TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Dekker, M. M., Daioglou, V., Pietzcker, R., Rodrigues, R., de Boer, H. S., Dalla Longa, F., Drouet, L., Emmerling, J., Fattahi, A., Fotiou, T., Fragkos, P., Fricko, O., Gusheva, E., Harmsen, M., Huppmann, D., Kannavou, M., Krey, V., Lombardi, F., Luderer, G., ... van Vuuren, D. (2023). Identifying energy model fingerprints in mitigation scenarios. *Nature Energy*, *8*(12), 1395-1404. https://doi.org/10.1038/s41560-023-01399-1

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

nature energy

Article

https://doi.org/10.1038/s41560-023-01399-1

Identifying energy model fingerprints in mitigation scenarios

Received: 28 February 2023

Accepted: 5 October 2023

Published online: 06 November 2023

Check for updates

Mark M. Dekker[1,2] ✉, Vassilis Daioglou[1,2], Robert Pietzcker[3], Renato Rodrigues[3], Harmen-Sytze de Boer[1], Francesco Dalla Longa[4], Laurent Drouet[5], Johannes Emmerling[5], Amir Fattahi[4], Theofano Fotiou[6], Panagiotis Fragkos[6], Oliver Fricko[7], Ema Gusheva[8], Mathijs Harmsen[1,2], Daniel Huppmann[7], Maria Kannavou[6], Volker Krey[7], Francesco Lombardi[8], Gunnar Luderer[3,9], Stefan Pfenninger[8], Ioannis Tsiropoulos[6], Behnam Zakeri[7], Bob van der Zwaan[4,10,11], Will Usher[12] & Detlef van Vuuren[1,2]

Energy models are used to study emissions mitigation pathways, such as those compatible with the Paris Agreement goals. These models vary in structure, objectives, parameterization and level of detail, yielding differences in the computed energy and climate policy scenarios. To study model differences, diagnostic indicators are common practice in many academic fields, for example, in the physical climate sciences. However, they have not yet been applied systematically in mitigation literature, beyond addressing individual model dimensions. Here we address this gap by quantifying energy model typology along five dimensions: responsiveness, mitigation strategies, energy supply, energy demand and mitigation costs and effort, each expressed through several diagnostic indicators. The framework is applied to a diagnostic experiment with eight energy models in which we explore ten scenarios focusing on Europe. Comparing indicators to the ensemble yields comprehensive 'energy model fingerprints', which describe systematic model behaviour and contextualize model differences for future multi-model comparison studies.

The European Union's ambition to reach climate neutrality in 2050 as part of the European Green Deal[1] requires a thorough transformation of the full energy–economy system. Insights required for this transition are obtained from various lines of research, such as analysis of the technical mitigation potential, the effectiveness of policy instruments and opportunities for system changes given the interests of stakeholders and institutional barriers[2]. An important part of quantitative information on mitigation pathways is obtained from model-based scenario analyses, such as those recently published in the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report[3,4].

Still a large spread is associated with the scenario output, which originates from many types of uncertainty[5,6]. Structural uncertainty stems from differences in numerous assumptions regarding, among others, technological innovation and uptake, market behaviour,

[1]PBL Netherlands Environmental Assessment Agency, The Hague, Netherlands. [2]Copernicus Institute of Sustainable Development, Utrecht University, Utrecht, Netherlands. [3]Potsdam Institute for Climate Impact Research, Potsdam, Germany. [4]TNO Energy and Materials Transition, Amsterdam, Netherlands. [5]RFF-CMCC European Institute on Economics and the Environment (EIEE), Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy. [6]E3-Modelling SA, Athens, Greece. [7]Energy, Climate and Environment (ECE) Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. [8]Faculty of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands. [9]Technical University Berlin, Berlin, Germany. [10]Faculty of Science, University of Amsterdam, Amsterdam, Netherlands. [11]SAIS Europe, Johns Hopkins University, Bologna, Italy. [12]KTH Royal Institute of Technology, Stockholm, Sweden. ✉e-mail: mark.dekker@pbl.nl

