weighted distributions for 2050 CO_2. **c**, Unweighted and diversity-weighted distributions for median peak temperature (MAGICCC). **d**, Quality weights $Q(i)$ for scenarios per IPCC scenario category. **e**, Unweighted and diversity-weighted distributions for 2050 CO_2. **f**, Unweighted and quality-weighted distributions for median peak temperature (MAGICCC). The legend for **b**, **c**, **e** and **f** is below the four panels, with the darker coloured distributions for each temperature representing weighted ones. The wide shaded ranges represent the 5th and 95th percentiles, with the narrow ones the IQR. Weighted and unweighted quantiles are computed using the same non-interpolating approach (Methods).

$$\mathrm{gw}_i = \frac{R(i)Q(i)}{D(i)} \quad (1)$$

where $0 \le R(i) \le 1$ is a measure of how relevant a scenario is to answer a specific question (relevance weighting), $0 \le Q(i) \le 1$ is a metric of scenario quality (quality weighting) and $D(i) \ge 1$ is a metric of scenario uniqueness (diversity weighting). This ensures that gw_i 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 Methods.

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 database[10] and look at the influence on key scenario assessment outcomes of the AR6[6]. 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 GHG emissions over the course of the twenty-first century[25] (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, $Q(i)$, based on the IPCC AR6 vetting procedure[26] (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 (Table 1). Owing to correlations between variables, we use eight representative variables for the main results (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 Supplementary Results 2.

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 more 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 spiky distribution, as scenarios from specific intercomparison projects and models have harmonized input data.

Visually, 2050 CO_2 emission and peak warming ranges of C1 and C2 show little change (Fig. 2b,c,e,f). Under diversity weighting, the median 2050 CO_2 emissions reduces slightly for C1 and C1a, while all categories show wider interquartile ranges (IQRs) and 5th–95th ranges (Fig. 2b). Median peak temperatures barely change for C1 and C1a but fall for C2; all categories have wider IQRs. For quality weighting (Fig. 2e,f), the impact on medians varies by categories.

For other outcomes of our diversity weighting, including net-zero years (Fig. 3a) and the trajectory of certain variables (Fig. 3b–l and Supplementary Figs. 2 and 3), some metrics show noteworthy changes. For C1 scenarios, the median year of reaching net-zero GHG emissions

