

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

In the format provided by the
authors and unedited

Contents

Contents	1
Supplementary Methods Tables	2
Supplementary Results 1: Implementation of Risk-Averse Relevance Weighting Variation	4
Supplementary Results 2: Sensitivity of climate action benchmarks to different variable weights and sigma values	7
Supplementary Results 3: Impact of different sigma values on spread of diversity weights	12
Supplementary Methods 1 Calculation of Correlation Adjusted Variable Weights	15
Supplementary Results 4: Impact of diversity weighting on database characteristics and exploration of scenario weights	16
References	18

Supplementary Methods Tables

Supplementary Table S1 | Entities used for jack-knife resampling along the model group dimension. Models are identified as in the IPCC AR6 scenario database's meta data¹ and model groups as indicated in this table.

Model Group	Model
AIM	AIM/CGE 2.1
AIM	AIM/CGE 2.2
AIM	AIM/Hub-Global 2.0
C-ROADS	C-ROADS-5.005
COFFEE	COFFEE 1.1
GCAM	GCAM 4.2
GCAM	GCAM 5.3
GEM	GEM-E3_V2021
IMAGE	IMAGE 3.2
MESSAGE	MESSAGE-GLOBIOM 1.0
MESSAGE	MESSAGEix-GLOBIOM 1.0
MESSAGE	MESSAGEix-GLOBIOM_1.1
MESSAGE	MESSAGEix-GLOBIOM_1.2
POLES	POLES ADVANCE
POLES	POLES EMF33
REMIND	REMIND 1.7
REMIND	REMIND 2.1
REMIND	REMIND-MAgPIE 1.5
REMIND	REMIND-MAgPIE 1.7-3.0
REMIND	REMIND-MAgPIE 2.1-4.2
REMIND	REMIND-MAgPIE 2.1-4.3
WITCH	WITCH 5.0
WITCH	WITCH-GLOBIOM 3.1
WITCH	WITCH-GLOBIOM 4.4

Supplementary Table S2 | Entities used for jack-knife resampling along the project dimension. Projects and studies are identified as in the IPCC AR6 scenario database's meta data¹.

Project
COMMIT
CD-LINKS
ENGAGE
Fujimori 2020
Holz 2018
SSP
Ou 2021
van Vuuren 2021
ADVANCE
EMF30
EMF33
Grubler 2018
NGFS2
Kikstra 2021
Strefler 2018
Strefler 2021a
Schultes 2021
Baumstark 2021
Kriegler 2018
Bertram 2018
Strefler 2021b
Soergel 2021
Luderer 2021
Levesque 2021
Guo 2021

Supplementary Table S3 | Summary of vetting criteria applied in IPCC AR6 and in this study. Adapted from Table 11 in ref. ².

	Reference Value	Range (IP range)
Historical emissions (2019 values)		
CO ₂ total (EIP + AFOLU) Emissions	44251 MtCO ₂ yr ⁻¹	±40% (±20%)
CO ₂ EIP emissions	37646 MtCO ₂ yr ⁻¹	±20% (±10%)
CH ₄ emissions	379 MtCH ₄ yr ⁻¹	±20% (±20%)
CO ₂ emissions EIP 2010-2020 % change	-	+0% to +50%
CCS from energy 2020	-	0-250 (100) MtCO ₂ yr ⁻¹
Historical energy production (2020 values)		
Primary Energy	578 EJ	±20% (±10%)
Electricity: nuclear	9,77 EJ	±30% (±20%)
Electricity: solar and wind	8.51 EJ	±50% (±25%)

Supplementary Results 1: Implementation of Risk-Averse Relevance Weighting Variation

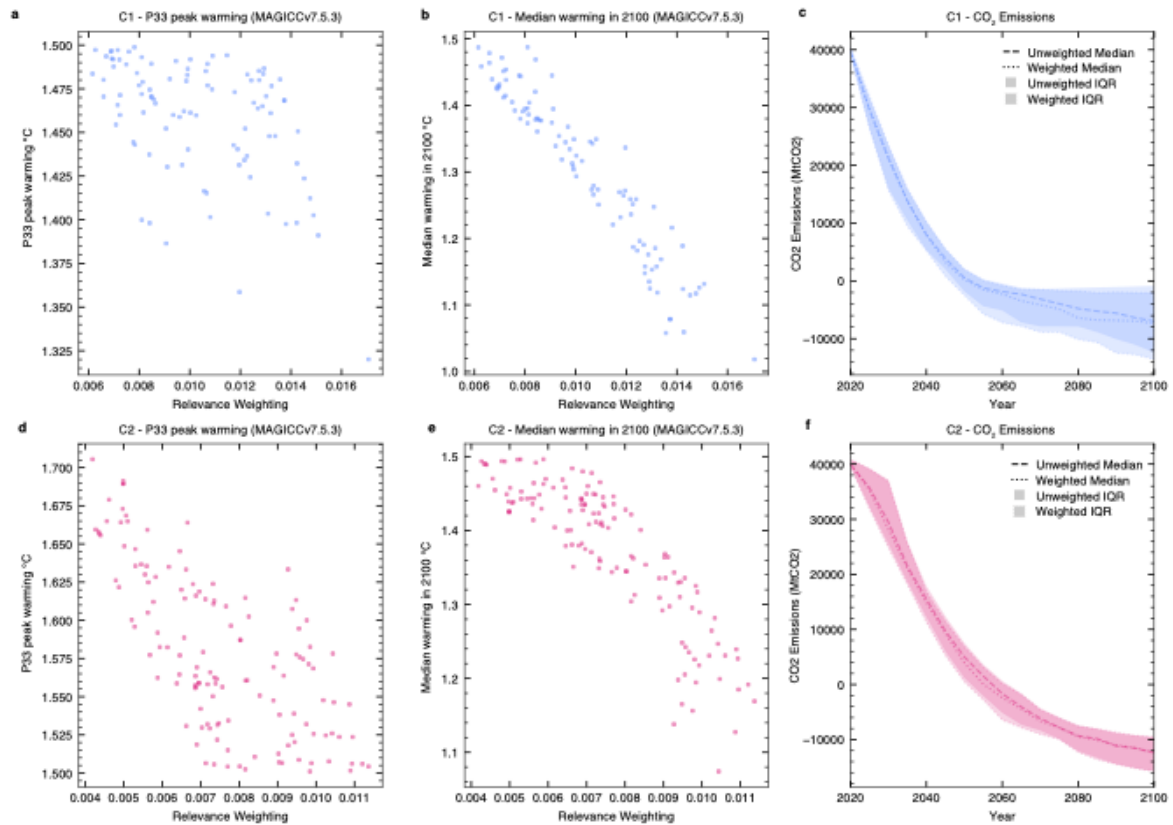
In the main text, our relevance weighting application to AR6 scenarios adopts a binary weighting based on assessed temperature category. Here we present an illustrative example of a weighting approach that maintains categorisation by temperature outcomes but places a higher weighting on ‘lower-risk’ scenarios within each. Here, our definition of risk is focused on global warming outcomes and is based only to the metric(s) used for temperature categorisation. Our risk-averse relevance weighting $R(i)$ is defined by

$$R(i) = I_c(i) \cdot \sum_{j=1}^n w_j \cdot \frac{1}{1 + \exp(k \cdot (m_{ij} - \widetilde{m}_j))}$$

