Climatic Change (2023) 176:80
https://doi.org/10.1007/s10584-023-03529-w

Revealing effective regional decarbonisation measures to limit global temperature increase in uncertain transition scenarios with machine learning techniques

Pei-Hao Li[1] · Steve Pye[1] · Ilkka Keppo[2] · Marc Jaxa-Rozen[3] · Evelina Trutnevyte[3]

Received: 26 April 2022 / Accepted: 7 April 2023

Abstract
Climate change mitigation scenarios generated by integrated assessment models have been extensively used to support climate change negotiations on the global stage. To date, most studies exploring ensembles of these scenarios focus on the global picture, with more limited attention to regional metrics. A systematic approach is still lacking to improve the understanding of regional heterogeneity, highlighting key regional decarbonisation measures and their relative importance for meeting global climate goals under deep uncertainty. This study proposes a novel approach to gaining robust insights into regional decarbonisation strategies using machine learning techniques based on the IPCC SR1.5 scenario database. Random forest analysis first reveals crucial metrics to limit global temperature increases. Logistic regression modelling and the patient rule induction method are then used to identify which of these metrics and their combinations are most influential in meeting climate goals below 2 °C or below 1.5 °C. Solar power and sectoral electrification across all regions have been found to be the most effective measures to limit temperature increases. To further limit increase below 1.5 °C and not only 2 °C, decommissioning of unabated gas plants should be prioritised along with energy efficiency improvements. Bioenergy and wind power show higher regional heterogeneity in limiting temperature increases, with lower influences than aforementioned measures, and are especially relevant in Latin America (bioenergy) and countries of the Organisation for Economic Co-operation and Development and the Former Soviet Union (bioenergy and wind). In the future, a larger scenario ensemble can be applied to reveal more robust and comprehensive insights.

Keywords Uncertainty · Energy transition · Machine learning · Scenario ensembles · Regional analysis

1 Introduction

To mitigate the threats posed by climate change, international delegates from 196 countries signed up to the Paris Agreement in 2015 to limit global warming to well below 2 °C, preferably to 1.5 °C, above the pre-industrial levels (UNFCCC 2015). In order to

✉ Pei-Hao Li
peihaoli@gmail.com

Extended author information available on the last page of the article

achieve these climate goals, it is essential to reach carbon neutrality by the middle of the century (IPCC 2018; IEA 2021), meaning that global energy systems must be dramatically transformed to reduce greenhouse gas (GHG) emissions, using a whole raft of radical policy interventions.

Integrated assessment models (IAMs) have been an important set of tools to explore how the transformation of the global energy system can be achieved to meet the climate targets. However, this system transformation is subject to multiple uncertainties, due to factors such as the rate of technology advancement, social acceptance, and political priorities (Rotmans and van Asselt 2001; van Asselt and Rotmans 2002; Kirchner et al. 2021). Analysis of a large ensemble of IAM scenarios provides an opportunity to capture a larger portion of that uncertainty than in conventional methods, like scenario matrix, and such scenario ensembles can help draw robust insights into long-term transitions (Trutnevyte et al. 2016; Guivarch et al. 2017). For example, more than 400 scenarios were collated for the evaluation of strategies to limit temperature increase to 1.5 °C in the Intergovernmental Panel on Climate Change (IPCC)'s special report SR1.5 (IPCC 2018).

In the past, however, studies aiming to gain insights from scenario ensembles have put a stronger focus on decarbonisation measures at the global level, paying less attention to regional heterogeneity of actions (Guivarch et al. 2016; Rogelj et al. 2018; Jaxa-Rozen and Trutnevyte 2021), most likely due to the challenges in investigating highly complex high-dimensional datasets. However, regional heterogeneity in mitigation may be significant due to variation in regional energy mixes, renewable energy potentials (e.g. bioenergy and solar power), costs, and access to technology (Shiraki and Sugiyama 2020; Azevedo et al. 2021). Understanding such differences could provide new insights into the need for regional support, opportunities in the low-carbon transition, or dependencies on regional action for meeting global climate goals (Brutschin et al. 2021).

Most current studies use visualisation (Meyer et al. 2021; DeAngelo et al. 2021) or simple statistics (Fujimori et al. 2019; Abernethy and Jackson 2022) to reveal the impacts of a limited number of decarbonisation measures on global temperature increases. These approaches can only reveal linear relationships between a small number of options and a climate goal of interest, without considering non-linear interactions between those options in complex energy systems. Hence, multiple-dimensional scenario ensembles cannot be well investigated by these simple approaches. Lately, more advanced methods, such as the patient rule induction method (PRIM), have been used to better reflect complex system dynamics in uncertain transition scenarios (Rozenberg et al. 2014; Guivarch et al. 2016). Nonetheless, in previous applications, only a small set of measures was selected for investigation based on expert judgement, which might be misleading and biased or might miss key influential factors. Therefore, a systematic approach that can simultaneously investigate the influences of numerous potential measures on global climate change is still lacking.

To assist policymakers and the research community to have a better understanding of regional actions in achieving global climate goals, we propose a systematic approach, based on machine learning techniques, to analysing high-dimensional scenario ensemble, comprehensively considering a wide range of plausible measures, along with non-linear interactions between those measures in the energy system transitions. A particular focus is on the influence of regional actions on reaching the climate goals of 2 °C and 1.5 °C. Hence, the contribution of this study is twofold.

- A systematic approach based on machine learning techniques is introduced to gain robust insights from high-dimensional uncertain system transition ensembles, comprehensively considering a wide range of potential regional factors.
- The role of regional measures in global climate change mitigations is explored, and their relative importance is revealed.

2 Overview of methods

In the literature, gaining insights into global decarbonisation from IAM scenario ensembles has tended to rely on the visualisation of plots or on descriptive statistics. However, this approach can only show relationships between two or three metrics at a time, ignoring the influences of other metrics. It therefore relies heavily on researchers' knowledge of global systems and interpretations of the plots. For instance, Meyer et al. (2021) used a series of box plots to analyse the influences of GDP, energy intensity, and emission intensity on carbon prices in the IPCC SR1.5 scenario ensemble individually, without showing the influences of multiple factors. Box plots were also widely used to reveal characteristics of energy systems for different temperature targets. With this approach, Gambhir et al. (2019) revealed major characteristics of deep mitigation scenarios in the IPCC AR5 database, such as low primary energy intensity and a high share of low-carbon electricity. Similarly, no interactions between factors have been investigated.

To better understand system dynamics in uncertain scenarios, simple statistical methods have also been applied to quantitatively analyse the influences of decarbonisation metrics. This approach can still only consider a limited number of pre-defined variables of interest. For instance, Abernethy and Jackson (2022) investigated the relationship between peak temperature and the year such temperature is reached across 213 scenarios using linear regression. Likewise, a linear regression model was established to show that cumulative CO_2 emissions decrease as carbon prices increase based on the SR1.5 scenarios (Fujimori et al. 2019). Linear regression models have also been applied to explore how regional energy characteristics (e.g. total consumption and primary energy use) determine the deployment of carbon dioxide removal (CDR) in the SR1.5 scenario ensemble (Diniz Oliveira et al. 2021). Most of these studies only used one or two factors based on researchers' judgements, without taking into account a wide range of potential factors.

Recently, more advanced techniques have been adopted to reveal complex, non-linear system dynamics and uncertainties in scenario ensembles. For example, to uncover diverse drivers to specific scenario outcomes, Guivarch et al. (2016) used a modified PRIM (Friedman and Fisher 1999) to identify various combinations of socio-economic drivers that can lead to two global decarbonisation archetypes, shared socio-economic pathways SSP4 and SSP5, based on 432 uncertain global transition scenarios. Similarly, PRIM has been successfully applied to discover challenges to mitigation and adaptation in the SSP framework (Rozenberg et al. 2014) and key drivers to various global copper price scenarios under deep uncertainty (Kwakkel et al. 2013). CART (classification and regression trees) (Breiman et al. 2017) is another widely adopted method to identify the main drivers for specific scenario outcomes. For instance, Gerst et al. (2013) used CART to determine how various conditions of technology innovation and energy intensity improvement lead to five typical global transition scenarios, from 200 scenarios, using a hierarchical clustering algorithm. Lately, Jaxa-Rozen and Trutnevyte (2021) used a spectral clustering algorithm to categorise 1392 global scenarios from IPCC databases as well other academic and influential grey literature into four typical scenario archetypes, including high electrification with low CO_2 emissions, middle ground, high energy use with high CO_2 emissions and low electrification.