preferences and changes in specific activities. Parametric uncertainties involve differences in parameter calibrations or are consequential to differences in sectoral granularity and regional and temporal scale. Fundamental modelling choices may also vary, such as those concerning mathematical formulation (for example, optimization versus simulation frameworks), model structure and foresight. Substantial differences can be recognized across the model outcomes[5], potentially even yielding contradictory observations[6]. To have a more accurate understanding of energy and climate policy scenario outcomes, it is important to have insights into both (1) where models differ substantially and where they agree and (2) how a model's output relates to the overall ensemble, yielding insights on outliers and discrepancies. This is specifically important because studies commonly use individual models rather than large ensembles in both scientific literature[7] and in policy reports (for example, in national policy studies). Only when single-model results are contextualized by the model's position in the larger ensemble, the reader would be able to have a complete and correct interpretation of the output. Additionally, such quantification of the model's position in the larger model ensemble allows for tracking model development[8].

Both questions require a stylized set of results across the model range in which only a well-defined number of assumptions is varied. To this end, multi-model comparison exercises have been effective. This is done within confined projects[3] and, for instance, in the long tradition[9] of studies by the Energy Modelling Forum[10,11], of which many scenario runs are collectively used in the Assessment Reports by the IPCC[10,11]. Still the observed large model differences—especially in estimates of costs[12], the diffusion of individual renewable energy technologies[13] and demand sector development[14,15]—motivate more research in this area and emphasize the importance of interpreting single-model results in light of larger model ensembles.

Many multi-model comparison studies test the (un)certainty of an outcome by looking at the range across models in scenarios designed for other purposes, for example, to describe the effect of current mitigation measures[6]. While insightful for that particular question, quantifying and evaluating overall model behaviour requires analysis beyond typical scenarios and typical variables—hence the importance of analysis of diagnostic scenarios, expressed in diagnostic variables or indicators. Such practice is well established in the climate and atmospheric sciences[16–18]. Examples of indicators are the equilibrium climate sensitivity and the transient climate response, which are associated with diagnostic scenarios in which the CO_2 concentration is doubled or quadrupled. In the emissions mitigation literature, examples of diagnostic multi-model studies are Kriegler et al. (2015) and Harmsen et al. (2021), which propose a limited set of diagnostic indicators that reflect crucial model behaviour aspects[8,19]. Using two diagnostic scenarios—with a constant and exponentially increasing carbon tax, respectively—Kriegler et al. condense the output of models into four core diagnostic indicators: (1) the relative abatement index, expressing the overall mitigation effort; (2) the carbon intensity over energy intensity, expressing the mitigation strategy; (3) the transformation index, expressing the required transformation; and (4) the cost per abatement value, expressing the policy costs per unit of marginal abatement. Harmsen et al. continued on this path by updating the calculations by Kriegler et al. for a larger number of models and model versions, adding an indicator on primary fossil fuel energy reduction and an analysis of model inertia. The latter yields a total of six key diagnostic indicators for energy system models and integrated assessment models (IAMs).

Some model aspects are more indicative of model behaviour than others: for example, model sensitivity to carbon taxes already reveals much about drivers of the model's output, much like equilibrium climate sensitivity is a core classifier in climate models. Still analysing individual model dimensions (or indicators) yields only a limited view due to the higher-dimensional intertwinement of mitigation, policy, energy supply and energy demand. For example, differences in solar power deployment under similar emissions levels can be understood only when information about wind power (as a potential competitor), carbon dioxide capture and storage (CCS) or energy demand reductions are provided. This motivates comparing multiple model dimensions at the same time. While Harmsen et al.'s six indicators are useful to classify models for each indicator individually (that is, one-dimensional), the aim of this paper is to characterize the overall model typology (that is, high-dimensional)—moving beyond a mere long list of 'individual' diagnostic indicators, towards developing a comprehensive overview or 'story' of the model's behaviour. This requires the analysis to go back and forth between many different aspects of the same model. In addition, a comprehensive overview of model behaviour would include an extension of the existing list of diagnostic indicators to a higher level of detail (Supplementary Information A.6).