https://doi.org/10.1038/s41558-026-02565-5

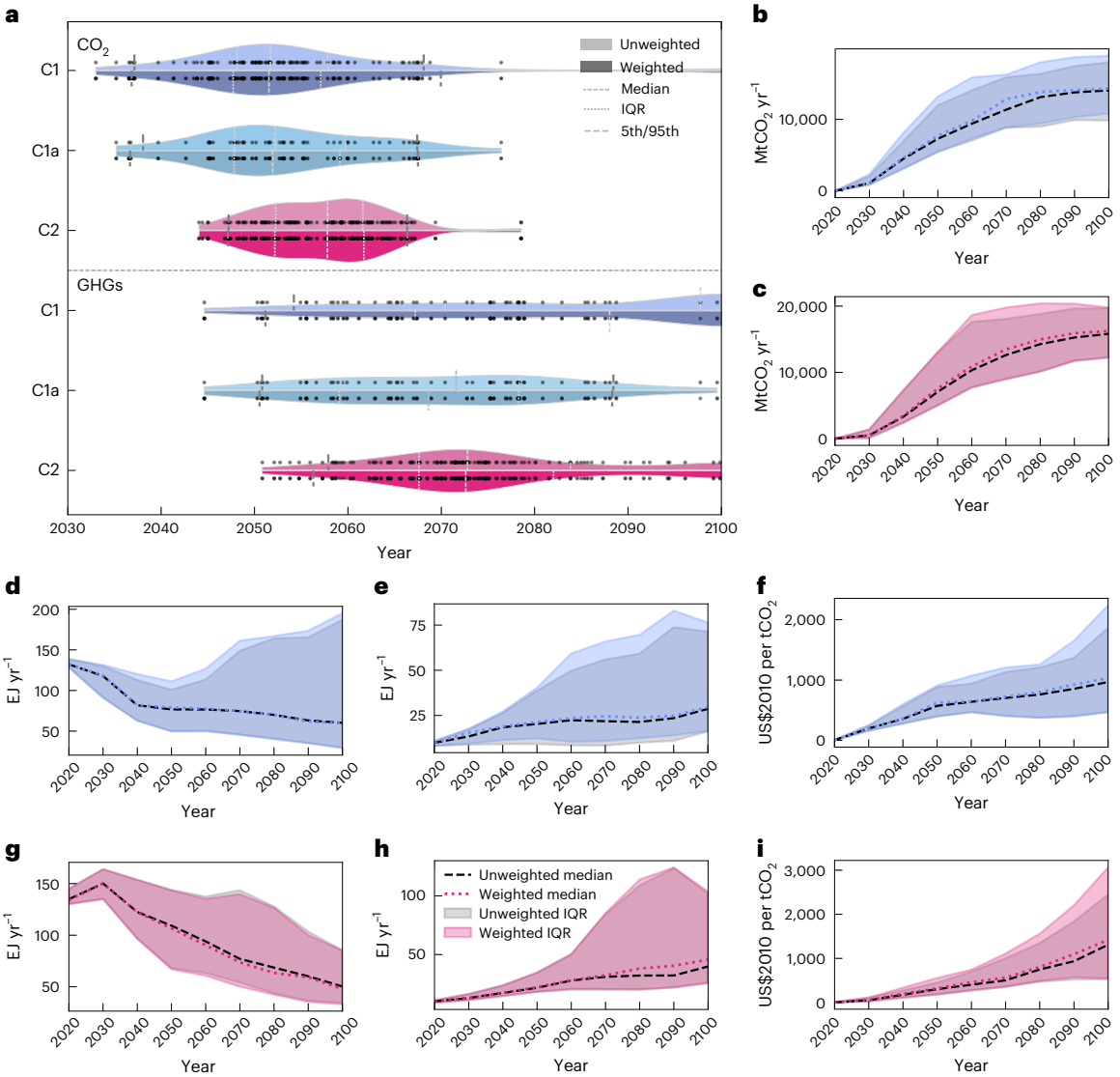


Fig. 3 | Key assessment quantities in diversity-weighted and unweighted scenario category ranges from the IPCC AR6 scenario database. a, Net-zero CO_2 and GHG years. **b**, Carbon sequestration from carbon capture and storage for C1. **c**, Carbon sequestration from carbon capture and storage for C2 **d**,**g**, Primary energy from gas time series for C1 (**d**) and C2 (**g**). **e**,**h**, Primary energy from nuclear time series for C1 (**e**) and C2 (**h**). **f**,**i**, Carbon price time series for C1 (**f**) and C2 (**i**). 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 **a**. For time series plots, the coloured shaded areas represent the IQR of the reweighted distribution, with grey shaded areas representing the unweighted IQR. The coloured dashed lines represent weighted medians, whereas black dashed lines represent unweighted medians, as shown in the legend in **h** and applicable to panels **b**–**i**. Weighted and unweighted quantiles are computed using the same non-interpolating approach (Methods).

is advanced by a decade (from 2098 to 2088); but for 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 CO_2 years remain virtually unchanged (C1/C1a <1 year earlier; C2 unchanged).

For many of the 15 variables reported by all scenarios, changes to time series are negligible (Supplementary Figs. 2 and 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 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 IQR, 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.