These studies using advanced techniques to gain robust insights into uncertainty on decarbonisation goals have majorly focused on the contributions of global system transitions but paid less attention to regional contributions. In addition, metrics were usually preselected based on existing knowledge and expert judgements rather than derived from the analysis itself. Such preselection of metrics, however, could be biased, ignoring other potentially influential metrics. This study thus uses a systematic approach to selecting potential metrics for the discovery of key drivers for global and regional decarbonisation.

3 Research approach

3.1 Methodological approach

As shown in Fig. 1, we first adopted the most recent scenario ensemble available that was used in the IPCC's special report on 1.5 °C (IPCC 2018). Proper metrics were then identified to represent decarbonisation strategies. These metrics were extracted from the ensemble at a regional level. Random forest analysis was used to choose a subset of regional metrics that are more likely to influence global climate goals for further analysis. This was followed by the application of logistic regression models to quantify the potential influence of those metrics on limiting the global temperature increase to either below 2.0 °C or below 1.5 °C. Visualisation of the transition scenarios was further used to explain the findings from the analysis. In parallel, the PRIM was used to reveal the trade-offs between key metrics to further confirm the findings from the logistic regression analysis. The adopted statistical methods were implemented with open-source, extensively documented Python packages, such as scikit-learn (Pedregosa et al. 2011) and EMA Workbench (Kwakkel 2017), and equivalent technical tools are available in other computational environments, such as R, to enable broad adoption and applications.

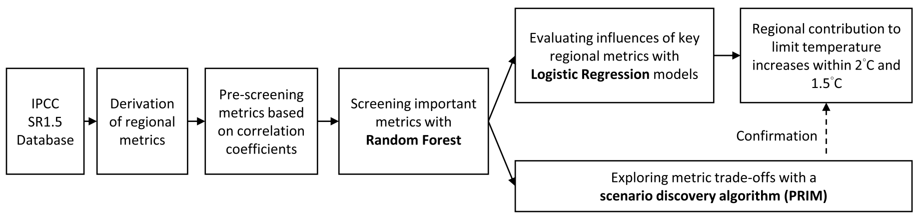


Fig. 1 Research flow

3.2 Scenario metrics

3.2.1 IPCC SR1.5 scenario database

This database, hosted by the International Institute for Applied Systems Analysis (Huppmann et al. 2019), includes 411 scenarios. Nineteen IAMs were used to generate those decarbonisation scenarios, which vary based on assumed climate ambition, and the wide range of socio-demographic and technological assumptions. This diversity of scenarios from these IAMs produces an ensemble that represents uncertainty across long-term global transitions for five aggregated regions as listed in Table 1. For simplification, scenarios are classified according to their expected global temperature increases by 2100 into three categories, including “Above 2 °C”, “2 °C” (covering the original “Higher 2.’0 °C” and “Lower 2.0 °C” as defined in the database), and “1.5 °C” (covering original “1.5 °C return with high overshoot”, “1.5 °C return with low overshoot,” and “Below 1.5 °C” as defined in the database). Those more detailed categories listed in the database are then not reflected in the following analysis.

The number of scenario attributes reported across the ensemble varies considerably, especially for sectoral attributes and alternative fuels (e.g. hydrogen). For this analysis, only those scenarios with regional outputs for most of the reported attributes were kept for further analysis, resulting in 236 scenarios out of the original set of 411 scenarios. The contribution of individual IAMs to the final set of scenarios can be found in the Appendix (Fig. A).

3.2.2 Regional metrics

In this study, 2010 is treated as the base year since attributes are more likely to be reported for this year than for more recent historical years. 2050 is used as the target year given its policy relevance in the UNFCCC negotiation process (to produce long-term low emission development strategies), and the conclusions of the IPCC special report that it is essential to reach net zero emissions around 2050 to avoid severe climate impacts (IPCC 2018).

To understand the contribution of regional decarbonisation measures to limit global temperature increases, a wide range of metrics were first identified, as listed in Appendix B. These cover three metric types (Table 2). The “structure” type includes metrics that show the technology shares in energy mixes in different regions, ranging from 0 to 1.

Table 1 Regions covered in the 1.5 °C database

Region	Abbreviation	Definition
ASIA	A	All Asian countries except those in OECD*
LAM	L	Latin America and the Caribbean
MAF	M	Middle East and Africa
OCED + EU	O	OECD 1990 countries as well as EU members and candidate countries
REF	R	Countries from the Reforming Economies of the Former Soviet Union

*OECD stands for the Organisation for Economic Co-operation and Development which is an intergovernmental organisation to stimulate economic growth and to improve the well-being of all

Table 2 Definition of regional metrics

Type	Metric	Definition
Structure	PriBio	Share of bioenergy in primary energy in 2050
	PriCoalCCS	Share of coal with CCS in primary energy in 2050
	PriGas	Share of gas without CCS in primary energy in 2050
	PriGasCCS	Share of gas with CCS in primary energy in 2050
	FinElc	Share of electricity in final energy consumption in 2050
	ElcBio	Share of bioenergy power without CCS in electricity generation in 2050
	ElcBioCCS	Share of bioenergy power with CCS in electricity generation in 2050
	ElcCoalCCS	Share of coal power with CCS in electricity generation in 2050
	ElcGas	Share of gas power in electricity generation in 2050
	ElcGasCCS	Share of gas power with CCS in electricity generation in 2050
	ElcGasCF	Average capacity factor of gas power plants (including plants with and without CCS) in 2050
	ElcNuc	Share of nuclear power in electricity generation in 2050
	ElcSol	Share of solar power in electricity generation in 2050
	ElcWnd	Share of wind power in electricity generation in 2050
Ratio	EngInt.Ratio	Ratio between energy intensities (Final Energy/GDP) in 2050 and 2010 (energy efficiency improvement) to show decarbonisation ambitiousness
Annual growth rate	-	Decarbonisation pace, such as annual growth rate of share of biomass in primary energy in the period between 2010 and 2050 (please refer to Appendix C for considered metrics)

*Only metrics mentioned in this study are listed here. For the complete list of metrics, please refer to Appendix C. CCS stands for carbon capture and sequestration technology

The "ratio" type represents the ratio of energy intensity (final energy consumption/GDP) in 2050 compared to that in 2010, indicating energy efficiency improvements across the time period. Finally, the "annual growth rate" type represents the pace of regional system changes, considering the technology shares in energy mixes in 2010 and 2050. Initially, 61 global metrics were derived, leading to 305 regional metrics. However, some attributes, such as hydrogen usage and sectoral energy shares, are rarely reported. As a result, only 26 metrics which had a more complete set of reported attributes were taken into account, reducing the total number of regional metrics to 130. For conciseness, only metrics mentioned in this study are listed in Table 2. The full list of these metrics can be found in Appendix C.

3.3 Statistical analysis of the metrics

3.3.1 Descriptive statistics

Scenarios are first divided into three temperature increase categories (i.e. above 2 °C, within 2 °C, and within 1.5 °C). The distribution of regional metrics across five regions is then shown in box and whisker plots for each of the scenario categories. As a result, the general trends of regional metrics over strengthening climate goals can be visualised for interpretation. These plots are then used to verify the findings from the following more advanced approaches.

3.3.2 Feature selection with correlation analysis and random forest (RF)

First of all, to reduce the number of regional metrics (130 in total) in the following analysis, Pearson correlation coefficients between each pair of regional metrics are calculated to screen out some of the regional metrics as metrics might have strong correlations with other metrics across regions. Only those without high correlations with other metrics and the first one from each group of highly correlated metrics are taken into account in the RF analysis for the feature importance evaluation. A correlation coefficient larger than 0.85 is regarded as highly correlated since the corresponding two metrics are more likely to increase or decrease together in uncertain scenarios. The correlation matrix of considered regional metrics can be found in Appendix D.

Then, the RF algorithm (Murphy 2012) was applied to further reduce the number of regional metrics to focus on those that are more likely to be influential in meeting climate goals. For a classification problem, first, this algorithm builds multiple decision trees to predict the class of a target variable (temperature increase of scenarios in this study) by reducing the variance (Gini index, explained below) of individual decision trees. Scenarios are used as training data, with regional metrics as predictors. Second, in the sequence of constructing individual decision trees, the algorithm not only randomly selects subsets of the input dataset (scenarios in this study) but also randomly chooses subsets of input variables (metrics in this study). Therefore, this algorithm can evaluate the importance of individual variables to a classification problem and has thus been used widely in many fields, including energy, construction, and ecology, for feature selection to improve prediction accuracy (Huo et al. 2021; Zhao et al. 2022; Kim et al. 2022). The procedure taken to determine three key RF parameters, including the number of trees, the maximum depth of a tree, and the maximum fraction of training samples, is explained in Appendix E.