Framework and fingerprints

Here we quantify typical model behaviour in a framework that we colloquially refer to as the model's 'fingerprint'. It assesses five key model behaviour dimensions: responsiveness, mitigation strategies, energy supply, energy demand and costs and effort. Each dimension is subsequently expressed in several diagnostic indicators, as shown in Fig. 1. The set of diagnostic indicators extends previous work[8,19] with new indicators to obtain a more complete and multi-dimensional perspective on model behaviour (Supplementary Information A.6 for a comparison with earlier work). For example, the 'responsiveness' dimension describes more than only tax response (R_1), adding the speed of response (R_{2-3}) and scenario sensitivities (R_4 and R_5). Other notable extensions on previous work are including carbon capture (M_3), the role of non-CO_2 (M_4), more details on the energy supply mix (Es_{1-7}), electrification of end-use sectors (Ed_{1-3}), hydrogen use (Ed_4) and demand shifts (C_3). In addition, all indicators are computed from a more extended set of diagnostic scenarios (Methods), containing scenario variation beyond only varying carbon price trajectories.

The framework's principle is to compare a model's scenario range (Fig. 1, yellow shaded) to the ensemble statistics in terms of its medians (η) and standard deviations (σ), yielding the range of typical (relative) model behaviour: the area towards the outer ring indicates above-median (up to $\eta + 2\sigma$) output and the area towards the circle centre indicates below-median (down to $\eta - 2\sigma$) output. Whereas scenario ensembles of energy models are rarely normally distributed, defining the range this way does ensure more intuitive equal ranges above and below the median. The diagnostic exercise performed to fill these diagrams includes eight different energy system models, involving ten diagnostic scenarios tailored to explore different model aspects. The scenarios (Methods and Table 3) have a high carbon price, aimed at approximating the Paris Agreement goals and the European Union's climate neutrality goal.

In Figs. 2 and 3, one can see the outcomes of the framework for the eight involved energy system model versions. For illustration purposes, we discuss the fingerprints of three model versions in detail before moving on to the more overarching patterns and typologies that the fingerprints reveal. Detailed model-by-model observations, as well as model acronym meanings, are provided in Supplementary Information C. In Fig. 2a, we see the fingerprint of the IMAGE model[20], a process-based IAM. The model shows medium levels of carbon emissions abatement (R_1) and carbon intensity reduction (M_1). Still it has a relatively low mitigation timescale (R_2), high maximum mitigation speed (R_3) and sensitivities (R_{4-5}) and relies relatively more on energy intensity reductions and carbon capture (M_{2-3}). Consequentially, primary fossil use is higher than the ensemble median, notably coal (Es_1). Relatively low values of solar and nuclear energy use (Es_3, Es_7) contrast with the high use of biomass (Es_6, in most scenarios) in the primary energy mix, and high electrification of transport (Ed_1), medium for industry (Ed_2) and low for buildings (Ed_3) is found. Transformation in