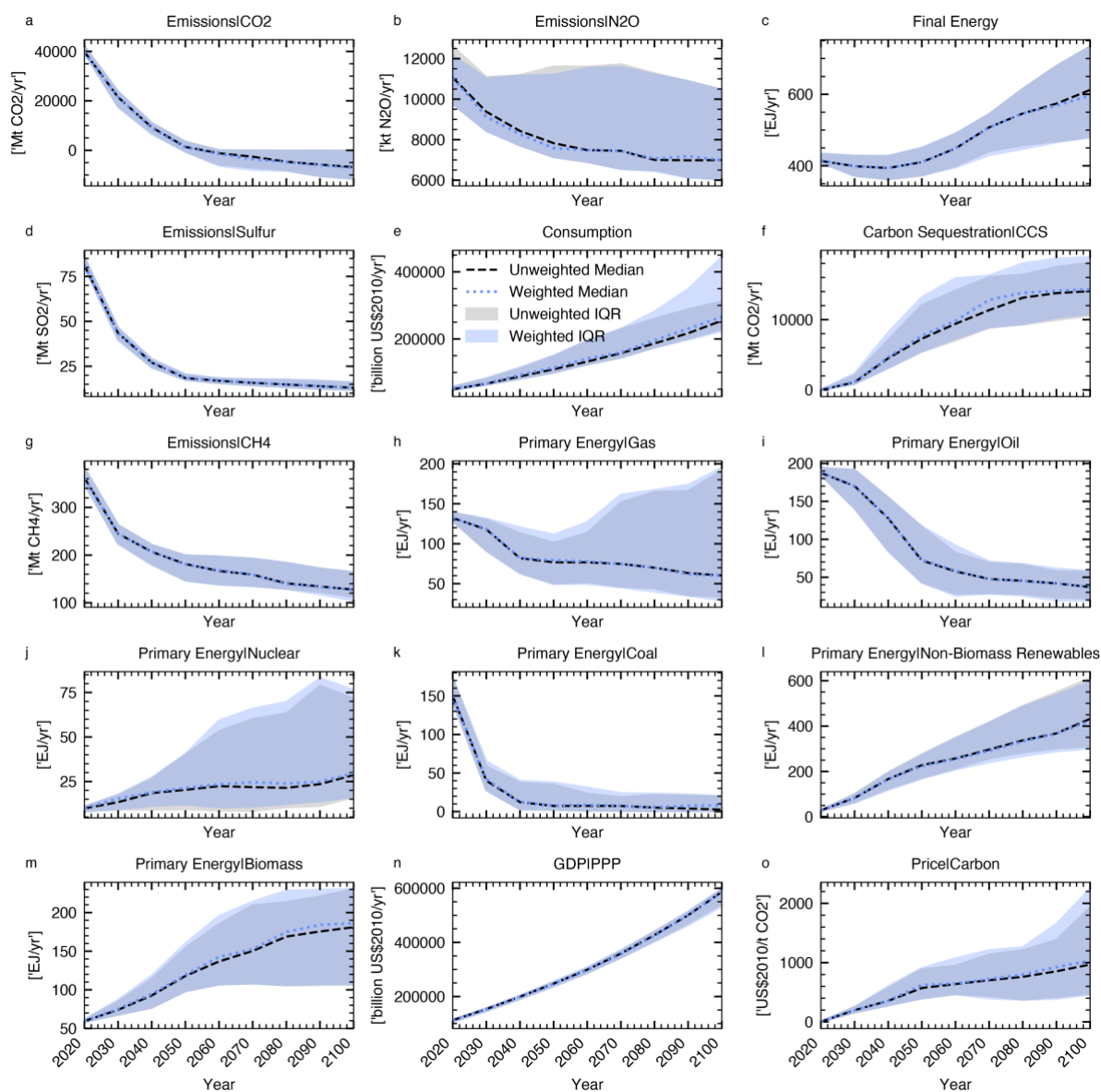
where $I_c(i)$ is a binary indicator (1 if scenario (i) is within the temperature category, 0 otherwise), w_j is the weight of the assessment metric for categorisation, j ; and m_{ij} is the scenario value of the assessment metric. This uses an S-curve, with the slope defined by k , greater down weighting occurs closer to the category threshold, and reduced upweighting further away from the median, $\widetilde{m}_j$.

We apply this relevance weighting approach to scenarios in the C1 and C2 temperature categories, both of which use the same temperature categorisation metrics (P33 peak warming and median warming in 2100) but differ based on their level of overshoot (Supplementary Figure 1). Under this example, we assume that a user wishes to main integrity of existing temperature categorisation; hence the explicit division between C1 and C2. When we apply a risk-averse relevance weighting, as would be expected, a more stringent CO₂ emissions trajectory is seen for both categories (Supplementary Figure 1c and f).

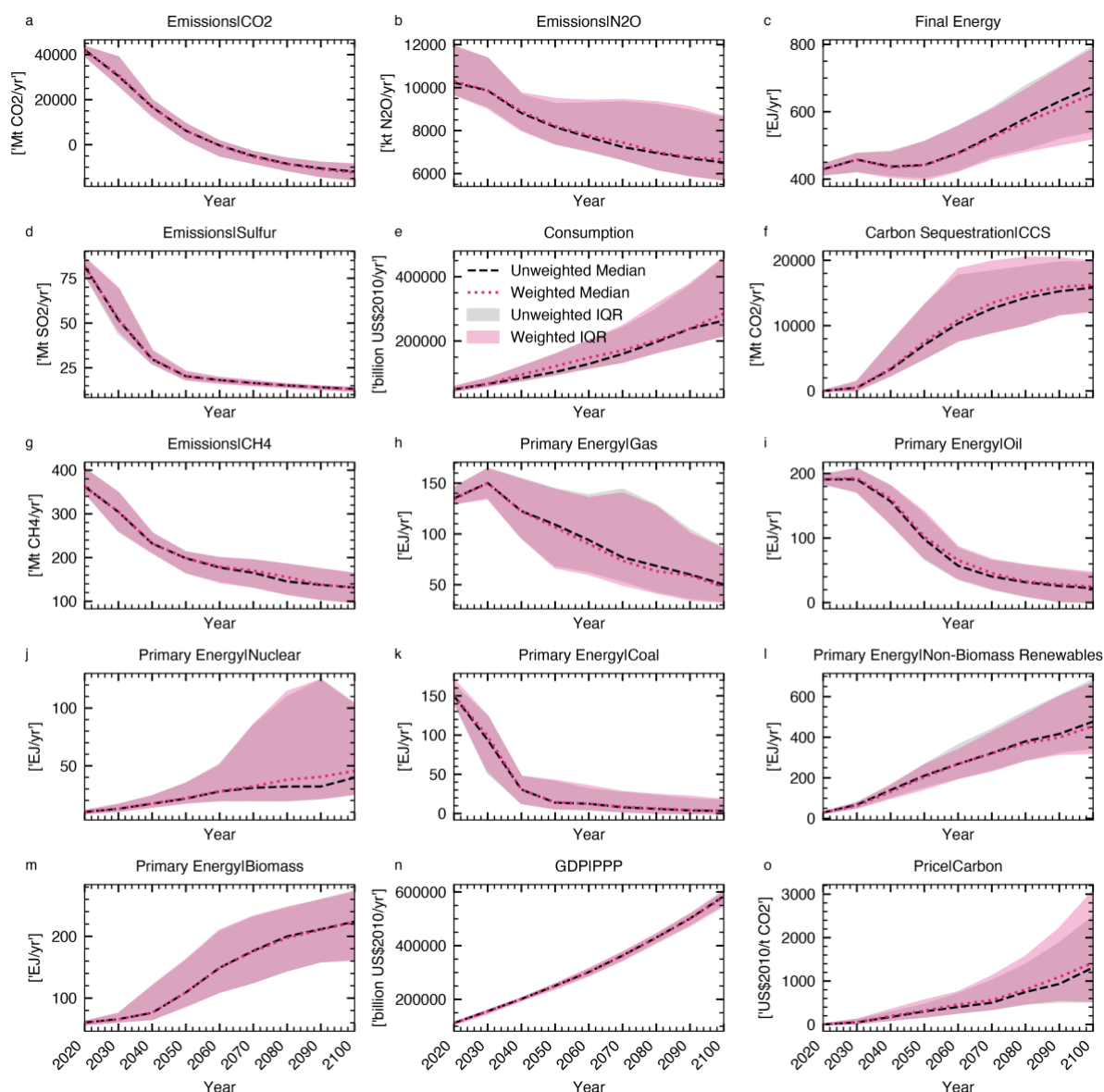
This is an illustrative example of how a continuous relevance weighting based on a user's needs could be applied. This could be further adapted based on the context, for example placing more weight on a particular temperature categorisation metric.