The Gini index is adopted to evaluate the performance of the constructed decision trees to classify uncertain scenarios into three temperature increase categories (i.e. above 2 °C, 2 °C, and 1.5 °C). The Gini index is defined as follows:

$$\sum_{c=1}^{C} \pi_c (1 - \pi_c) \tag{1}$$

where c is a category, which is the temperature target in this study; C is the total number of categories to be predicted, which is three (i.e. above 2 °C, 2 °C, and 1.5 °C); π_c is the probability of an entry (scenario) in category c being correctly classified into category c; $(1 - \pi_c)$ is the misclassification probability for category c. Hence, a higher Gini index indicates a decision tree having a better classification performance. In other words, a subgroup of scenarios after divisions at input variables (metrics) is more consistent in categories.

Finally, the importance of a metric can then be evaluated by the average reduction in the Gini index by splits over the given metric, averaged over all trees. A high importance value for a metric means a change to the metric is more likely to divide scenarios into subgroups of scenarios with more consistent temperature increase categories. After ranking by feature importance, only metrics with high importance values, summing up to not higher than 0.8, were regarded as important for the following analysis.

Nonetheless, the importance can only be used to rank the potential influences of metrics on global temperature increases but cannot explicitly reveal the contribution of metrics to reach specific temperature targets (i.e. 2 °C and 1.5 °C), as the relationship between importance values is not linear. This algorithm, hence, is majorly applied to select key metrics for further consideration.

3.3.3 Logistic regression model

Since the scenarios are created by multiple IAMs with distinctive assumptions and modelling mechanisms, the temperature increases in 2050 could vary considerably, even with similar input assumptions. There is no clear cut between scenarios in different temperature increase categories, consequently. A probabilistic classification method that is not too sensitive to the overlaps between climate categories is thus desirable to evaluate the influences of regional metrics. The logistic regression model (James et al. 2021) is used to determine the influences of those metrics identified with high importance by the RF algorithm. The model is one of the most widely-adopted probabilistic classification algorithms and has a relatively simple formulation as a nonlinear function with the mean parameters in terms of a linear combination of input assumptions (Murphy 2012). Its straightforward formulation makes the interpretation easier.

Logistic regression models use a logistic function, as defined by Eq. (2), for binary classification to predict the probability of a scenario belonging to a specific category, given values of input metrics in the scenario:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \tag{2}$$

where X is the vector of chosen metrics, such as those listed in Table 2; β_0 is the intercept and β_1 is the vector of regression coefficients for input metrics. $p(X)$ is the probability of a considered scenario falling into a category of concern (i.e. climate target in this study), with a value ranging from 0 to 1. The continuous value explicitly shows the prediction uncertainty, which makes the logistic function an ideal method for classifying the climate target category of highly uncertain transition scenarios.

Two binary classification problems are considered, one for the probability of limiting the temperature increase to below 2 °C and the other for the probability of limiting the temperature increase to below 1.5 °C. Scenarios in the categories of 2 °C and 1.5 °C are regarded as reaching the below 2 °C target.

First, the temperature increases of the scenarios are regarded as target variables in two cases respectively, using the metrics identified by the RF as predictors. Second, a maximum likelihood approach (James et al. 2021) is then used to estimate the regression coefficients for the two models. To avoid potential biases from the random training dataset (80% of the highly uncertain scenarios), the model training procedure is repeated 10 times for each of the two climate targets (i.e. limiting temperature increase to 2 °C and 1.5 °C) to determine the average coefficient of individual metrics across ten built models. A positive average coefficient means an increase in the corresponding metric can increase the possibility of reaching the temperature target under consideration. A negative average coefficient suggests the opposite. A higher coefficient indicates that the change of that metric is more likely to have stronger impacts on reaching specific climate targets.

Due to the non-linear relationship between input metrics and the probability function, as shown in Eq. (2), the rate of change in the probability per unit change in the metrics depends on the current values of input metrics.

In other words, the estimated coefficients cannot show the linear influences of the corresponding metrics on the probability of direct comparisons (James et al. 2021). Hence, this study only uses the coefficients to rank the importance of individual metrics to attain insights into the role of regional measures in global decarbonisation.

3.3.4 Patient rule induction method (PRIM)

To highlight trade-offs across the most influential scenario metrics within each temperature increase category, we then use the PRIM with those metrics deemed important by the RF. Since PRIM does not impose requirements on data distributions, the PRIM is well-suited for the unstructured ensemble of decarbonisation scenarios used in our analysis (Bryant and Lempert 2010). This method not only provides additional policy-support information but also further confirms the crucial roles of key metrics identified by the logistic regression analysis before.

The PRIM uses a hill-climbing optimization algorithm to identify combinations and ranges of predictor variables associated with a certain value or region of an output variable, by sequentially "peeling" ranges of each predictor variable (Friedman and Fisher 1999). This yields a "box" described by k restrictions on the value of predictor variables X_i (Eq. 3), where i is the metric under consideration. These restrictions are defined as minimum and maximum value bounds in the case of continuous predictor variables:

$$\bigcap_{i=1}^{k} min_i \le X_i \le max_i \tag{3}$$

To identify combinations and ranges of regional metrics associated with above 2 °C, 2 °C, and 1.5 °C scenarios, we apply the PRIM separately for each type of metric and for each temperature increase category. We consider the metrics as predictor variables and the membership of a scenario in a specified temperature category as the output variable. Three standard statistical learning performance measures are used to assess resulting "boxes": recall, precision, and F1 score (van Rijsbergen 1979). The recall is the fraction of all scenarios within a temperature increase category that shares the combinations of metrics described by the box. Precision is the fraction of scenarios described by the box that fall within the specified temperature category. The F1 score combines these two metrics using their harmonic mean.

Our analysis is conducted with a constrained application of the PRIM (Kwakkel 2017). For simplicity, we choose boxes describing three restricted metrics for each type of metrics and each temperature increase category. We then select boxes that have a score above 0.9 on at least one of the recall and precision performance measures, combined with a F1 score above 0.7 to ensure balanced performance between recall and precision. The identified three metrics forming the boxes are thus more likely to be crucial for scenarios in the corresponding temperature category.

4 Results and discussion

4.1 Descriptive statistics

4.1.1 Overview of regional decarbonisation

The distributions of regional metrics across all scenarios by temperature increase (i.e. above 2 °C, 2 °C, and 1.5 °C) are shown in Figs. 2, 3, and 4. The relationship between metric distribution and temperature increase reveals the general elements of the transitions for limiting temperature increases.

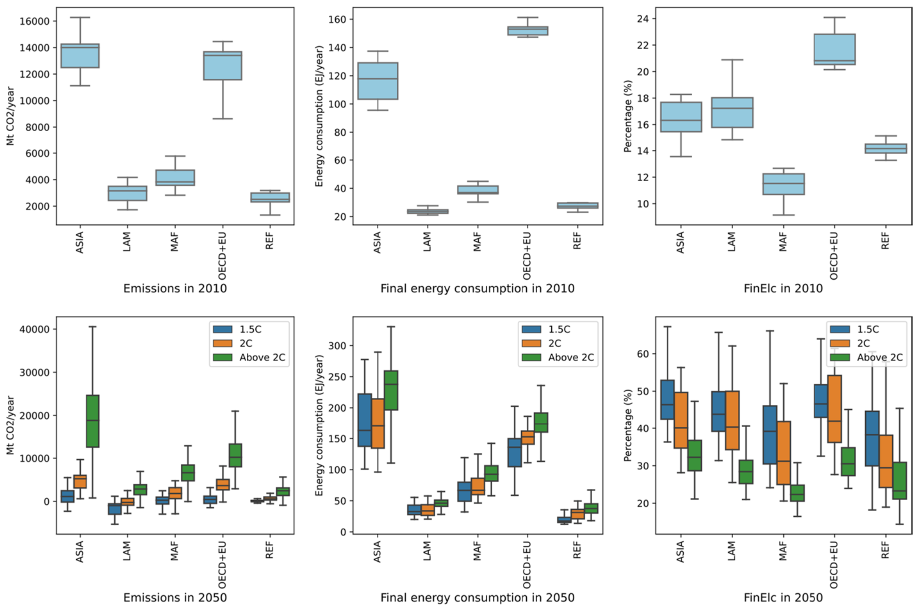


Fig. 2 Distribution of CO_2 emissions, final energy consumption, and share of electricity in final energy consumption in 2010 and 2050 by temperature increase. Metrics in 2010 represent uncertain assumptions on initial status in models and are thus not illustrated by temperature increase. (FinElc, share of electricity in final energy consumption)