https://doi.org/10.1038/s41558-026-02565-5

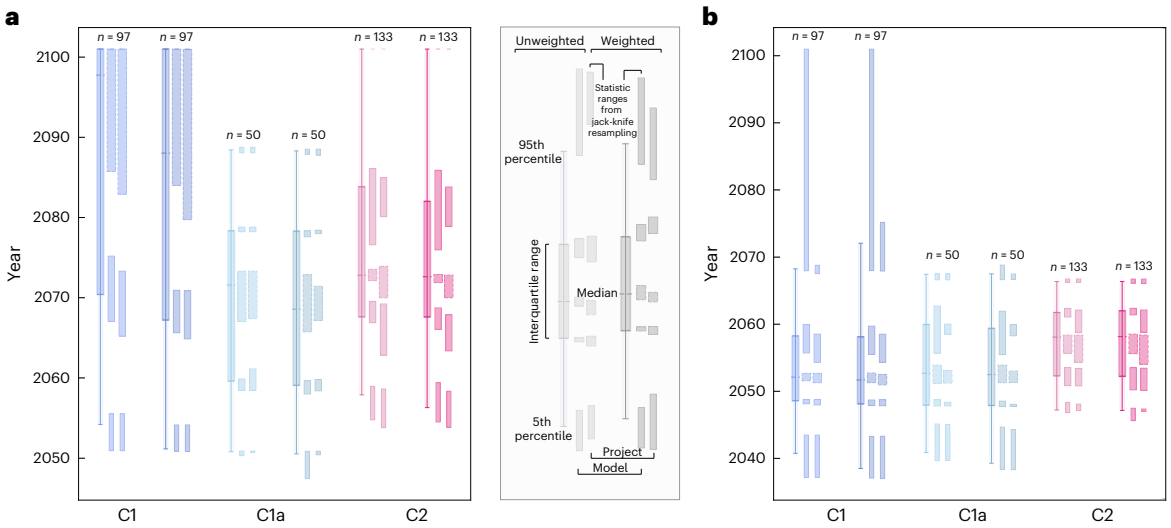


Fig. 4 | Interplay between weighting of scenario ensembles and jack-knife resampling outcomes. a, Net-zero GHG years. **b**, Net-zero GHG years CO_2. Statistics are documented in the legend, which should be used for interpreting both **a** and **b**. For each statistic of the unweighted and the weighted ensemble, the jack-knife 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 1 and 2 for a list. Both weighted and unweighted quantiles are computed using the same non-interpolating approach (Methods).

We explore further diversity-weighting procedures and their impact in Supplementary Results 2 and 3. These further explorations highlight which changes broadly persist across inputs (for example, reduction in net-zero GHG years, higher carbon capture and storage) and which vary depending on the choice of weighting inputs (for example, primary energy from gas). Further, some variables such as carbon capture and storage and primary energy from nuclear appear more sensitive to changes in sigma values (Supplementary Figs. 4 and 8). The impact of continuous quality weighting is shown in Supplementary Figs. 5 and 6.

Exploring robustness

It is well-established that IAMs behaviour is to a large degree determined by their framework logic, structure and input assumptions[27–29]. Equally, model intercomparison studies use harmonized scenario design assumptions, dominating key scenario insights[30,31]. 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 ensembles[17,32]. Such a test calculates how outcomes V_d vary across the ensemble, E when iteratively removing scenarios according to each instance d in resampling dimension D (for example, 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 then test the sensitivity to contributions by projects (Fig. 4 and Supplementary Tables 1 and 2). 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 CO_2 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.

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 (Fig. 5b,d). We apply the Herfindahl–Hirschman index (HHI)[33], 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 (Methods). For model framework and model type, when compared with 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 Fig. 11). Sensitivity tests show these general trends are robust to a range of alternative inputs to the diversity-weighting scheme (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 ten diversity-weighted scenarios for both C1 and C2 originate from a model framework and/or project with low prevalence (Fig. 5a,c). Likewise, the bottom ten 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 $GtCO_2$ differences (400–650) in full-century carbon budgets (Fig. 5). In our C2 top ten, there are three scenarios which are also carbon budget