Supplementary Figure 1 | Figures showing implementation of risk-averse relevance weighting. Panels a and b shows C1 scenarios overall relevance weighting plotted against peak warming and median peak warming respectively. Panel c shows the weighted and unweighted CO₂ emissions trajectories for C1 scenarios. Panels d, e and f mirror the preceding panels but for the C2 scenario subset.



Supplementary Figure 2 | Timeseries of variables C1 diversity reweighted Panels a-o show each of 15 variables reported by all scenarios, comparing the median and interquartile range for the weighted and unweighted distributions. The black dotted lines represent unweighted medians, the blue dotted lines weighted medians. Grey and coloured shaded areas are the interquartile ranges for the unweighted and weighted distributions, respectively (key shown in panel e, IQR=Interquartile range).



Supplementary Figure 3 | Timeseries of variables C2 diversity reweighted Panels a-o show each of 15 variables reported by all scenarios, comparing the median and interquartile range for the weighted and unweighted distributions. The black dotted lines represent unweighted medians, the blue dotted lines weighted medians. Grey and coloured shaded areas are the interquartile ranges for the unweighted and weighted distributions, respectively (key shown in panel e, IQR=Interquartile range).

Supplementary Results 2: Sensitivity of climate action benchmarks to different variable weights and sigma values

Here we explore a range of different variable weighting schemes combined with different sigma values. This sensitivity analysis is not intended to be comprehensive; but demonstrates how adjusting certain weighting parameters alters the outcomes. Further we can identify changes in ensemble medians that maintain the directionality regardless of inputs used.

We explore the following weighting of variables used:

- 'Expert': variable weights based on expert judgment across 15 core variables reported by all scenarios. Variable weighted reported in Table 1.
- 'Correl': correlation adjusted weights, with only 8 variables used (See Supplementary Methods 1 and Table 1)
- 'Energy_only': inclusion of only the energy category variables under our expert weight scheme, so the energy group weight becomes 1.
- 'Emissions_only' inclusion of only the emissions category variables under our expert weight scheme, so the emissions group weight becomes 1.

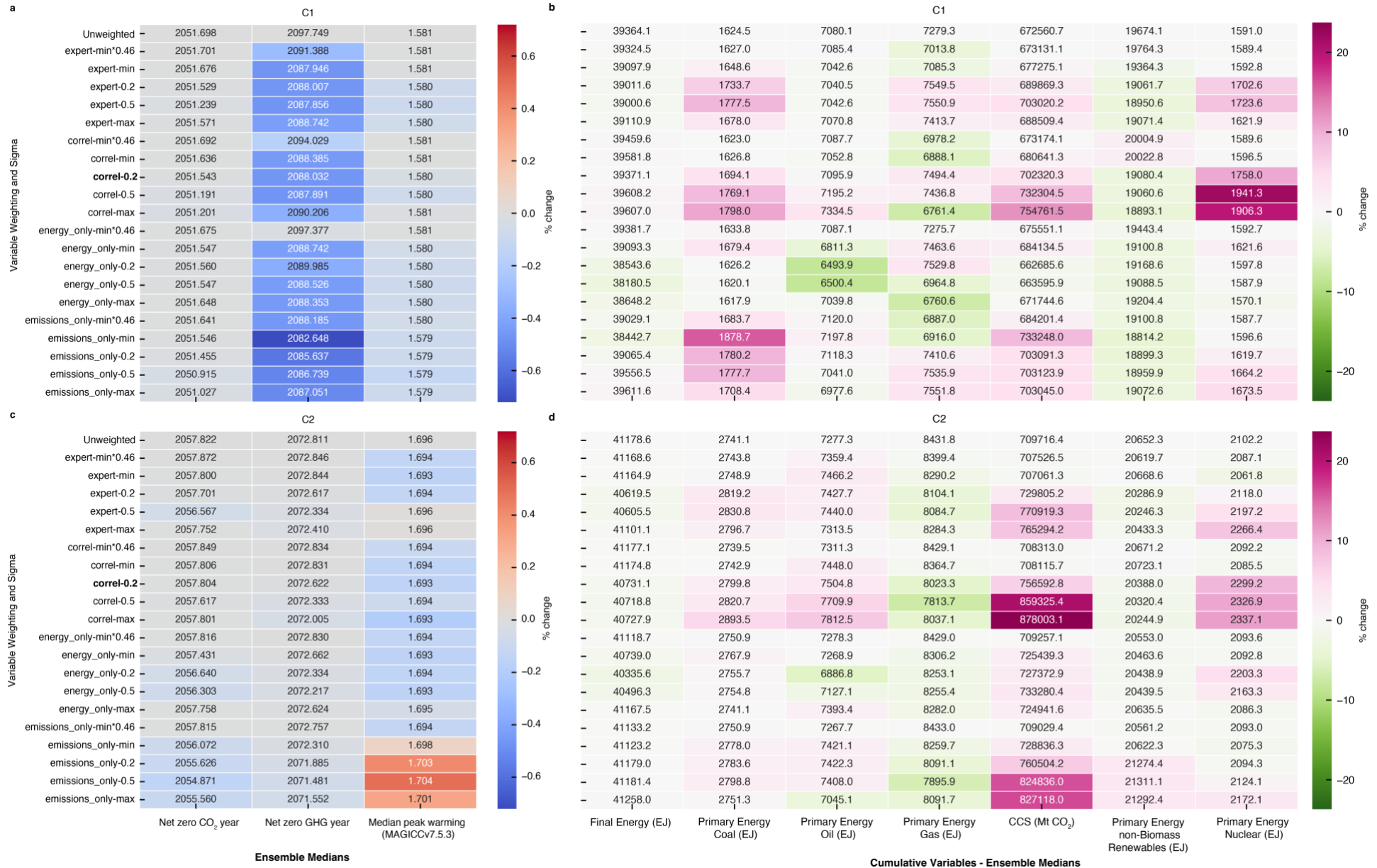
For each of these sets of variable weights, we test 5 sets of sigma values between a fraction of the minimum of model differences of the same RCP-SSP combination to the maximum (see Supplementary Results 3).