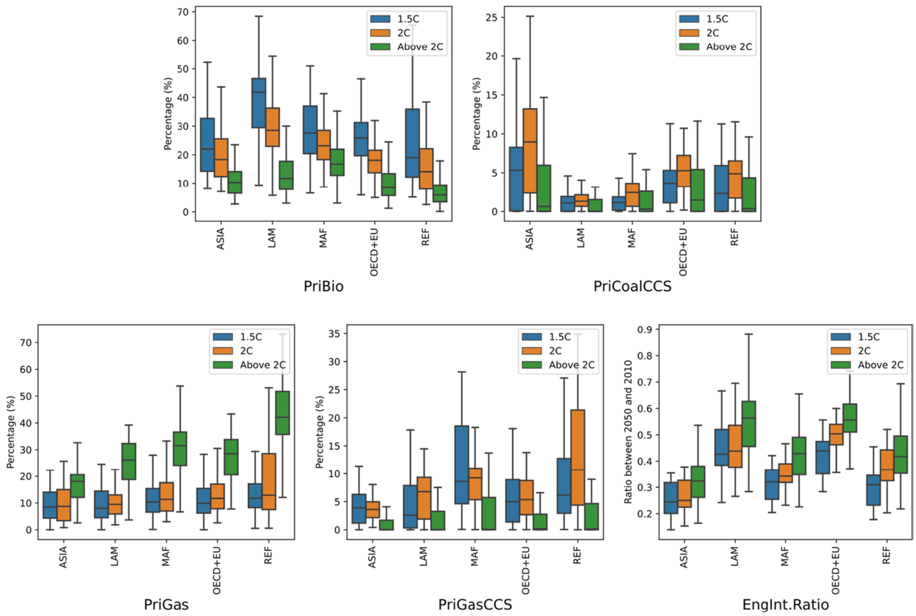


Fig. 3 Distribution of primary energy mix in 2050 and energy intensity ratio between 2050 and 2010 by temperature increase (PriBio, share of biomass; PriCoalCCS, share of coal usage with CCS; PriGas, share of gas; PriGasCCS, share of gas usage with CCS; EngInt.Ratio, ratio of energy intensity between 2050 and 2010)

The efforts to transform regional systems to meet the three climate goals vary considerably across regions, as shown in Fig. 2. This is driven to a large extent by the differences in the current systems and their final state in 2050. Among the five regions, OECD + EU is the least carbon-intensive (about 0.08 Mt CO2/PJ) region in 2010 due to its higher electrification level (around 22%), which is about 20% lower than most of the rest regions (i.e. ASIA, LAM, and MAF) and about 6% lower than REF in terms of carbon emissions per unit final energy consumption. By 2050, final energy consumption increases considerably across five regions due to economic growth and increasing populations, compared to the base year. ASIA and OECD + EU remain the two regions that consume the most final energy; ASIA increases by a factor of two while OECD + EU consumes around the same amount of final energy in 2050 compared to 2010. All five regions dramatically ramp up their electrification levels to around 40% of the final energy supply in 2050.

4.1.2 Regional structural transformation

The contribution of detailed regional metrics to different temperature levels is illustrated in Figs. 3 and 4. The results for growth rate metrics are in Appendix F as the findings are similar to those for structure metrics. Across decarbonisation measures, it is clear that some metrics play a more important role in meeting temperature goals based on energy mixes. For instance, to reach the 1.5 °C target, the share of primary bioenergy supply (PriBio) increases to more than 20% in most regions and up to 40% in LAM. Meanwhile, more than 20% of electricity generation should be from wind power (ElcWnd). On the demand side,

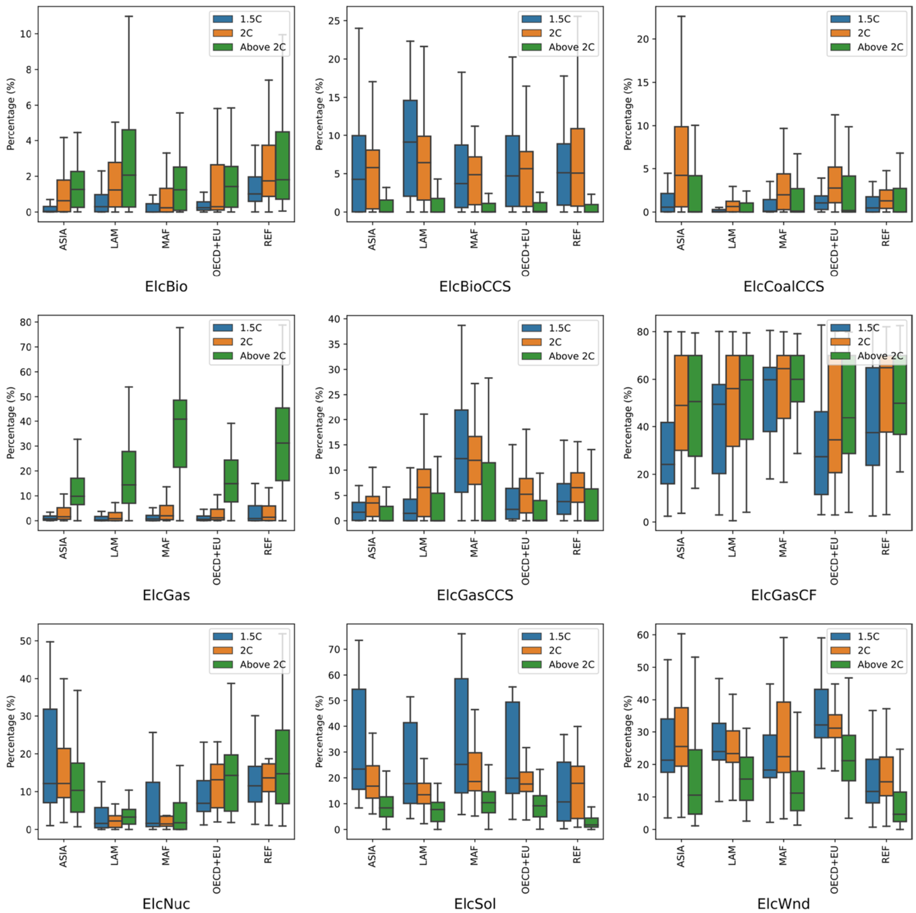


Fig. 4 Distribution of power generation technologies in 2050 by temperature increase (ElcBio, share of biomass power; ElcBioCCS, share of biomass power with CCS; ElcCoalCCS, share of coal power with CCS; ElcGas, share of gas power; ElcGasCCS, share of gas power with CCS; ElcGasCF, average capacity factor of gas power plants; ElcNuc, share of nuclear power; ElcSol, share of solar power; ElcWnd, share of wind power)

energy efficiency is improved considerably as climate goals strengthen, reducing energy intensity (EngInt.Ratio) by 60% in LAM and OECD + EU and by 70% in ASIA, MAF, and REF, compared with 2010 levels. Understandably, metrics associated with fossil fuels are very constrained in both 1.5 °C and 2 °C scenarios. For example, the share of gas supply (PriGas) does not exceed 15% where carbon capture and storage (CCS) is not used. Even with CCS, gas usage for primary energy supply (PriGasCCS) and power generation (ElcGasCCS) remains at a low level (less than 10%) in most regions.

These metrics also reveal specific regional characteristics, with some mitigation actions deployed in a more heterogenous pattern than others. Those more heterogenous imply that certain decarbonisation measures may be more feasible in some regions than others, owing to higher resource availability, technology maturity, or even political preferences. For instance, in the 1.5 °C scenarios, bioenergy accounts for a much higher share of primary energy (more than 40%) in LAM than in other regions. Moreover, bioenergy can be used in power plants fitted with CCS (ElcBioCCS) to provide upwards of 10% of electricity generation in LAM, which is much higher than observed elsewhere.

In ASIA, nuclear power (ElcNuc) can contribute to more than 20% of electricity generation whereas other regions have lower shares. This is also the only region that uses coal with CCS (PriCoalCCS) to satisfy up to 10% of primary energy demand. Meanwhile, gas usage with CCS (PriGasCCS and ElcGasCCS) increases to a much higher level in MAF than other regions in the 1.5 °C scenarios. This diversity in regional action results from global climate targets determining where regional decarbonisation potentials can be utilised.

4.1.3 Difficulties in identifying influential metrics

The distribution plots often provide a clear distinction between the role of metrics in relation to different temperature goals. However, these box plots also overlap, obscuring the role of specific actions in limiting temperature rise. Significant overlaps between above 2 and 2 °C scenarios can be found in several metrics, including the share of coal consumption with CCS (PriCoalCCS), share of bioenergy power without CCS (ElcBio), share of coal power with CCS (ElcCoalCCS), capacity factor of gas power plants (ElcGasCF), and share of nuclear power (ElcNuc). This might be due to the larger carbon budgets for the 2 °C target allowing models to have more choices for the decarbonisation measures to pick, resulting in the wide variation of these metrics. With the more ambitious 1.5 °C goal, the difference in metric distributions between temperature targets is more distinctive, as seen for bioenergy usage and solar power (PriBio, ElcBio, and ElcSol). Nevertheless, due to the uncertainties and trade-offs between decarbonisation measures, overlaps between 2 and 1.5 °C scenarios across metrics are still evident.