https://doi.org/10.1038/s41558-026-02565-5

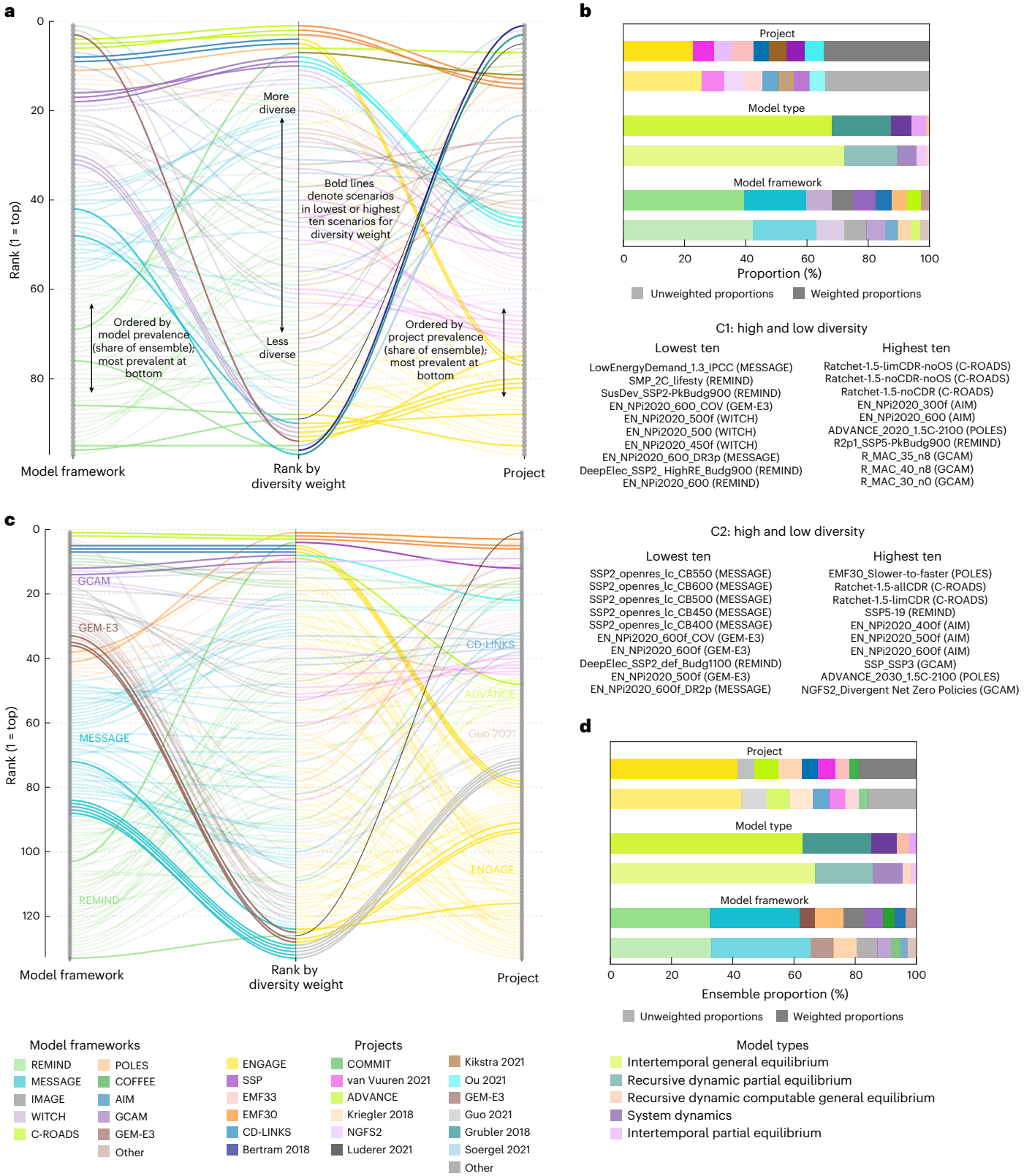


Fig. 5 | Insights into how scenarios diversity weighted according to model frameworks and projects. a, Connected rank plot showing C1 scenarios grouped by model (left ranking), diversity weight (central ranking) and project (right ranking). **b**, The unweighted and weighted proportions of each project, model type and model framework for C1 scenarios. **c**, Equivalent connected rank plot from **a**, but for C2 scenarios. **d**, Stacked bars showing the unweighted and weighted proportions of each project, model type and model framework for C2 scenarios. For the rank plots (**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 for scenarios in the highest or lowest ten scenarios by diversity weighting. For the stacked bars (**b**,**d**) the lighter shaded bars are unweighted and the darker shaded bars are diversity weighted. Only the eight largest shares are shown, with others grouped together as 'Other', as visible on the key.

variants; however, these come from the AIM model framework (third least prevalent) and have wider carbon budget differences (100 $GtCO_2$).

Reweighting for improved scenario assessment

The weighting framework presented here is designed to be simple and transparent. Although our illustrative application to AR6 scenarios is not 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 Agenda[5,34,35]. Alternatively, it could be used to consider questions regarding feasibility of scenario pathways across several dimensions[16,36]. 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 emissions[37].

Beyond global IAM emission scenarios, the framework could also be applied to ensembles of national or sectoral pathways[38]. 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 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 judgement exists in the identification and choice of the weighting approaches, distance metrics or relevant indicators and in interpretation of weighted ensembles. Here we demonstrate how some expert judgements 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 judgements. Acknowledging this, the structured nature of the framework can help to ensure that the necessary expert judgement is well-guided and transparent.

Moreover, transparency is essential: scenario data have a range of applications, some of which may incentivize selective presentation. Without a clear framework and best-practice guidance, reweighting 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 composition[8]. 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 statistics[12]. 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, while 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 demand[39,40], are likely to be weighted higher and contribute more to category statistics under a weighted assessment. The drawback of this is that potential and emerging scenario approaches might over-emphasize scenarios with technologies at the boundaries of realism. Indeed, diversity weighting in isolation could incentivize the creation of such scenarios in unintended ways. There is a need to explore extremes and new approaches[41], 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. The approach presented here 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.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41558-026-02565-5.