Firstly, there ensemble changes that are directionally maintained regardless of the weighting inputs, and those that are not (Supplementary Fig. 4). These include, earlier median net-zero GHGs for C1, increased cumulative CCS for C2, reduced cumulative gas for C2, or increased cumulative coal use for C1 and C2. The median net zero CO₂ year for C1 and C2 are earlier across all tested inputs, but for most the movement is minor (Supplementary Fig. 4a and c) Adjustments to variable weights and sigma values naturally changes the emphasis of what defines diversity in the database. observing changes that remain across a range of inputs highlights signals that are robust across variables, while others are sensitive to methodological choices and would vary depending on the question being asked.

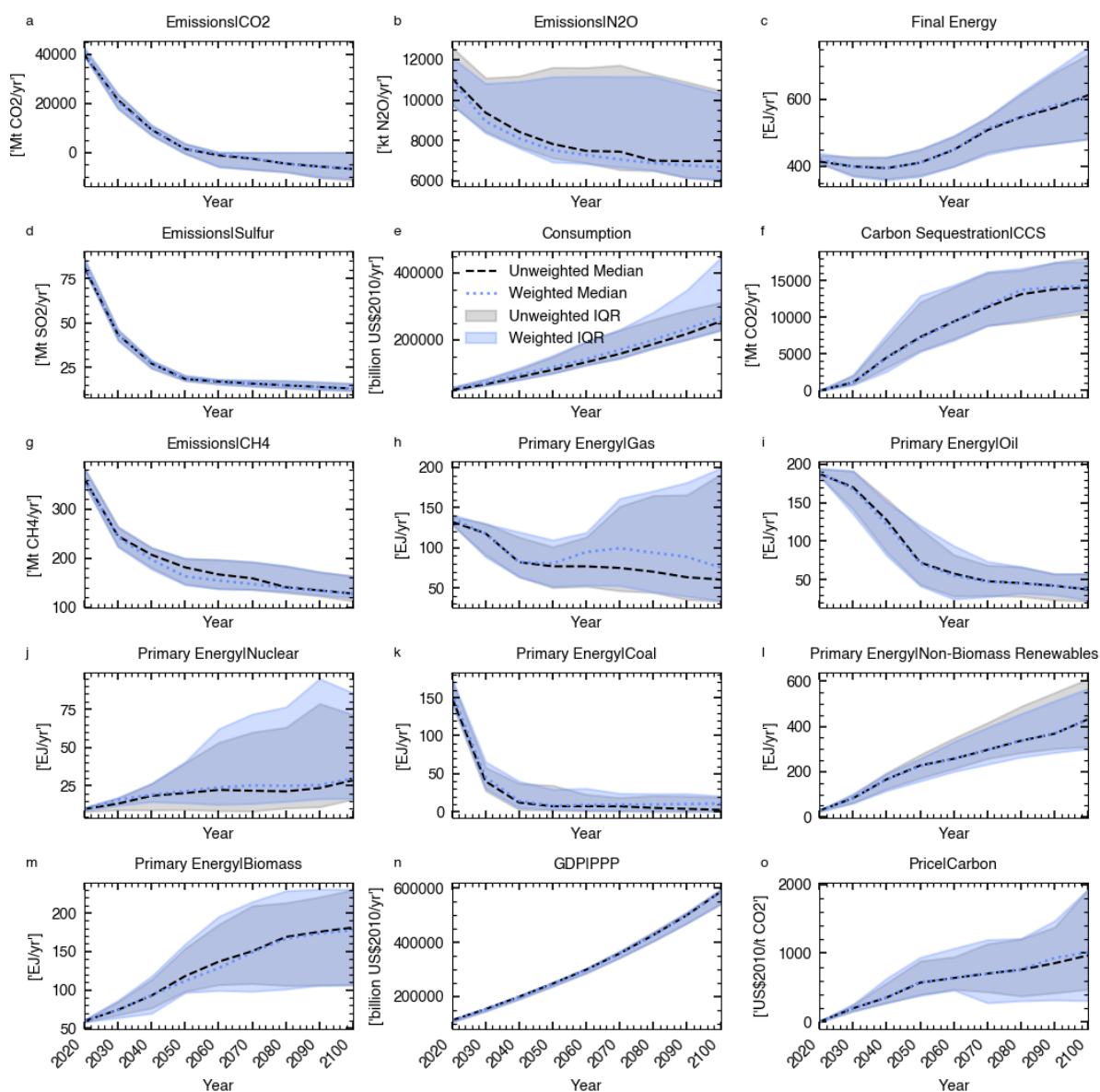
We can see in the data patterns relating to choice of sigma values, for example some quantities exhibiting bigger changes at the high or low ends of our sigma choices. This is clear for CCS for example, which has a greater upward shift at higher sigma values across our variable weight choices. Our selection of groups of sigma values at intervals means, that the spread of diversity weights for variable adjusts differently between sigma intervals (Supplementary Figure 7). As such, some variables exhibit an increased spread above our central chosen sigma (0.2). This choice was based on an even spread in diversity weights across all variables. Opting for a different choice represents a greater unevenness in the spread across our 15 variables (Supplementary Results 3). So, while stronger impacts for specific quantities can be seen at lower or higher sigma values, these reflect a weighting configuration that disproportionately emphasises certain variables which see a greater spread at higher or lower sigma values. Depending on the research question, however, if fewer variables are of interest, sigma values could be adjusted to maximise spread in those dimensions.

Finally, we can observe some features in the data between variable weight schemes. For example, when applying the emissions only variable weights, we see that for all but the lowest sigma interval, there are larger reductions in the C2 CO₂ and GHG net zero years, when compared to using a variable weight scheme that includes all variable types. Under this scheme, diversity is only being defined by differences in emissions trajectories, and therefore scenarios diverging on their emissions pathways will achieve greater weight. For C2, this leads to a more pronounced effect on our test quantities.