https://doi.org/10.1038/s41560-023-01399-1

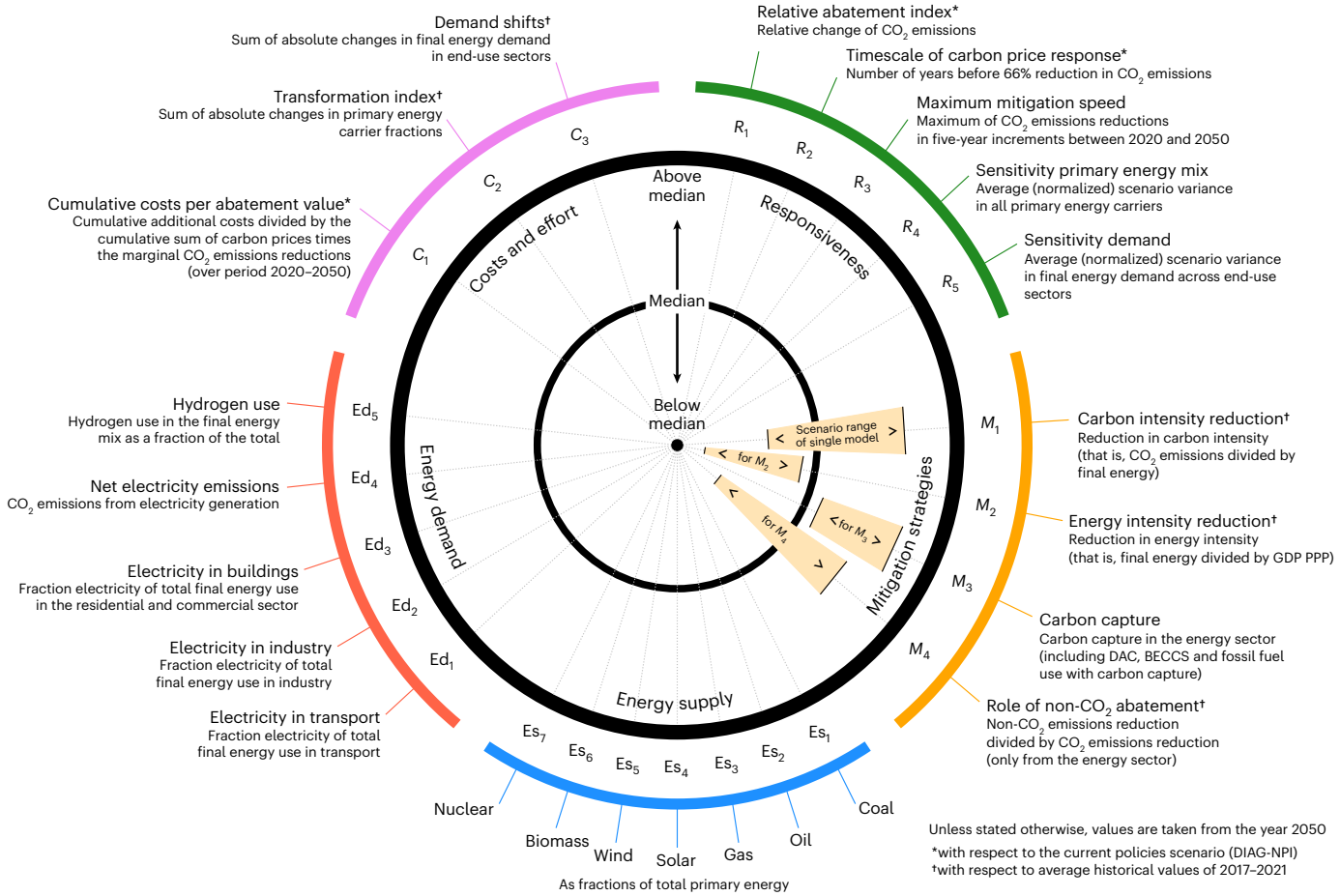


Fig. 1 | Framework of model fingerprints. Explanation of the framework that is used to compute model fingerprints. The five diagnostic indicator dimensions are shown in colours: responsiveness (green), mitigation strategies (yellow), energy supply (blue), energy demand (red) and costs and effort (purple). The framework uses the statistics of the ensemble to visualize the results of a single model, resulting in a circular diagram per model. The inner circle represents the median per indicator, calculated from the ensemble containing all model-scenario combinations—that is, including all other models as well. The outer circle and the centre indicate the medians ± two standard deviations, respectively. Example ranges (yellow shaded, not based on data) for indicators M_1–M_4 are shown: the ranges show the range that the respective model covers across its scenarios. These ranges (Figs. 2–3) exclude the scenario that includes only current implemented policies (referred to as the DIAG-NPI scenario, Table 3), which is used only as a reference for R_1, R_2 and C_1. Note that differences in this reference may cause differences in the indicators; this is discussed in Supplementary Information A.3, which is why we use the comparison with historical values as much as possible (whenever a reference is needed). The shape or indicator arc widths do not have any quantitative meaning themselves. For M_4, direct air capture (DAC) and bioenergy with carbon capture and storage (BECCS) are included, and for the energy intensity in M_2, we use gross domestic product (GDP) at purchasing power parity (PPP).

the energy supply and demand sides depend on the scenario, while the costs remain approximately average ($C_{1–3}$).

Figure 2b shows the fingerprint of the REMIND model[21], which is a Ramsey-type general equilibrium growth IAM—quite different from IMAGE. Indeed, the fingerprint also looks very different indicating substantial differences between the output of REMIND on the one hand, and those of IMAGE on the other hand. Striking are the higher levels of abatement (R_1), carbon intensity reductions (M_1) and fast reaction to carbon pricing (low R_2, high R_3). Consistent with these observations, primary fossil use (Es_1–Es_3) is limited, whereas variable renewable energy and electrification in transport and buildings (D_1, D_2) are higher than ensemble medians. For industry, electrification is quite sensitive to scenario assumptions. The associated transformation index (C_2) is expectedly high, while the costs (C_1) are lower than the ensemble median.