References

1. IPCC *Special Report on Emissions Scenarios* (eds Nakicenovic, N. & Swart, R.) (Cambridge Univ. Press, 2000).
2. Morita, T. et al. in *Climate Change 2001: Mitigation* (eds Metz, B. et al.) 115–166 (IPCC, Cambridge Univ. Press, 2001).
3. Fisher, B. et al. in *Climate Change 2007: Mitigation* (eds Metz, B. et al.) 169–250 (IPCC, Cambridge Univ. Press, 2007).
4. Clarke, L. et al. in *Climate Change 2014: Mitigation of Climate Change* (eds Edenhofer, O. et al.) 413–510 (IPCC, Cambridge Univ. Press, 2014).
5. Rogelj, J. et al. in IPCC *Special Report on Global Warming of 1.5 °C* (eds Masson-Delmotte, V. et al.) 93–174 (WMO, 2018).
6. Riahi, K. et al. in *Climate Change 2022: Mitigation of Climate Change* (eds Shukla, P. R. et al.) 295–408 (IPCC, Cambridge Univ. Press, 2022).
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* (IIASA, 2018); https://doi.org/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. in IPCC *Special Report on Global Warming of 1.5 °C* (eds Masson-Delmotte, V. et al.) 2SM-1–2SM-50 (WMO, 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*. https://doi.org/10.1080/17583004.2018.1477374 (2018).
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. Dynam. Discuss*. https://doi.org/10.5194/esd-2020-23 (2020).
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**, 135 (2022).
26. Guivarch, C., Kriegler, E. & Portugal-Pereira, J. in *Climate Change 2022: Mitigation of Climate Change* (eds Shukla, P. R. et al.) 1841–1908 (IPCC, Cambridge Univ. Press, 2022).
27. Kriegler, E. et al. Diagnostic indicators for integrated assessment models of climate policy. *Technol. Forecast. Soc. Change* **90**, 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. Riahi, K. et al. Cost and attainability of meeting stringent climate targets without overshoot. *Nat. Clim. Change* **11**, 1063–1069 (2021).
31. Bertram, C. et al. Feasibility of peak temperature targets in light of institutional constraints. *Nat. Clim. Change* **14**, 954–960 (2024).
32. Rogelj, J. et al. Zero emission targets as long-term global goals for climate protection. *Environ. Res. Lett.* **10**, 105007 (2015).
33. Rosenbluth, G. in *Business Concentration and Price Policy* 57–99 (National Bureau of Economic Research, 1955).
34. Roy, J. et al. in IPCC *Special Report on Global Warming of 1.5 °C* (eds Masson-Delmotte, V. et al.) 445–538 (WMO, 2018)
35. Deprez, A. et al. Sustainability limits needed for CO_2 removal. *Science* **383**, 484–486 (2024).
36. 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).
37. 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).
38. *National Scenarios* (IAMC, 2025); https://www.iamconsortium.org/scientific-working-groups/national-scenarios/
39. 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).
40. Millward-Hopkins, J. Inequality can double the energy required to secure universal decent living. *Nat. Commun.* **13**, 5028 (2022).
41. McCollum, D. L., Gambhir, A., Rogelj, J. & Wilson, C. Energy modellers should explore extremes more systematically in scenarios. *Nat. Energy* **5**, 104–107 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

Methods

Here we provide a detailed explanation of our weighting framework. We explain in more detail how relevance, quality and diversity weighting can be applied. The final sections are related to our demonstration of the weighting framework to the AR6 database.

Weighting based on scenario relevance

Analysing and re-using scenario data is only sensible once a corresponding research question is defined. A key step in generalized scenario weighting is therefore to define question-specific relevance weights $R(i)$ (equation (1)). This relevance-weighting term, $R(i)$, can take several forms depending on the research question. For example, it could be strictly binary, including or excluding a scenario i based on meeting the question-specific condition C:

$$R(i) = \begin{cases} 1 & \text{if } C \\ 0 & \text{else} \end{cases} \tag{2}$$

A straightforward example is whether a scenario limits global warming to within certain temperature bounds[15]. This condition is critical if the intention is to explore characteristics aligned with keeping warming well below 2 °C or 1.5 °C with a specific likelihood[15,25,42].

However, binary weighting of scenarios based on temperature outcome is only fully defensible if uncertainty around the climate outcome of a scenario can be unambiguously quantified with a single probability distribution. This is typically never the case[43]. Under such conditions, $R(i)$ could also take the form of a continuous function. Consider an example with a threshold ϑ provided for a given scenario metric m_i. $R(i)$ can be constructed such that scenarios within the threshold are weighted with unity and scenarios beyond the threshold are weighted on the basis of their distance to the threshold, for example, using a stretched exponential function with scaling factor α and stretching exponent β to determine the relevance weight:

$$R(i) = \begin{cases} 1 & \text{if } m_i < \vartheta \\ e^{-\left(\left(\alpha \frac{m_i - \vartheta}{\vartheta}\right)^{\beta}\right)} & \text{else} \end{cases} \tag{3}$$

Alternatively, scenarios that are within a defined threshold ϑ, could be weighted on the basis of their distance from it. For example, a user may want adherence to temperature thresholds but may want to apply a risk-based relevance weighting. Here scenarios further from thresholds achieve a higher $R(i)$, as they are deemed to have less risk of breaching the threshold.