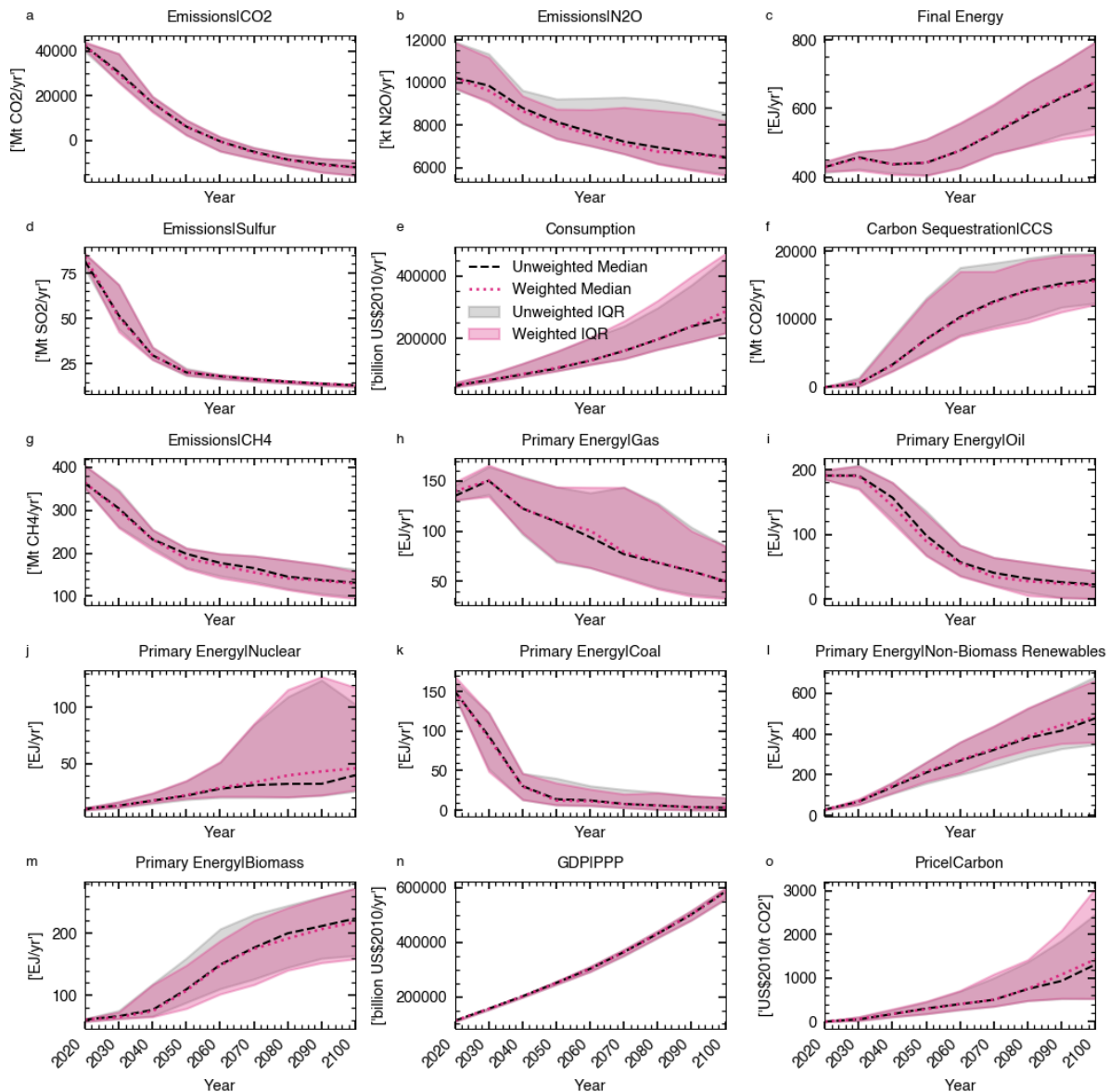
Whilst we observe some features in our data that are robust to adjustments in sigma values and variable weights, as would be expected, these results highlight that how diversity is defined may impact the outcomes for the ensemble



Supplementary Figure 4 | Heatmaps showing impact of alternative diversity weighting inputs on ensemble outcomes (medians). Panel **a** and **c** showing impact on net zero years and peak temperature for C1 and C2 scenarios respectively; panels **b** and **d** showing impact on a range of variables, median values and cumulative 2020-2100 for C1 and C2 scenarios respectively. Cell colours are derived from the % difference from the corresponding unweighted values, shown in the top row. For panels a and c, red denotes increases when compared with unweighted ensemble medians, and blue decreases. For panels b and d, pink = increases and green = decreases when compared with unweighted medians. The results presented in the main paper are identifiable as the bold listed item 'correl-0.2' on the x axis. For guidance on weighting inputs see description above in Supplementary Results 2.



Supplementary Figure 5 | Timeseries of variables C1 quality reweighted Panels a-o show each of 15 variables reported by all scenarios, comparing the median and interquartile range for the weighted and unweighted distributions. The black dotted lines represent unweighted medians, the blue dotted lines weighted medians. Grey and coloured shaded areas are the interquartile ranges for the unweighted and weighted distributions, respectively (key shown in panel e, IQR=Interquartile range).



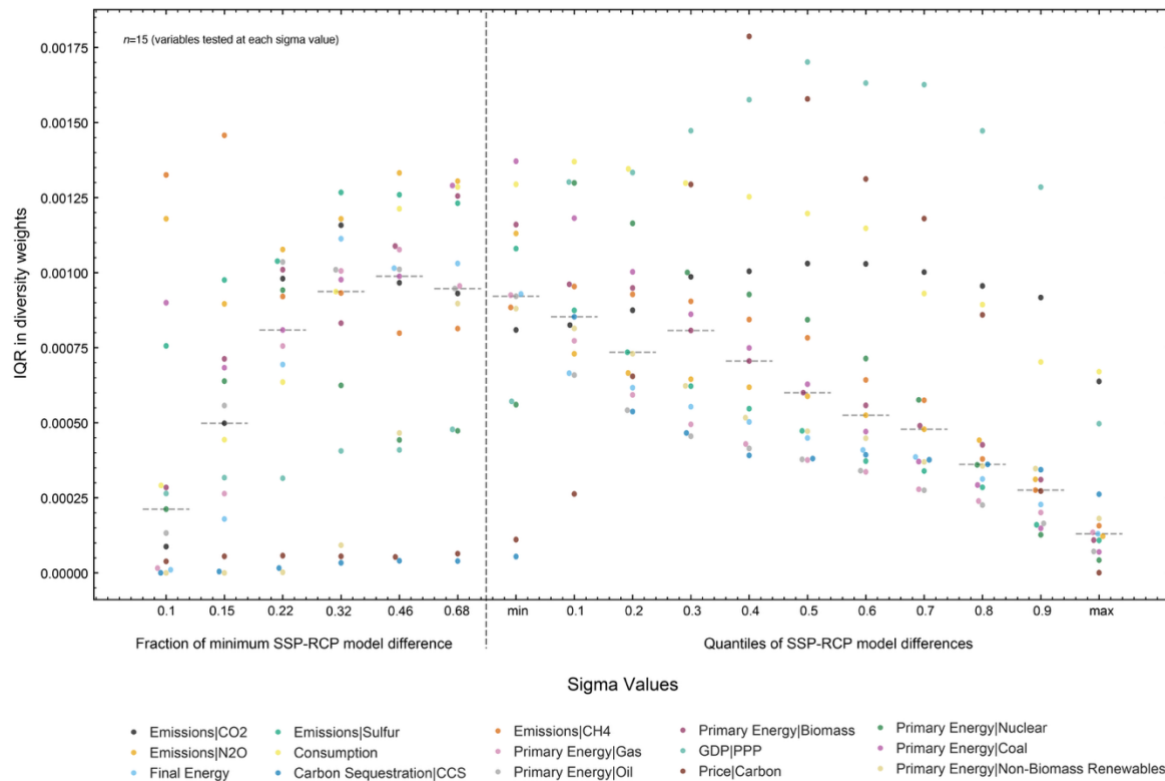
Supplementary Figure 6 | Timeseries of variables C2 quality reweighted Panels a-o show each of 15 variables reported by all scenarios, comparing the median and interquartile range for the weighted and unweighted distributions. The black dotted lines represent unweighted medians, the blue dotted lines weighted medians. Grey and coloured shaded areas are the interquartile ranges for the unweighted and weighted distributions, respectively (key shown in panel e, IQR=Interquartile range).

Supplementary Results 3: Impact of different sigma values on spread of diversity weights

The choice for what constitutes scenario similarity is crucial for how the diversity weighting approach operates in practice. Sigma values that are very small relative to ensemble variation, will see only very few scenarios fall within it, this will result in no or minimal reweighting for most scenarios. Conversely, large sigma values relative to ensemble variation will result in scenarios effectively not being reweighted as they are all treated as similar.