A third and final example we will discuss here is shown in Fig. 3b—the PRIMES model[22], which is an energy system model that provides projections of energy demand, supply, prices and investments. This model does not react heavily to the various scenario assumptions, reflected in the low scenario sensitivity of the primary energy mix (R_4) and demand (R_5) and in the narrow scenario ranges shown in the fingerprint diagram. PRIMES projects near-median levels for most indicators, with a few exceptions: a relatively low mitigation speed potential (R_3), high energy intensity reduction (M_2), high relative non-CO_2 reductions (M_4), low solar energy use (Es_4) and notably a very high hydrogen use (Ed_5).

Model typology

A summary of the general tendencies of the models is shown in Table 1. We use a similar (but extended) approach to this as in previous literature[8], where a general classification was given of each of the models. The assessments in the table are directly based on the fingerprints in Figs. 2 and 3 and quantify the extent to which the models deviate from the ensemble.

The spread in Table 1 indicates the significance of the differences in reported model deviations. For example, in response to carbon taxing in REMIND (+1.8σ) versus that in TIAM-ECN (−1.6σ), the energy intensities in WITCH (+2.1σ) and TIAM-ECN (−1.7σ) or the relative mitigation of non-CO_2 emissions in IMAGE and REMIND (−1.1σ and −1.2σ) was compared to that in TIAM-ECN (+1.8σ). Whereas such values are