In summary, the role of these metrics in limiting temperature increases is highly uncertain, with evident overlaps between temperature targets. This implies that strengthening certain decarbonisation measures might not be necessarily the most effective way to reach more ambitious climate goals due to other more influential alternative measures. However, these trade-offs between metrics cannot be easily discovered with descriptive statistics, as shown here. Regional heterogeneity further complicates the strategies required for the 1.5 °C target since a few measures might be more feasible or effective in specific regions. Similarly, the contribution of individual regional metrics to climate goals is not easily discerned with simple visualisation. Therefore, more systematic statistical approaches that can simultaneously consider system dynamics in the models and the trade-offs between metrics in uncertain scenarios are useful to reveal the decarbonisation patterns at the regional level, as discussed below.

4.2 Random forest analysis

The ranking of importance of regional metrics determined by the built RF model (with 96% classification accuracy) is shown in Fig. 5 for structure metrics. Characters in brackets indicate regions. Regions that are highly correlated for a specific metric are listed together for clarification. For example, ElcGas(A.M.O.R) indicates gas generation in ASIA, MAF, OECD + EU, and REF is highly correlated. More information on metric correlations can be found in Appendix D. Only those with high feature importance are listed in the figures and are taken into account in the logistic regression analysis. The selected metrics are more likely to be critical for meeting decarbonisation targets.

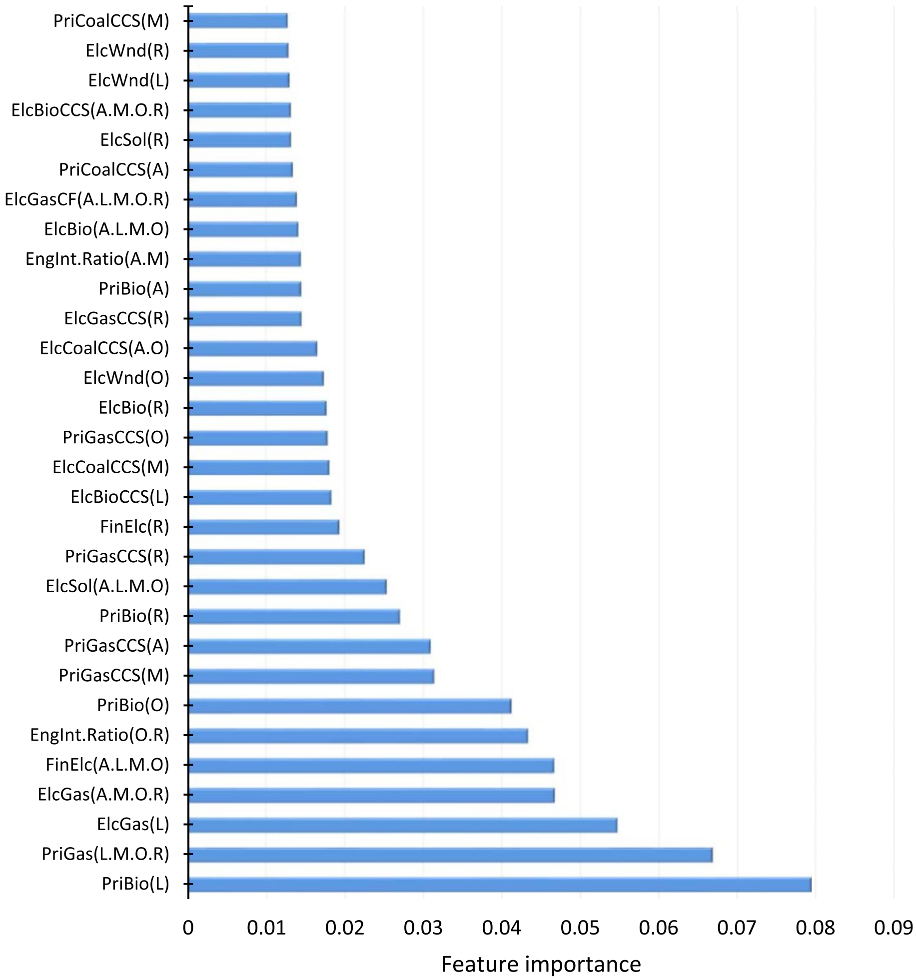


Fig. 5 Feature importance of structure metrics to global temperature increases (above 2 °C, 2 °C, or 1.5 °C). L, A, M, O, and R refer to Latin America and the Caribbean, Asia, the Middle East and Africa, OECD+EU, and the Reforming Economies of the Former Soviet Union (REF) respectively. (EngInt.Ratio, ratio of energy intensity between 2050 and 2010; FinElc, share of electricity in final energy consumption; PriBio, share of biomass use; PriGas, share of gas use; PriCoalCCS, share of coal use with CCS; PriGasCCS, share of gas use with CCS; ElcBio, share of biomass power; ElcGas, share of gas power; ElcSol, share of solar power; ElcWnd, share of wind power; ElcBioCCS, share of biomass power with CCS; ElcCoalCCS, share of coal power with CCS; ElcCoalCCS, share of coal power with CCS; ElcGasCF, average capacity factor of gas power plants)

The RF analysis reveals that unabated gas generation (ElcGas) and share of electricity (FinElc) across all regions and share of bioenergy in LAM (PriBio(L)) are the most likely to affect final temperature increases in 2050, supported by previous figures. This is because of two reasons. First, metrics with a strong positive or negative correlation with increasing temperature targets have been emphasised. For example, Fig. 5 reveals the high influences of PriBio, that clearly increases as climate goals strengthen (Fig. 3). The potential high influences of FinElc are also identified (Fig. 5) to reflect its strong negative correlation with temperature targets (Fig. 2). Second, metrics with evident differences under different temperature targets mean the deployment of the measures also has a clear impact on global decarbonisation. For instance, ElcGas is also recognised as an important metric (Fig. 5) due to its large fall from above 2 to 2 °C and below 1.5 °C scenarios, as shown in Fig. 4.

Regional heterogeneity in meeting climate goals has also been shown. For example, actions in LAM that are likely to have a high impact on global temperature increases include bioenergy supply (PriBio), gas power (ElcGas), and BECCS power (ElcBioCCS). In ASIA and MAF, gas supply with CCS (PriGasCCS) has stronger impacts.

The RF analysis only ranks the importance of metrics to global decarbonisation, without explicitly showing the level of contribution of these metrics to reach specific climate goals (2 °C or 1.5 °C). The logistic regression analysis is thus essential to gain further insights into regional transition strategies to reach specific global goals (see Section 4.2).

4.3 Logistic regression analysis

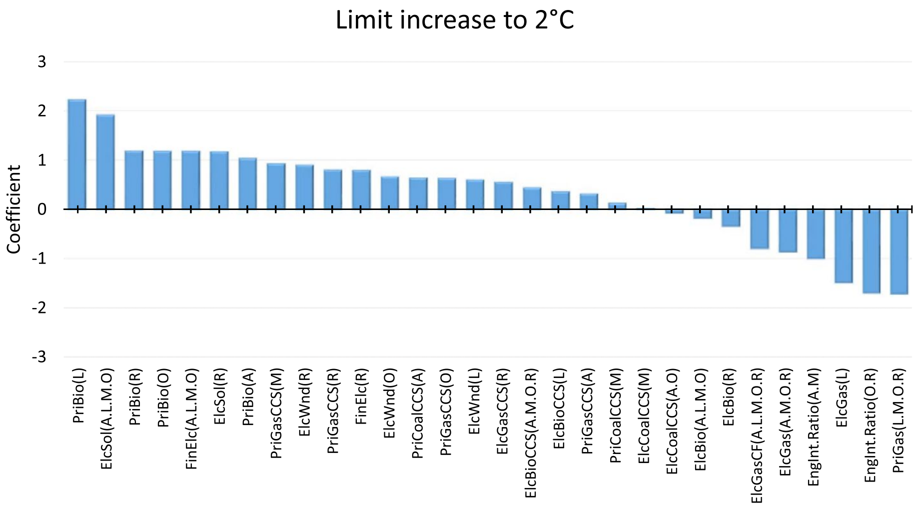
Logistic regression models were then applied to determine those regional metrics indicating actions most effective in reaching specific climate goals across the scenarios (Fig. 6). The potential influences of metrics were analysed for two climate goals (i.e. below 2 °C and below 1.5 °C). The average prediction accuracies of the trained models for the two goals are 85% and 87% respectively. Average regression coefficients of the built models reflect the findings from descriptive statistics (Figs. 2, 3, and 4) but also go further by quantifying and ranking the impacts of individual regional metrics.