Weighting based on scenario quality

The quality of a scenario is central to consider whether to continue using its information for secondary analysis. However, an assessment of quality is subjective and depends on the research question being explored. For example, accurate historical emissions might be important for understanding how emissions relate to global warming targets, while accurate historical energy-system capacities might be important if characteristics of the energy transformation are gleaned from these pathways. Alternatively, quality weighting could be derived from the characteristics of certain models. For example, their representation of specific technologies of interest. Quality weighting could also be assigned using quantifiable feasibility criteria[16,44], obtained from literature, expert judgement or elicitation[45]. Practically, this could mean down-weighting scenarios with technology pathways or societal changes judged by experts to be outside plausible achievability.

Methods that have earlier been applied to account for climate data projection quality[20–23] can be adapted to our context, by comparing emissions and energy trends in scenarios to recent observations. Scenario quality can be accounted for by implementing a continuous weighting $Q_j(i)$ between 0 and 1 for each scenario i, based on a set of j distance criteria $f_j(d)$ of k quality metrics defined between modelled data for a metric ($v_{k,i}$) of a scenario and expert assessment of a set of measures of quality (E_k) (equation (4)).

$$Q(i) = \prod_{j,k} f_j\left(d\left(v_{k,i}, E_k\right)\right) \tag{4}$$

The IPCC SR1.5 and AR6 assessments did not use a formal method of scenario weighting but applied a scenario quality filtering[7,8,18,26], with each scenario effectively assigned $Q(i) = 0$ (excluded) or $Q(i) = 1$ (included). For example, SR1.5 excluded scenarios with negative CO_2 emissions from agriculture, forestry and other land-use (AFOLU) in 2020 because of the perceived implausibility of AFOLU CO_2 emissions becoming negative within 2 years of the publication of the report[7,18]. In addition, SR1.5 excluded a further 30 scenarios from its analysis of global and sectorial emissions evolutions (Table 2.4 and Fig. SPM.3 in ref. 5) for having GHG emissions in 2010 outside the range of historical estimates. Other parts of the report that explore scenario dimensions orthogonal to global and sectorial emissions evolutions, include all scenarios. In both above-mentioned cases, $f_j(d)$ would take a functional form that returns 1 when a scenario variable in a given year falls inside a set range and zero when not.

Weighting based on scenario diversity

Scenario diversity provides a metric of distance or proximity between scenarios. It requires expert judgement and a clear research question to identify relevant variables to consider. For example, to explore emissions statistics consistent with limiting global warming to a specific level, considering the diversity in emission evolutions of scenarios could be sufficient. For questions about possible energy-system configurations compatible with a specific climate goal, another set of variables could be selected for determining that diversity.

We use an adapted version of the method to estimate effective repetitions in Earth system model projections[20,21] that uses a Gaussian function:

$$D(i) = \left(1 + \sum_{i' \neq i} \exp\left(-\frac{S_{ii'}^2}{\sigma_S^2}\right)\right) \tag{5}$$

where $S_{ii'}$ is a similarity distance metric between two scenarios i and i' for a single variable and is taken as a root-mean-square difference between two time series; and σ_S is the 'radius of scenario similarity'[21]. Parameter σ_S defines how close models need to be to be effectively down-weighted and is an assessment choice. If two scenarios i and i' are exactly the same, $S_{ii'} = 0$ and $D(i) = 2$, weighting each model as 1/2. Equation (5) can be generalized to compare scenario similarity across N variables:

$$D_n(i) = 1 + \sum_{n=1}^{N} \sum_{i' \neq i} b_n \exp\left(-\frac{S_{ii',n}^2}{\sigma_{S,n}^2}\right) \tag{6}$$

where b_n are relative contributions of each variable to the total diversity weighting that sum to 1 across all variables n. Besides selecting $\sigma_{S,n}$ for each variable n, equation (6) also requires the selection of the appropriate constants b_n. The simplest case could be to set $b_n = 1/N$, but some variables may be more important than others in relative terms.

Other ways of defining $D(i)$ could also draw on an analysis of the underlying energy model fingerprint of mitigation scenarios[46] or alternative ways of conceptualizing distance, for example, through principal component analysis[47].

Dependencies between weighting applications

Our method uses several weights which can interact with one another. For instance, the definition and calculation of quality $Q(i)$ and diversity $D(i)$ weights is informed by the scope of relevance weights $R(i)$ being applied. Further, quality and relevance can be multidimensional and

individual dimensions can be weighted differently. As a result, a set of final scenario weights and the analysis outcomes of the weighted dataset may change based on the scope of the research question. Our weighting method allows this dependency to be communicated in a structured and transparent way.