https://doi.org/10.1038/s41560-023-01399-1

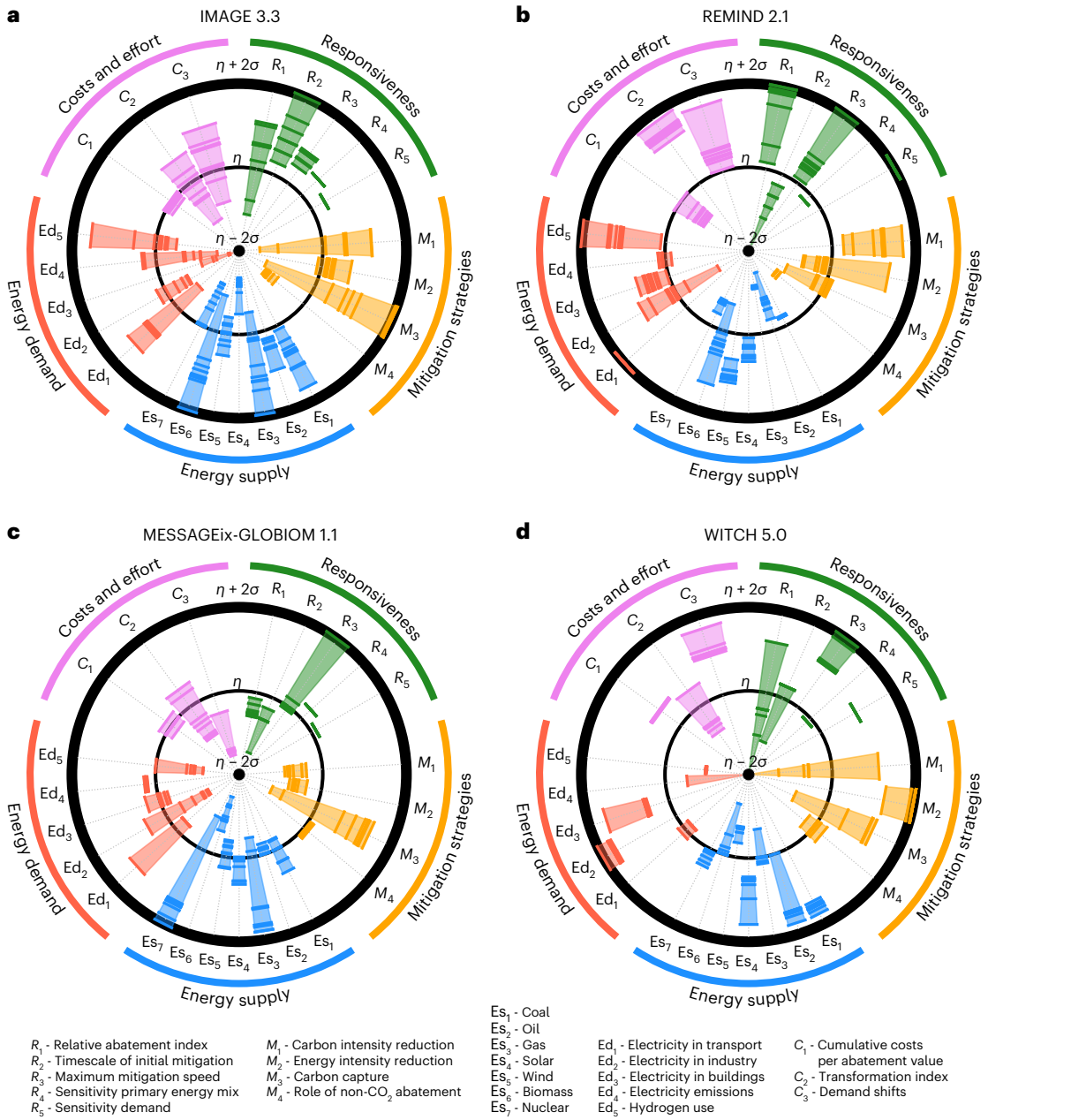


Fig. 2 | Model fingerprints. a–**d**, Model typologies or 'fingerprints' of the IMAGE (**a**), REMIND (**b**), MESSAGEix-GLOBIOM (**c**) and WITCH (**d**) models for Europe in 2050. The axis for each indicator ranges between the ensemble median (η) ± two standard deviations (*σ*). These medians and standard deviations are computed from the full ensemble, that is, the eight models and nine linearly increasing-price scenarios (excluding the current-policies scenarios), and the coloured shaded scenario ranges are for each individual model indicated by the panel titles. Data close to the centre reflect below-median results. Data towards the outer ring reflect above-median results. More details on the indicators are in Fig. 1.

expected for individual model runs (as the values are expressed in standard deviations), this table reports the *medians per model*, revealing that the model outputs are indeed substantially different in certain respects. In other words, inter-model differences are high compared to intra-model (that is, scenario) differences. Note that Table 1 contains combinations of diagnostic indicators. For example, the averaging electrification across the end-use sectors conceals the high model differences in individual sector electrifications such as transport in REMIND (+2.2*σ*) and industry in WITCH (+1.7*σ*). For more detail, we refer the reader to Figs. 2 and 3 and Supplementary Information B.1.

To provide structure to the discussion, we sorted the models by their tax response (second column, R_1): that is, carbon emissions in 2050 under a carbon tax scenario compared with that in the current-policies scenario of the same model. REMIND is by far the most responsive (green); it shows almost two standard deviations higher relative abatement and, in addition, has a high scenario response. This high responsiveness coincides with high carbon intensity reduction, renewables and electrification. It emphasizes carbon intensity reduction with emphasis on CO_2 emissions while projecting average numbers for energy intensity reduction and carbon capture. Euro-Calliope partially shows similar results, also having a (moderately) high tax response, carbon intensity reduction and renewables rollout but to a lesser extent than REMIND. From a system point of view, these models show a consistent fingerprint; the high tax response (green) is achieved by replacing fossils (yellow) by renewables (blue) and high electrification (orange). IMAGE can also be regarded as a model with moderately