The most striking difference between the two cases of temperature targets is coefficient distribution. Coefficients are higher in the case of reaching 2 °C, compared with those for 1.5 °C. This implies that an increase in any single decarbonisation measure has a lower contribution to reach the 1.5 °C target than it does to the 2 °C target, as much more carbon emissions should be reduced for 1.5 °C. For instance, the coefficient of bioenergy supply (PriBio) in LAM for 2 °C is around 2.2, compared to 1.4 for 1.5 °C. A unit increase of PriBio in LAM thus leads to a higher probability of reaching 2 °C than it does when the target is further tightened to 1.5 °C. Hence, multiple actions should be taken simultaneously to increase the probability of reaching the 1.5 °C target.

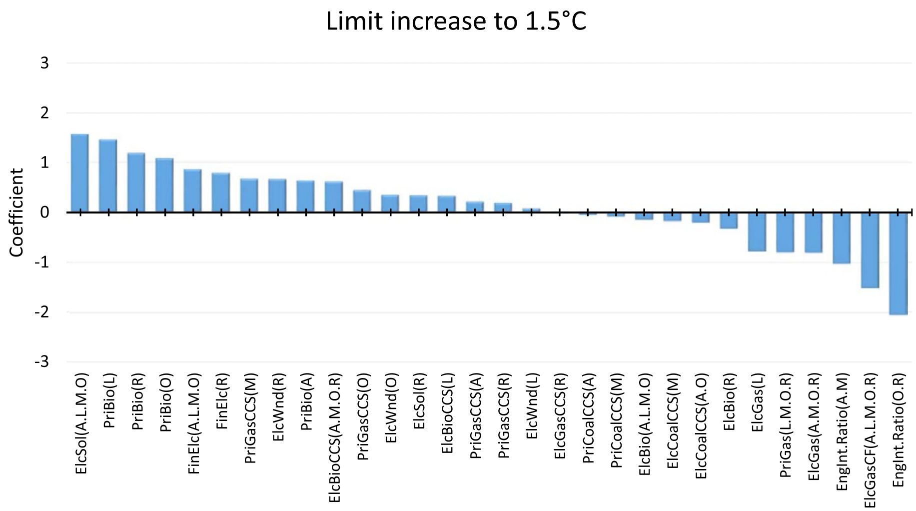
This analysis shows that specific metrics have a particularly strong contribution to climate goals across most regions. For example, electrification level (FinElc) and solar power generation (ElcSol) in ASIA, LAM, MAF, and OECD + EU have been found the most influential in limiting global temperature increases for both 2 °C and 1.5 °C. This is due to two reasons: (1) the higher levels of deployment of these measures relative to others and (2) the clear increase in the deployment of these measures over strengthening targets, as shown in Figs. 2, 3, and 4.

Even though bioenergy supply across three regions (LAM, REF, and OECD + EU) all show strong impacts, this metric has higher regional heterogeneity due to regional conditions, with LAM being the most influential region. With high bioenergy potentials in LAM, bioenergy is used not only for electricity generation, but also to replace fossil fuels in the end-use sectors. This can be shown by its low correlation coefficient with ElcBioCCS (0.5) and high negative correlation coefficients with PriGas (− 0.7) in LAM (Appendix D).

Conversely, energy intensity (EngInt.Ratio) and unabated gas usage (including ElcGas and PriGas) see high negative coefficients, whereby their reduction sees temperature targets more likely to be met. These two metrics also have clear regional heterogeneity in respect of their contribution. The metric EngInt.Ratio in OECD + EU and ASIA has been



(a)



(b)

Fig. 6 Average regression coefficients represent the potential influences of corresponding metrics on reaching **a** 2 °C and **b** 1.5 °C. L, A, M, O, and R refer to Latin America and the Caribbean, Asia, the Middle East and Africa, OECD + EU, and the Reforming Economies of the Former Soviet Union (REF) respectively. (EngInt.Ratio, ratio of energy intensity between 2050 and 2010; FinElc, share of electricity in final energy consumption; PriBio, share of biomass use; PriGas, share of gas use; PriCoalCCS, share of coal use with CCS; PriGasCCS, share of gas use with CCS; ElcBio, share of biomass power; ElcGas, share of gas power; ElcSol, share of solar power; ElcWnd, share of wind power; ElcBioCCS, share of biomass power with CCS; ElcCoalCCS, share of coal power with CCS; ElcCoalCCS, share of coal power with CCS; ElcGasCF, average capacity factor of gas power plants)

shown to be more influential than most supply-side measures given its particularly high negative coefficients for both temperature targets (Fig. 6). This reflects the clear decrease in energy intensity (i.e. final energy per unit of GDP) over strengthening temperature targets in Fig. 3 and highlights the importance of energy demand reduction to reduce pressure on energy supply and associated carbon emissions.

The drops in gas supply (PriGas) and electricity generation from gas (ElcGas) from scenarios above 2 to below 2 °C are particularly evident, ranging from 10 to 30%, as shown in Fig. 4. Without CCS, gas consumption in LAM, MAF, OECD + EU, and REF (PriGas(L.M.O.R)) decreases dramatically for ambitious temperature targets. However, the contribution of PriGas in ASIA is not recognised as significant since the gap between scenarios above 2 °C and more stringent targets is smaller in ASIA than observed for other regions (Fig. 3). To further limit the temperature increase to below 1.5 °C, it is particularly crucial to reduce the operational capacity of existing gas power plants (ElcGasCF) to around 30%, indicated by the high coefficient in Fig. 6 and a huge drop in Fig. 4. This implies numerous gas power plants will become stranded assets.

ElcBio in all regions has a negative coefficient, indicating that bioenergy use in power generation without CCS should decrease under stringent targets, with REF having a particularly high coefficient. Instead, bioenergy use should increase in combination with CCS for negative emissions, as suggested by the positive coefficients of ElcBioCCS in all regions. This is particularly the case for 1.5 °C across ASIA, MAF, OECD + EU, and REF (ElcBioCCS(A.M.O.R)), with the 10th largest coefficient in the case for 1.5 °C but the 17th for 2 °C (Fig. 6).

The analysis also highlights that strengthening some selected measures might not be always effective or desirable to meet temperature targets due to interlinkages. There are two specific cases observed. Firstly, some metrics, such as ElcWnd(L) and ElcGasCCS(R), might be only effective to reach the 2 °C target (suggested by their large coefficients) but become ineffective to further reduce temperature increase from 2 to 1.5 °C. This is shown by the extremely small coefficients across specific metrics, and the large overlaps between metric distributions in the earlier box plots in Fig. 4. For ElcWnd, this might be because wind potentials are fully exploited even in the 2 °C scenarios driven by the cost-effectiveness and maturity of the technology. Further wind power deployment for 1.5 °C might considerably increase marginal costs for required system balancing measures or might be limited to feasible development sites. Secondly, some actions need to be increased by 2 °C (suggested by positive coefficients for metrics) but may need to be reduced by 1.5 °C (suggested by negative coefficients). This is evident for metrics associated with fossil fuel consumption with CCS. For 2 °C, residual emissions from CCS can still be incurred but not under the more limited carbon budgets for 1.5 °C. This situation can be found in distributions of these metrics in ASIA, LAM, MAF, and REF, including PriCoalCCS(A), PriGasCCS(L), PriCoalCCS(M), and PriGasCCS(R). The corresponding power generation metrics (i.e. ElcGasCCS) see similar trends since these two types of metrics are highly correlated (with correlation coefficients around 0.8). Compared to other metrics, these are less influential based on their lower metric values.

4.4 PRIM analysis

As the logistic regression analysis already hinted at interlinkages among the metrics, the PRIM results (Fig. 7) can further reveal the combinations of and trade-offs between the most important regional metrics for each temperature target. The PRIM results are aligned with the findings from the logistic regression analysis and thus confirm the robustness of the findings. Here, we provide the analysis for one representative region of a group of highly correlated regions for simplification, as metrics in those regions have very similar trends.