Sigma choice will reflect the needs of the user in their application of the weighting procedure. A user may want to implement a specific and known set of sigma values informed by a definition of scenario similarity. This could be guided by distances between certain scenarios. In this context, even if the sigma's are small in relation to ensemble variation, it could be applied as a robustness check across an ensemble. However, an alternative approach is to test the impact of a range of sigma values with the final choice determined by the impact on the ensemble.

Here, we perform an illustrative example of the latter approach, exploring the impact of a range of sigma values on the spread of weights, by variable. To do this, we stay with our conceptual definition of scenario similarity: the RMS of SSP-RCP combinations, run by different models. We test a range of sigma values between the minimum and the maximum of the SSP-RCP combinations, at 10 percentile intervals to explore the impact on the spread of diversity weights (Supplementary Figure 7). To explore the point at which most variables have seen a collapse in their IQR, we expanded our assessment below the minimum, SSP-RCP model differences testing fractions of the minimum, in a log-spaced grid from $0.1 \cdot \min$ to $1 \cdot \min$ (6 intervals) (Supplementary Figure 7)



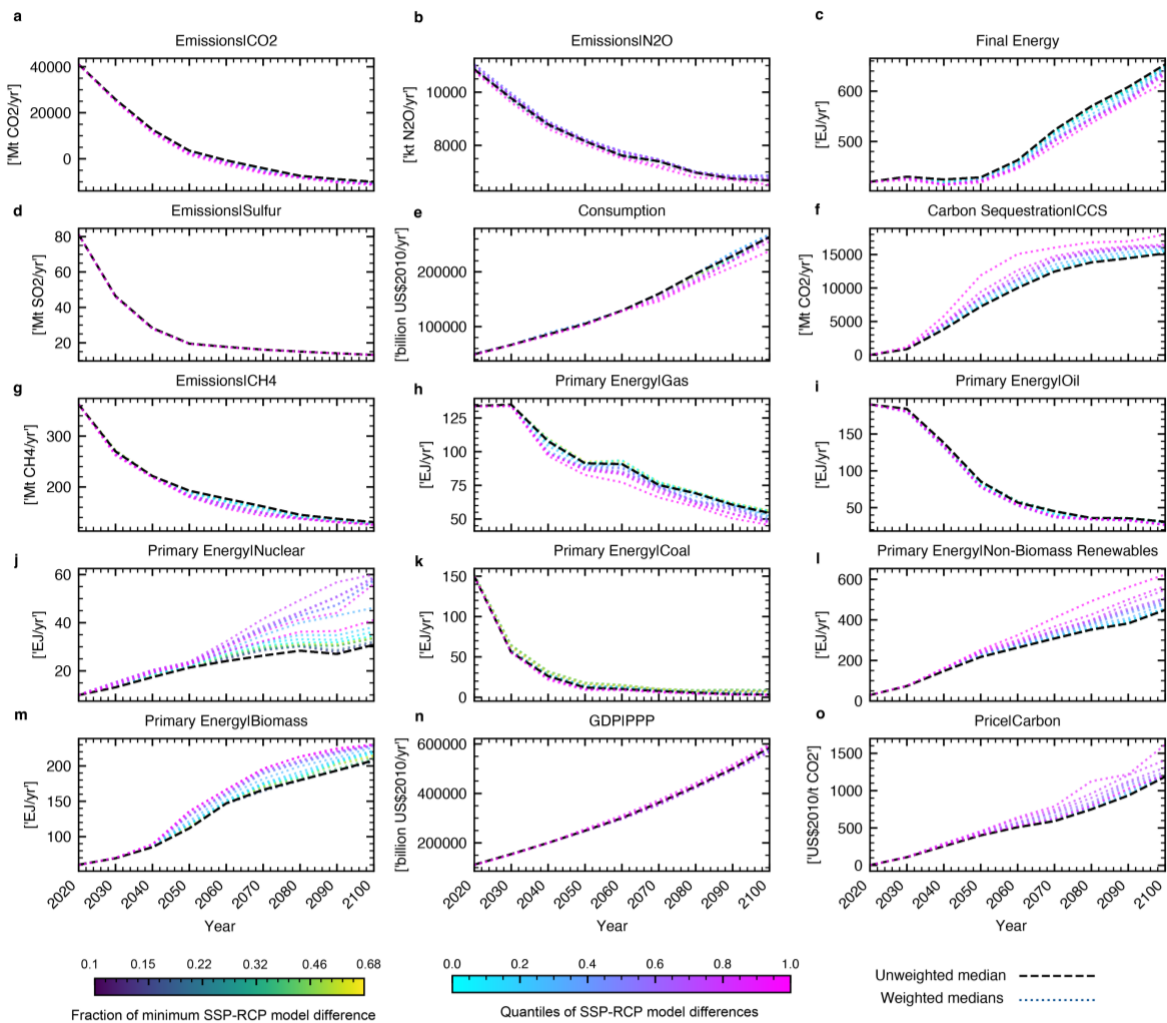
Supplementary Figure 7 | Impact of sigma values on spread of diversity weights by weighting variable. Swarm plots showing the interquartile range in the variable-specific diversity weights at a range of sigma values between the maximum and minimum of RMS difference of SSP-RCP combinations run by the same model. Values were taken at 10 percentile intervals and adjusted in tandem with each other. We also show fractions of the minimum, in a log-spaced grid from $0.1 \cdot \min$ to $1 \cdot \min$ (6 intervals). Median lines shown by grey dotted lines.

As would be expected, the IQR in the diversity weights varies by variable, with the highest median IQR at our $0.46 \cdot \min$ set of values (Supplementary Figure 2). The median IQR declines sharply towards the $0.1 \cdot \min$ values, however the range increases with some variables (e.g., CH_4 , N_2O) with IQR appearing to be increasing beyond our minimum sigma values. Above the 30th percentile of SSP-RCP model differences, the medians steadily decline, with all variables declining by the max values. Although the maximum median IQR $0.46 \cdot \min$, there is a

large range in the IQR values, with some variables, e.g., Carbon Price and CCS where the respective values at this interval have minimal reweighting effect as they are too small. At a higher interval, around the 20th and 30th percentiles, although there is a lower median IQR, the data indicate that all variables are seeing some reweighting. This test of sigma values within a specified range and definition of scenario similarity allows us to stick with the conceptual choice but ensure that the values used will lead to effective reweighting.

Next, taking the diversity weights calculated for each of our variables, before combining, we can explore the impact at the variable level for each set of sigma values. This is important for understanding how sigma values influence each variable and the final reweighted outcomes in the ensemble. Looking at C1 and C2 scenarios (Supplementary Figure 3), we can see that for many variables, regardless of which sigma values used, reweighting would indicate moves likely to represent a strengthening of mitigation action (e.g., more low carbon primary energy and CCS (Supplementary Figure 3, panels j, l, m and f). However, this is not universally true, for example for primary energy from gas and coal, the lower sigma values see the median gas trajectory higher than the unweighted ensemble median.

An important point to recognise is that we perform our diversity weighting and sigma selection across the whole AR6 database. The selection of sigma values to ensure a spread of diversity weights across the ensemble may result in sigma values that have little effect for our subset.