https://doi.org/10.1038/s41560-023-01399-1

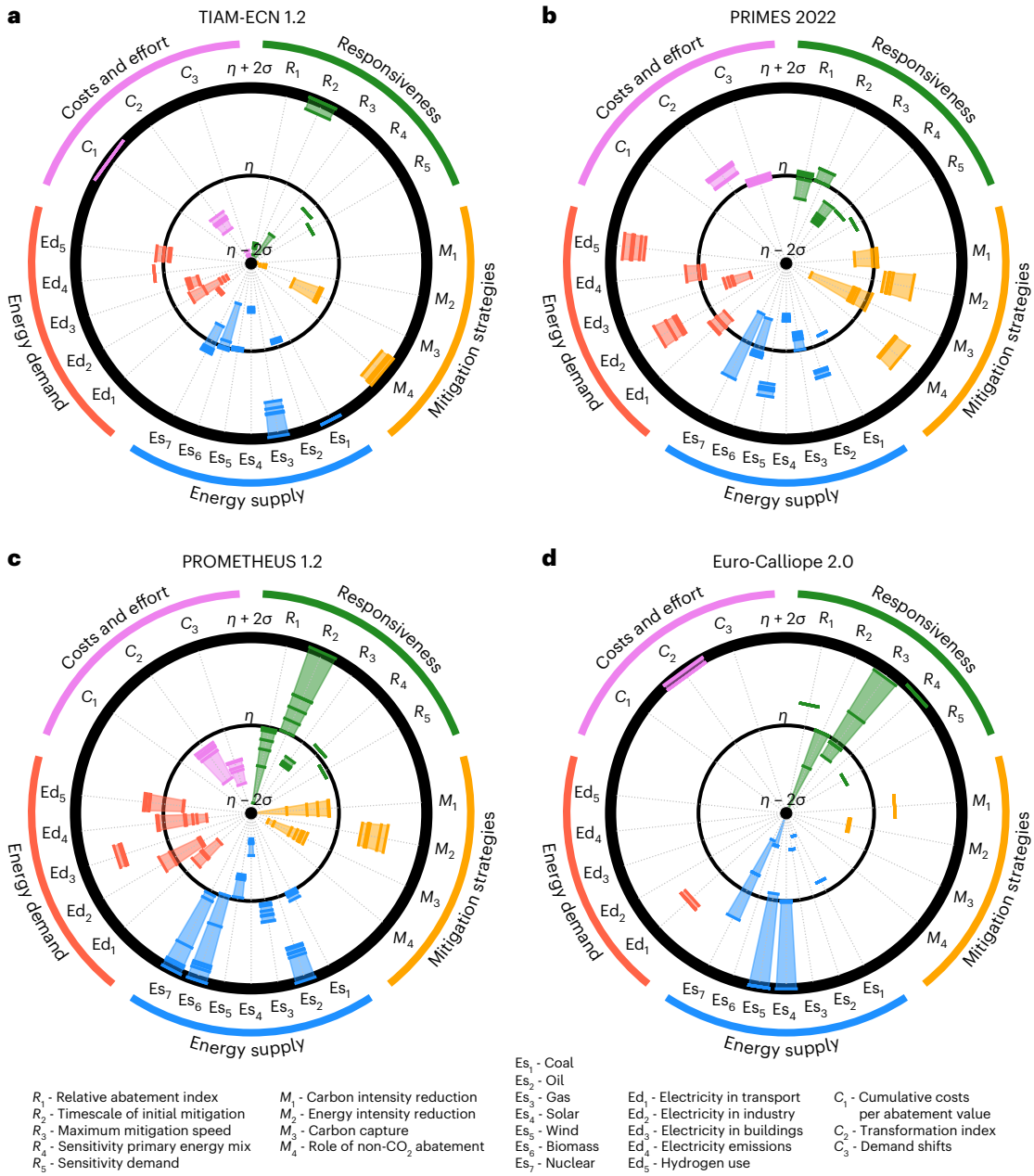


Fig. 3 | Model fingerprints. a–**d**, Model typologies or 'fingerprints' of the TIAM-ECN (**a**), PRIMES (**b**), PROMETHEUS (**c**) and sector-coupled Euro-Calliope (**d**) models, for Europe in 2050, similar to Fig. 2. The axis for each indicator ranges between the ensemble median (η) ± two standard deviations (σ). These medians and standard deviations are computed from the full ensemble, that is, the eight models and nine linearly increasing-price scenarios (excluding the current-policies scenarios), and the coloured shaded scenario ranges are for each individual model indicated by the panel titles. Data close to the centre reflect below-median results. Data towards the outer ring reflect above-median results. More details on the indicators are in Fig. 1.

high responsiveness, projects relatively high carbon intensity reductions and, like REMIND, indicates relatively low non-CO_2 reductions compared with CO_2. However, as observed in Fig. 2, it differs from REMIND and Euro-Calliope by relying more on carbon capture (+1.4σ) and fossils (+0.7σ).