For scenarios in the temperature category above 2 °C ($n = 108$), the PRIM identifies a box with high precision for structure metrics (Fig. 7, top subplot; $P = 0.96$), so that 96% of the scenarios that combine these restricted ranges of values fail to achieve the 2 °C target. This enables insights into specific combinations of regional scenarios that would "lock in" undesirable global outcomes. Scenarios with relatively limited shares of solar generation in ASIA (ElcSol(A.L.M.O) < 21%), combined with limited shares of primary energy in LAM for biomass (PriBio(L) < 20%), but higher for gas in LAM (PriGas (L.M.O.R) > 12%), nearly always lead to a temperature increase above 2 °C. This combination is consistent with regression coefficients (Fig. 6a) in which solar generation share in most regions and primary biomass share in LAM have the highest positive values towards the target above 2 °C, while the primary gas share in most regions has the second-highest negative value. The PRIM results additionally emphasise the parallel regional transformations required: a high share of solar generation globally alone would not prevent outcomes above 2 °C if combined with a limited increase in biomass share in LAM and a limited decrease in global gas share in primary energy.

For the scenarios below 2 °C ($n = 128$), the PRIM finds a box with lower precision for structure metrics (Fig. 7, middle subplot; $P = 0.78$) but captures nearly all of these scenarios ($R = 0.98$ or 126 of 128 scenarios). This box thus highlights combined regional prerequisites for achieving 2 °C in the context of the scenario ensemble. Nearly all scenarios below 2 °C combine relatively low gas generation in ASIA (ElcGas(A.M.O.R) < 23%), relatively high energy intensity improvement in OECD (EngIntRatio (O.R) < 61%), and relatively low share of gas in primary energy for LAM (PriGas(L.M.O.R) < 27%). Compared with the box with scenarios above 2 °C, the results indicate that a certain level of unabated global gas use could potentially be compensated by sufficiently low energy intensity in OECD and global gas generation.

Fig. 7 Combinations of ranges of metrics associated with each temperature increase category, identified using the Patient Rule Induction Method (PRIM). The normalised uncertainty range shows the full range of metrics across the entire ensemble of decarbonisation scenarios. Grey lines denote constrained ranges of metrics associated with each temperature increase category. (R, recall performance measure; P, precision performance measure; F1, F1 score.) The definition of metrics can be found in Table 2

a

	Above 2°C scenarios	
Metric	Min. value (%)	Max. value (%)
ElcSol(A.L.M.O)	0.052	20.8
PriBio(L)	3	20.2
PriGas(L.M.O.R)	11.8	39

b

	Below 2°C scenarios	
Metric	Min. value (%)	Max. value (%)
ElcGas(A.M.O.R)	0	22.9
EngIntRatio(O.R)	28	60.9
PriGas(L.M.O.R)	0.0018	27

c

	Below 1.5°C scenarios	
Metric	Min. value (%)	Max. value (%)
ElcGas(A.M.O.R)	0	5.2
EngInt.Ratio(O.R)	28	56.4
FinElc(A.L.M.O)	36.3	71

As the number of scenarios below 1.5 °C is limited ($n=53$) compared to the number of regional metrics, the precision of the box identified for structure metrics (Fig. 7, bottom subplot; $P=0.56$) is lower than for the other temperature categories. Nonetheless, the box captures nearly all 1.5 °C scenarios ($R=0.98$, or 52 of 53 scenarios). As such, the combination of minimal gas generation in ASIA (ElcGas (A.M.O.R) < 5.2%), even stronger energy intensity improvement in OECD (EngInt.Ratio (O.R) < 56%), and relatively high electrification in ASIA (FinElc(A.L.M.O) > 36%) are additionally required to achieve 1.5 °C in the context of this scenario ensemble. Also, the additional effort required for 1.5 °C in relation to 2 °C is apparent in smaller feasible ranges for energy intensity in OECD and especially electricity generation from gas, compared to the 2 °C PRIM box.

5 Conclusions

To benefit from large existing IPCC SR1.5 ensembles of global climate mitigation scenarios to draw regional insights, a novel approach has been developed with three machine learning techniques of random forest algorithm, logistic regression model, and the patient rule induction method, to reveal the links between decarbonisation measures in five global regions and reaching two climate goals: below 2 °C and below 1.5 °C). Influential regional decarbonisation strategies, identified through random forest algorithm and logistic regression model, have also been verified and explained by descriptive statistics on the proposed metrics for three temperature increase categories, including above 2 °C, 2 °C, and below 1.5 °C. The patient rule induction method has further shown the trade-offs between key regional metrics and revealed combinations of regional actions that are associated with the temperature targets considered. More powerful than just descriptive statistics, the developed approach can emphasise more influential regional measures for ambitious temperature targets, along with the strategies of those measures in specific regions in terms of increasing or decreasing deployment. Moreover, the developed approach can also be applied to investigate high-dimensional uncertain system transition scenarios in other areas at various scales to gain robust insights to inform policymakings. It is worth noting that the regional heterogeneity among metrics is revealed by correlation analysis on regional metrics and the importance of individual regional metrics identified by the RF algorithm and logistic regression model.

Several insights into regional decarbonisation strategies have been found from the analysis:

1. Specific global actions should be prioritised across all regions to effectively limit temperature increases, recognised as the most influential measures by the analysis. These decarbonisation measures are the deployment of solar power and electrification. Particularly, governments should propose effective policies to decommission unabated gas plants prematurely, which is one of the crucial measures for the 1.5 °C target.
2. Evidently, energy efficiency improvement can play an equally or even more important role than most other metrics in reaching 2 °C and 1.5 °C targets, which is especially crucial in OECD and countries of the Former Soviet Union for the 2 °C target and even across all regions for the 1.5 °C target.
3. However, some actions are specific to a region and not others, highlighting regional dependencies in meeting global temperature targets. Such insights are particularly relevant as they raise questions about dependency on regions, and potentially questions of

equity in meeting global climate objectives. For instance, primary bioenergy supply in LAM is particularly influential (and to a lesser extent, bioenergy in OECD and countries of the Former Soviet Union), implying that without these regional actions, the said targets would be harder to meet. This may particularly relate to the role of bioenergy used with CCS as an option for CDR, offsetting emissions in other regions. The strong scale-up of wind power in OECD and countries of the Former Soviet Union could also be effective for climate targets, albeit with lower influences.

4. Some region-specific actions, on the other hand, need to be limited to increase the chance to reach ambitious climate goals, including reducing gas use in power generation in Latin America, or bioenergy without CCS for power in countries of the Former Soviet Union.
5. Finally, influential combinations of regional actions are essential to effectively limit global temperature increases, instead of only focusing on individual regional actions. Limiting temperature increase to below 2 °C requires a low share of unabated gas usage (for primary energy consumption and power generation) in most regions and a significant energy intensity improvement in OECD. Relatively high electrification and an even lower share of unabated gas power in global energy systems are further needed to reach the 1.5 °C target.

The findings from the use of these methods reveal how the changes to actions (represented by metrics) in a scenario ensemble can raise the probability of reaching 2 °C and 1.5 °C targets, and whether actions in specific regions have a stronger bearing on outcomes. This information can help policymakers decide how to focus on more effective measures in specific regions or globally.

Nonetheless, this study only considers three temperature increase categories without reflecting on more detailed decarbonisation categories, including climate goals with and without overshoot. Those more detailed categories can also be considered in future studies, along with decarbonisation strategies by 2100. Second, this study uses a relatively small sample size of scenarios from the IPCC SR1.5 scenario database due to the missing reported sectoral attributes and alternative fuels in scenarios. This also limits the number of metrics that can be considered. Consequently, some extreme cases in the long-term transitions might not be well represented, and decarbonisation measures have not been fully considered in this study, such as land-based CDR and hydrogen consumption. In the future, a larger uncertain scenario ensemble, especially the IPCC AR6 ensemble, with more completely reported attributes, covering advanced decarbonisation measures, model characteristics, and sociodemographic factors, can be analysed with the proposed approach to reveal more robust and comprehensive insights. Furthermore, other probabilistic classification algorithms can also be applied to investigate metric importance to verify the findings of this study.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10584-023-03529-w.

Author contribution All authors (PHL, SP, IK, MJR, and ET) contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by PHL and MJR. The first draft of the manuscript was written by PHL and MJR. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by University of Geneva This work received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 821124

(NAVIGATE). PHL is also partially supported by EnergyREV Plus (EP/S031898/1) and UKERC 4 (EP/S029575/1).

Data availability The global transition scenario dataset analysed in this study is available on the website of IAMC 1.5 °C Scenario Explorer hosted by IIASA. https://data.ene.iiasa.ac.at/iamc-1.5c-explorer.

Declarations

Ethics approval and consent to participate Not applicable because this study does not involve humans and/or animals. Consent to participate is not applicable since this study does not report the results of studies involving humans and/or animals.

Consent for publication Not applicable because this study does not report the results of studies involving humans and/or animals.

Competing interests The authors declare no competing interests.