Supplementary Figure 8 | Impact of sigma values on median timeseries for all C1 and C2 scenarios, applying variables weights to specific variables. Panels a-o show each of our weighting variables comparing the median of reweighted variables across a range of sigma values, with the unweighted median. Coloured lines represent

the reweighted median lines, with the minimum to the maximum of the RCP-SSP model differences shown at each 10 percentile intervals and fractions of the minimum, in a log-spaced grid from 0.1*min to 1*min (6 intervals)

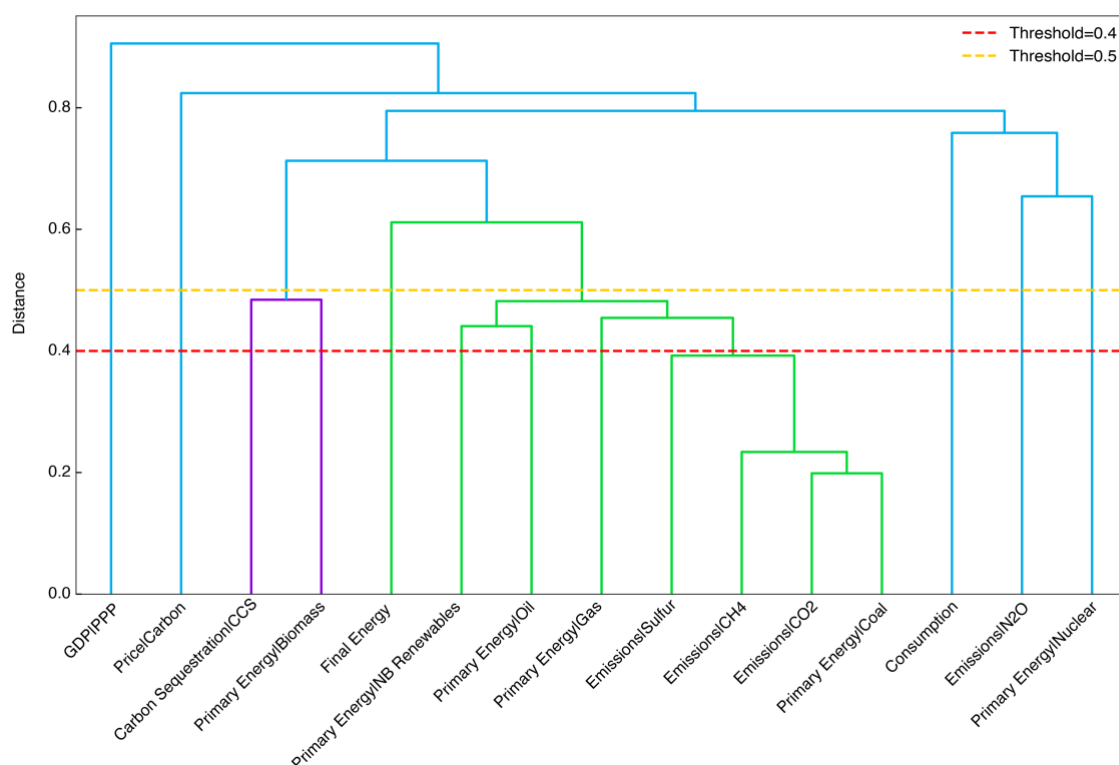
Supplementary Methods 1 | Calculation of Correlation Adjusted Variable Weights

Some variables in the database are correlated, meaning that a set of weights might amplify certain signals. Here we perform some simple analysis to derive a set of illustrative variable weights that reduce this effect. To adjust for correlation between our variables, we computed the correlation between our pairs of variables across all vetted scenarios at decadal intervals (2020-2100). We used the mean of the decadal correlations to provide single correlation values for all variable pairs. Variables with correlation above 0.5 are shown in Supplementary Table S4.

Supplementary Table S4 | Weighting variable pairs with correlations high than 0.6.

Pair	Correlation
Emissions CH4 - Emissions CO2	0.756
Emissions CH4 - Emissions Sulfur	0.624
Emissions CO2 - Final Energy: 0.663	0.663
Emissions CH4 - Primary Energy Coal	0.776
Emissions CO2 - Primary Energy Coal	0.801
Emissions CO2 - Primary Energy Gas	0.659
Emissions CO2 - Primary Energy Oil	0.671
Emissions Sulfur - Primary Energy Coal	0.608

The absolute correlations were converted into distance metric ($\text{Distance} = 1 - |\text{correlation}|$) to be used in the hierarchical clustering process. We used the Scipy hierarchical clustering package to perform clustering³. Clusters were formed using the average linkage method, merging clusters on the mean pairwise distance between cluster members. Looking at clusters between thresholds of 0.4 and 0.5, 0.4 provides only one cluster of more strongly correlated variables, whilst raising the threshold to 0.5 adds further variables to clusters (Supplementary Figure 9). We adopt 0.5 for our illustrative use case as it represents greater redundancy removal to be observed in our diversity reweighting results. For each cluster with multiple variables, we select an illustrative variable for each cluster, using the one with the highest weighting in our original weighting scheme. The result is a set of variable weights that reduce redundancy and for variables present, we adopt an even weighting across the variables as we have preserved emphasis on originally higher weighted variables in selection of our representative cluster variables.



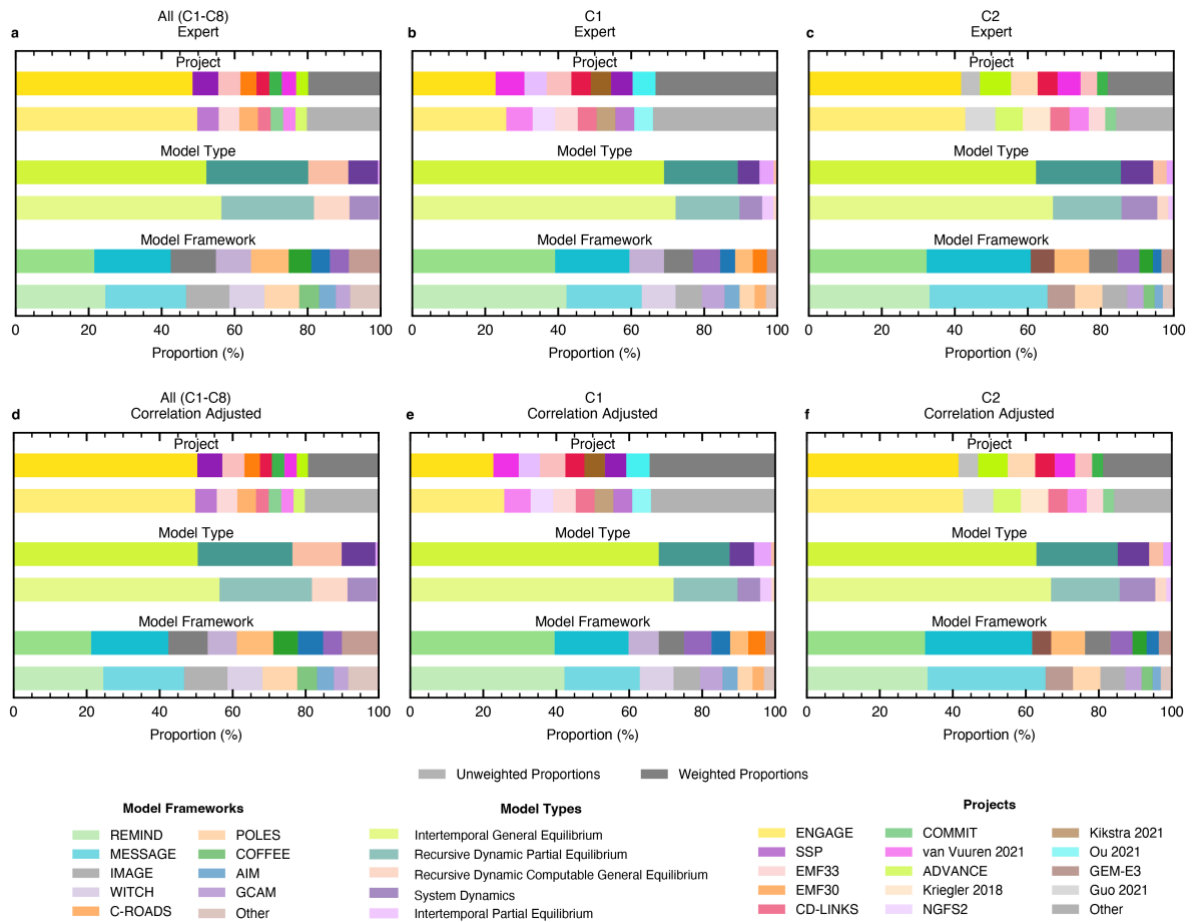
Supplementary Figure 9 | Dendrogram visualising hierarchical clustering of variables. Colours are used to describe variables that eventually fall into clusters at different distance thresholds. Dotted lines are used to show the difference in clustering at a threshold of 0.4 and 0.5.