Consider two analyses of interest. The first interrogates a scenario dataset based on temperature outcomes. In this case, $Q(i)$ may comprise weights based on proximity to historical emissions inventories. However, for a research question focused on investments needed to achieve a given temperature limit, $Q(i)'$ would additionally weight scenarios on the basis of their historical investment and near-term investment outcomes. In this case, $Q(i)$ differs from $Q(i)'$ and secondary analysis outcomes will reflect these different research demands.

Application of weighting to AR6 scenarios

Here we provide the weighting specifications for the illustrative application of the framework to the IPCC AR6 database[10]. Integrating expressions for all weighting components, the application-specific scenario-weighting equation becomes:

$$\mathrm{gw}_i = \frac{R(i)Q(i)}{D(i)} = R(i)Q(i)/\left(1+\sum_{n=1}^{N}\sum_{i'\neq i} b_n \exp\left(-\frac{S^2_{ii',n}}{\sigma^2_{S,n}}\right)\right) \quad (7)$$

Relevance weighting. For analysis of scenarios that limit global warming to 1.5 °C with no or limited overshoot (IPCC AR6 category C1), or scenarios that return warming to 1.5 °C after a high overshoot (IPCC AR6 category C2), respectively, we apply binary question-specific relevance weights $R(i)$ that select scenarios with the specific global warming characteristics as defined in IPCC AR6[6].

$$R_{\mathrm{C1}}(i) = \begin{cases} 1, & P(\mathrm{d}T_{\text{21st century}} > 1.5^{\circ}\mathrm{C}) \leq 67\% \wedge P(\mathrm{d}T_{2100} < 1.5^{\circ}\mathrm{C}) > 50\% \\ 0, & \text{else} \end{cases} \quad (8)$$

And separately:

$$R_{\mathrm{C2}}(i) = \begin{cases} 1, & P(\mathrm{d}T_{\text{21st century}} > 1.5^{\circ}\mathrm{C}) > 67\% \wedge P(\mathrm{d}T_{2100} < 1.5^{\circ}\mathrm{C}) > 50\% \\ 0, & \text{else} \end{cases} \quad (9)$$

C1a scenarios are a subset of C1 scenarios in which net-GHG emissions reach zero levels in the second half of the century.

Quality weighting. We test a continuous quality weighting $Q(i)$, adapted from the IPCC AR6 procedures and reported in Table 11 in ref. 26 (reproduced in Supplementary Table 3). For each quality criterion, j, that specified a range, we use the distance of the modelled value, k, for each scenario i from the reference value, $d_{k,i}$, normalizing by the IQR of the scenarios

$$\widetilde{d_{k,i}} = \frac{d_{k,i}}{\mathrm{IQR}(d_k)} \quad (10)$$

We then use a Gaussian function to produce the continuous weight, w, for each criterion, j:

$$w_j(i) = \exp\left[-\left(\widetilde{d_{k,i}}\right)^2\right] \quad (11)$$

This yields a weight of 1 when there is perfect agreement with the reference value. We treat each criterion evenly and combining to give a single quality weighting:

$$Q(i) = \sum_j w_j(i) \quad (12)$$

Diversity weighting. Diversity weighting $D(i)$ accounts for variations in 15 variables across four key dimensions (emissions, economy, mitigation strategy and energy; Table 1). These 15 variables are part of the minimum data requirement for a scenario to be considered as part of the IPCC AR6 ensemble and therefore available for each scenario. The illustrative variable weights b_n for each variable are a subjective choice and here chosen to put equal weights on each of the four key dimensions. Subweighting of the variables in each group attempts to limit the overall influence of variables that are similar in nature and could bias the total, or puts weight on more important variables. For example, in the energy category, the six primary energy variables combined are worth the same as the single variable of final energy. In the emissions category, the aggregate weighting of all non-CO_2 emissions variables equals that of CO_2, recognizing that CO_2 is the primary anthropogenic driver of climate change[48].

Although not applicable here, scenarios that do not report all variables can also be included by rescaling coefficients b_n for the variables that are available. This rescaling is first achieved within each variable group or across all remaining groups if one variable group would end up having no reported variables. The sum of b_n is always unity.