References

Abernethy S, Jackson RB (2022) Global temperature goals should determine the time horizons for greenhouse gas emission metrics. Environ Res Lett 17:024019. https://doi.org/10.1088/1748-9326/ac4940

Azevedo I, Bataille C, Bistline J et al (2021) Net-zero emissions energy systems: what we know and do not know. Energy Clim Chang 2:100049. https://doi.org/10.1016/j.egycc.2021.100049

Breiman L, Friedman JH, Olshen RA, Stone CJ (2017) Classification and regression trees. Classification and Regression Trees 1–358. https://doi.org/10.1201/9781315139470/CLASSIFICATION-REGRESSION-TREES-LEO-BREIMAN-JEROME-FRIEDMAN-RICHARD-OLSHEN-CHARLES-STONE

Brutschin E, Pianta S, Tavoni M, Riahi K, Bosetti V, Marangoni G, Van Ruijven BJ (2021) A multidimensional feasibility evaluation of low-carbon scenarios. Environ Res Lett 16(6):064069. https://doi.org/10.1088/1748-9326/abf0ce

Bryant BP, Lempert RJ (2010) Thinking inside the box: a participatory, computer-assisted approach to scenario discovery. Technol Forecast Soc Chang 77:34–49. https://doi.org/10.1016/J.TECHFORE.2009.08.002

DeAngelo J, Azevedo I, Bistline J, Clarke L, Luderer G, Byers E, Davis SJ (2021) Energy systems in scenarios at net-zero CO2 emissions. Nat Commun 12(1):6096. https://doi.org/10.1038/s41467-021-26356-y

Diniz Oliveira T, Brunelle T, Guenet B, Ciais P, Leblanc F, Guivarch C (2021) A mixed-effect model approach for assessing land-based mitigation in integrated assessment models: A regional perspective. Glob Change Biol 27:4671–4685. https://doi.org/10.1111/gcb.15738

Friedman JH, Fisher NI (1999) Bump hunting in high-dimensional data. Stat Comput 9(2):123–143. https://doi.org/10.1023/A:1008894516817

Fujimori S, Rogelj J, Krey V, Riahi K (2019) A new generation of emissions scenarios should cover blind spots in the carbon budget space. Nat Clim Chang 9:798–800

Gambhir A, Rogelj J, Luderer G et al (2019) Energy system changes in 1.5 °C, well below 2 °C and 2 °C scenarios. Energ Strat Rev 23:69–80. https://doi.org/10.1016/j.esr.2018.12.006

Gerst MD, Wang P, Borsuk ME (2013) Discovering plausible energy and economic futures under global change using multidimensional scenario discovery. Environ Model Softw 44:76–86. https://doi.org/10.1016/j.envsoft.2012.09.001

Guivarch C, Lempert R, Trutnevyte E (2017) Scenario techniques for energy and environmental research: an overview of recent developments to broaden the capacity to deal with complexity and uncertainty. Environ Model Softw 97:201–210. https://doi.org/10.1016/J.ENVSOFT.2017.07.017

Guivarch C, Rozenberg J, Schweizer V (2016) The diversity of socio-economic pathways and CO2 emissions scenarios: insights from the investigation of a scenarios database. Environ Model Softw 80:336–353. https://doi.org/10.1016/j.envsoft.2016.03.006

Huo W, Li W, Zhang Z, Sun C, Zhou F, Gong G (2021) Performance prediction of proton-exchange membrane fuel cell based on convolutional neural network and random forest feature selection. Energy Convers Manag 243:114367. https://doi.org/10.1016/j.enconman.2021.114367

Huppmann D, Kriegler E, Krey V, Riahi K, Rogelj J, Calvin K, Humpenoeder F, Popp A, Rose S, Weyant J, Bauer N, Bertram C, Bosetti V, Doelman J, Drouet L, Emmerling J, Frank S, Fujimori S, Gernaat D, Grubler A, Guivarch C, Haigh M, Holz C, Iyer G, Kato E, Keramidas K, Kitous A, Leblanc F, Liu J-Y, Löffler K, Luderer G, Marcucci A, McCollum D, Mima S, Sands R, Sano F, Strefler J, Tsutsui J, Van Vuuren D, Vrontisi Z, Wise M, Zhang R (2019) IAMC 1.5°C Scenario Explorer and Data hosted by IIASA. https://doi.org/10.5281/ZENODO.3363345. Accessed 19 Feb 2023

IEA (2021) Net Zero by 2050: A Roadmap for the Global Energy Sector. Paris, France, International Energy Agency

IPCC (2018) Summary for Policymakers. 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 [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 3-24. https://doi.org/10.1017/9781009157940.001.

James G, Witten D, Hastie T, Tibshirani R (2021) An introduction to statistical learning. Springer US, New York

Jaxa-Rozen M, Trutnevyte E (2021) Sources of uncertainty in long-term global scenarios of solar photovoltaic technology. Nat Clim Chang 11:266–273. https://doi.org/10.1038/s41558-021-00998-8

Kim D, Pham K, Oh JY, Lee SJ, Choi H (2022) Classification of surface settlement levels induced by TBM driving in urban areas using random forest with data-driven feature selection. Autom Constr 135:104109. https://doi.org/10.1016/j.autcon.2021.104109

Kirchner M, Mitter H, Schneider UA, Sommer M, Falkner K, Schmid E (2021) Uncertainty concepts for integrated modeling-Review and application for identifying uncertainties and uncertainty propagation pathways. Environ Model Softw 135:104905. https://doi.org/10.1016/J.ENVSOFT.2020.104905

Kwakkel JH (2017) The exploratory modeling workbench: an open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. Environ Model Softw 96:239–250. https://doi.org/10.1016/J.ENVSOFT.2017.06.054

Kwakkel JH, Auping WL, Pruyt E (2013) Dynamic scenario discovery under deep uncertainty: the future of copper. Technol Forecast Soc Chang 80:789–800. https://doi.org/10.1016/j.techfore.2012.09.012

Meyer M, Löschel A, Lutz C (2021) Carbon price dynamics in ambitious climate mitigation scenarios: An analysis based on the IAMC 1.5°c scenario explorer. Environ Res Commun 3(8):081007. https://doi.org/10.1088/2515-7620/ac02ad

Murphy KP (2012) Machine learning : a probabilistic perspective. MIT Press

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J (2011) Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825–2830

Rogelj J, Popp A, Calvin KV, Luderer G, Emmerling J, Gernaat D, Fujimori S, Strefler J, Hasegawa T, Marangoni G, Krey V (2018) Scenarios towards limiting global mean temperature increase below 1.5 °C. Nat Clim Chang 8:325–332. https://doi.org/10.1038/s41558-018-0091-3

Rotmans J, van Asselt MBA (2001) Uncertainty management in integrated assessment modeling: towards a pluralistic approach. Environ Monit Assess 69(2):101–130. https://doi.org/10.1023/A:1010722120729

Rozenberg J, Guivarch C, Lempert R, Hallegatte S (2014) Building SSPs for climate policy analysis: a scenario elicitation methodology to map the space of possible future challenges to mitigation and adaptation. Clim Chang 122:509–522. https://doi.org/10.1007/s10584-013-0904-3

Shiraki H, Sugiyama M (2020) Back to the basic: toward improvement of technoeconomic representation in integrated assessment models. Clim Chang 162:13–24. https://doi.org/10.1007/s10584-020-02731-4

Trutnevyte E, Guivarch C, Lempert R, Strachan N (2016) Reinvigorating the scenario technique to expand uncertainty consideration. Clim Chang 135(3):373–379. https://doi.org/10.1007/S10584-015-1585-X

UNFCCC (2015) The Paris Agreement. https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement. Accessed 19 Feb 2023

van Asselt MBA, Rotmans J (2002) Uncertainty in Integrated Assessment Modelling. Clim Chang 54(1):75–105. https://doi.org/10.1023/A:1015783803445

van Rijsbergen CJ (1979) Information retrieval, 2nd ed. Butterworths, London ;;Boston

Zhao Y, Zhu W, Wei P, Fang P, Zhang X, Yan N, Liu W, Zhao H, Wu Q (2022) Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period. Ecol Indic 135:108529. https://doi.org/10.1016/J.ECOLIND.2021.108529

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

Authors and Affiliations

Pei-Hao Li[1] · Steve Pye[1] · Ilkka Keppo[2] · Marc Jaxa-Rozen[3] · Evelina Trutnevyte[3]

Evelina Trutnevyte
evelina.trutnevyte@unige.ch

1 UCL Energy Institute, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, UK

2 Department of Mechanical Engineering, Aalto University, Otakaari 1 B, 02150 Espoo, Finland

3 Renewable Energy Systems, Institute for Environmental Sciences (ISE), Section of Earth and Environmental Sciences, University of Geneva, Uni Carl Vogt, Boulevard Carl Vogt 66, CH-1211 Geneva, Evelina, Switzerland