Supplementary Results 4: Impact of diversity weighting on database characteristics and exploration of scenario weights

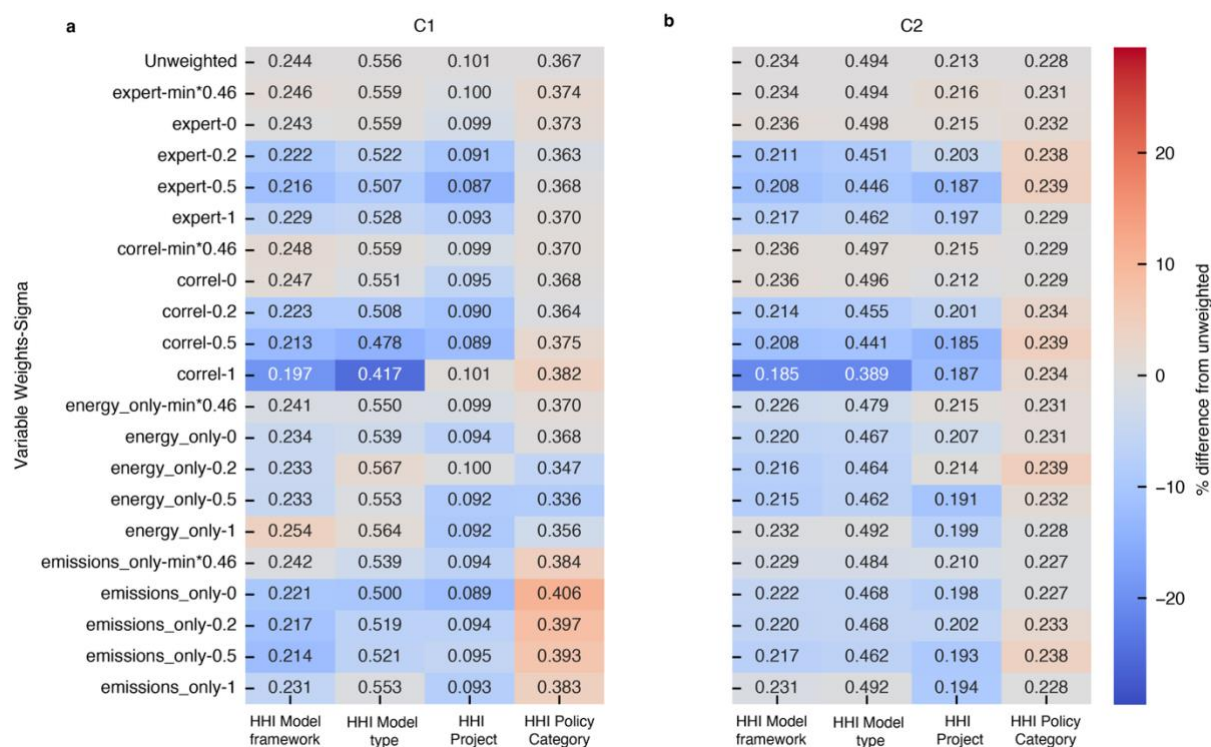
Here, we explore how a range of reweighting approaches impact the distribution of projects, model types and model frameworks and policy category.

Firstly, visualising how database characteristics change in unweighted and weighted ensembles, we show a some two examples of weighting procedures for our 1.5 degrees scenarios and the database as a whole. We show sigma using the 20th percentile of RMS RCP-SSP model differences, identified as having a weighting impact across variables (see Supplementary Results 2). We combine it with our ‘expert’ variable weighting (see Table 1) (Supplementary Figure 4 panels a-c) and our correlation adjusted weighting (panels d-f) (Supplementary Methods 1). For these weighting schemes, there are modest but visible differences in the unweighted and weighted proportions of project, model type and model framework. Although not universal, it appears the reweighting procedure under these conditions results in more balanced distribution of project, model type and model frameworks. For example, for all subset examples and both the expert and correlation adjusted weights, we see the REMIND model framework with a reduced share of scenarios. For project distribution there is a reduction in the share of the ‘ENGAGE’ project in all but one of our examples. Under our correlation adjusted variable weights, across the whole database, the share of ENGAGE scenarios increases slightly.

Secondly, we test a range of weighting inputs and their impact on the concentration of model frameworks, model types, projects and policy categories. Here we explore the same weighting inputs as in Supplementary Results 2. Although not universal, improvements in project, model framework and model type are robust to a range of alternative inputs for C1 and C2. However, when sigma values are high or low (uneven impact across variables) there are instances where these quantities are worsened. Diversity in scenario policy category is worsened under almost all weighting approaches.



Supplementary Figure 10 | How diversity reweighting impacts distribution of projects, model types and model frameworks in C1-C8, C1, and C2 scenarios. Each panel displays the unweighted and weighted proportions of each project, model type and model framework for a subset of scenarios, and for a reweighting procedure. Panels a and d show C1-8 scenarios; panels b and e show C1 scenarios; and panels c and f show C2 scenarios. The top row of panels show sigma of 20th percentile of RCP-SSP model differences, see Supplementary Results 2) with our expert variable weighting. The second row of panels show the same sigma values but with our correlation adjusted variable weighting (see Supplementary Methods 1).



Supplementary Figure 11 | Heatmaps showing impact of a range of weighting inputs on Herfindahl-Hirschman Index (HHI). Lower scores indicate improved evenness of model framework, model type, project of policy category. For guidance on weighting examples tested, see Supplementary Results 2. Colours of heatmap cells are based on % difference in HHI value when compared to the unweighted equivalent in the top row. Panel **a** C1 scenarios, and panel **b** C2 scenarios.

References

- Byers, E. *et al.* AR6 Scenarios Database. Zenodo <https://doi.org/10.5281/zenodo.5886912> (2022).
- 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).
- Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).