Acknowledging the subjectivity in variable weighting, we also explored using only energy variables, as well only emissions variables from our selection (Supplementary Results 2). To account for correlation between variables, we adjusted our variable weights using a correlation matrix and hierarchical clustering (Supplementary Methods 1). This results in a set of 'correlation-adjusted' variable weights, with a reduced number of eight variables used following correlation assessment and clustering. We adopt an even weighting across the variables as we preserve emphasis on originally higher weighted variables in selection of our representative cluster variables (Table 1 and Supplementary Methods 1). We present our correlation-adjusted weights in the main text but also show results from our full set of variables in Supplementary Results 2.

Our conceptual choice of $\sigma_{S,n}$ (radius of similarity) between two scenarios is guided by expert judgement and should be seen as an informed yet illustrative choice rather than a conclusive one. We define $\sigma_{S,n}$ as the root-mean-square difference between different IAMs running the same shared socioeconomic pathways and representative concentration pathways (SSP–RCP) combination available from ref. 9 using 10-year timesteps from 2020 to 2100. This choice is made on the basis that two independent IAMs running the same scenario are structurally different yet subject to a set of common constraints (exogenously provided population and gross domestic product (GDP) projections and year-2100 radiative forcing). We explore a range of values between the minimum and maximum of the SSP–RCP model differences, selecting a value that achieves a high median spread in diversity weights across our variables, while ensuring a relatively even impact across all variables (Supplementary Results 2). For no-policy, no-mitigation or business-as-usual scenarios, we remove the mitigation group variables from determining the distances of $\sigma_{S,n}$ as these are zero by definition in these scenarios.

Calculation of weighted quantiles. Quantiles are calculated without interpolation. Reported quantile values are the lowest scenario values equal to or above the quantile. When applying to weighted distributions, quantiles are calculated using the same approach, but rather than scenarios having an equal weight in the distribution, scenarios contribute according to its assigned weight. Therefore, the cumulative sum of weights is used to identify the quantile threshold.

Calculation of Herfindalh–Hirschman index. To quantify changes in model or project concentration in our unweighted and weighted ensembles, we use the HHI[33]. This has historically been used for measuring market concentration in economics but can be applied more

broadly to a dataset to understand dominance of components. It is defined by the sum of squared shares of each component:

$$\mathrm{HHI} = \sum_{i=1}^{N} c_i^2 \tag{13}$$

Where c_i is the component i in relation to the total. A higher HHI value indicates increased dominance of single components.

Data availability

All data used for this study are available via Zenodo at https://zenodo.org/records/18164313 (ref. 49).

Code availability

The code used for the results in this study can be found at https://github.com/hamishbeath/scenario-reweighting.

References

42. 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).
43. Forster, P. et al. in *Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. et al.) 962–967 (IPCC, Cambridge Univ. Press, 2021).
44. 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).
45. Colson, A. R. & Cooke, R. M. Expert elicitation: using the classical model to validate experts' judgments. *Rev. Environ. Econ. Policy* **12**, 113–132 (2018).
46. Dekker, M. M. et al. Identifying energy model fingerprints in mitigation scenarios. *Nat. Energy* **8**, 1395–1404 (2023).
47. Dekker, M. M. et al. Spread in climate policy scenarios unravelled. *Nature* **624**, 309–316 (2023).
48. IPCC: Summary for Policymakers. In *Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. et al.) 3–31 (IPCC, Cambridge Univ. Press, 2021).
49. Beath, H. & Chris, S. Tools for debiasing scenario ensembles. hamishbeath/scenario-reweighting: V0.1.1. *Zenodo* https://doi.org/10.5281/zenodo.18213995 (2026).

Acknowledgements

H.B. acknowledges funding from the Hitachi-Imperial Centre for Decarbonisation and Natural Climate Solutions. C.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 no. NE/S007415/1. M.J.G. was supported by funding from the European Research Council under the European Union Horizon 2020 research and innovation programme (GENIE—grant agreement no. 951542). J.R. and H.B. were supported by the European Union 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 no. G-2023-201305841). M.J.G. is also affiliated with Pacific Northwest National Laboratory, which did not provide specific support for this paper.

Author contributions

C.S., J.S.K., M.J.G. and J.R. conceived the original study, with H.B. and M.M.D. conceptualizing further study elements. C.S. performed initial experiments, analysed their data and produced visualizations. Code was initiated by C.S. and rewritten and greatly extended by H.B. for use in the present study. J.R. and H.B. led the writing of the study, while M.M.D., C.S., M.J.G. and J.S.K. contributed parts. H.B., C.S., J.S.K, M.M.D, M.J.G. and J.R. reviewed the paper before submission. H.B. led the experiments, their analysis and the subsequent revisions of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41558-026-02565-5.

Correspondence and requests for materials should be addressed to Hamish Beath or Joeri Rogelj.

Peer review information *Nature Climate Change* thanks John Bistline, Benjamin Sanderson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.