This consistent fingerprint can also be observed at the other end of the spectrum. This is clearly expressed in TIAM-ECN output, which has a low tax response, low carbon intensity reduction, high fossil use and low electrification. MESSAGEix also shows this mirror image to some extent, but it is closer to the rest of the ensemble than TIAM-ECN in this respect. While this model initially rapidly mitigates emissions (reflected in a low R_2—approximating REMIND), the abatement becomes relatively slow over time, yielding an overall medium to low relative abatement (R_1) of −0.5σ. Note that this is partially consequential to differences in what years are represented in each model year—for MESSAGE, 2050 represents the five preceding years. These abatement projections are accompanied by relatively low (sectoral average) electrification of −0.4σ and higher fossil use of +0.5σ and also a low carbon and energy intensity reductions (−0.8σ and −0.7σ, respectively; Supplementary Information B.1).

Besides MESSAGEix, the three other models with medium tax response—PRIMES, PROMETHEUS and WITCH—are the models with the highest energy intensity reductions while having low (WITCH) or medium (PRIMES, PROMETHEUS) carbon intensity reductions. This is most clearly visible in WITCH output, projecting energy intensity reductions of over two standard deviations higher. The three

https://doi.org/10.1038/s41560-023-01399-1

Table 1 | Overview of model typologies

	Responsiveness		Mitigation strategies			Energy supply	Electrification	Costs
	Carbon tax sensitivity	Scenario sensitivity	Reduction CI/EI	Carbon capture	Non-CO_2 reduction compared to CO_2	Fossils/ renewables/ other	Low/medium/ high	Cheap/ medium/ expensive
	R_1	R_{4-5}	M_{1-2}	M_3	M_4	Es_{1-7}	Ed_{1-3}	C_1 inverted
REMIND IT-GE	**High (+1.8σ)**	**High (+1.3σ)**	**CI reduction (+1.2σ)**	Medium (0.0σ)	**Low (−1.2σ)**	Renewables (+0.7σ)	High (+0.9σ)	Cheap (−0.6σ)
Euro-Calliope IT	Medium (+0.5σ)	**High (+1.2σ)**	CI reduction (+0.5σ)	–	–	**Renewables (+1.3σ)**	– (–)	–
IMAGE RD-PE	Medium (+0.5σ)	Medium (+0.5σ)	CI reduction (+0.5σ)	**High (+1.1σ)**	**Low (−1.1σ)**	Fossils (+0.7σ)	Medium (−0.5σ)	Medium (0.0σ)
PRIMES IT-PE	Medium (+0.1σ)	Medium (−0.3σ)	EI reduction (+0.4σ)	Medium (−0.4σ)	High (+1.0σ)	Renewables (+0.1σ)	Medium (0.0σ)	–
PROMETHEUS RD-PE	Medium (−0.3σ)	Medium (0.0σ)	EI reduction (+0.7σ)	Low (−0.9σ)	–	**Other (+1.9σ)**	Medium (0.0σ)	–
WITCH IT-GE	Medium (−0.4σ)	Medium (+0.4σ)	**EI reduction (+2.1σ)**	**High (+1.1σ)**	Medium (0.0σ)	**Fossils (+1.0σ)**	High (+0.7σ)	Expensive (+0.6σ)
MESSAGEix-GLOBIOM IT-GE	Medium (−0.5σ)	Medium (+0.2σ)	EI reduction (−0.7σ)	**High (+1.1σ)**	Medium (+0.1σ)	Fossils (+0.5σ)	Medium (−0.4σ)	Medium (0.0σ)
TIAM-ECN IT-PE	**Low (−1.6σ)**	Medium (−0.4σ)	**EI reduction (−1.7σ)**	Medium (−0.3σ)	**High (+1.8σ)**	**Fossil (+1.4σ)**	Low (−0.9σ)	**Expensive (+2.8σ)**

The columns refer to individual or combined diagnostic indicators and the cells below it show how many standard deviations the model median differs from the ensemble median (Methods). Missing values are denoted with dashes. 'Medium' implies a deviation within ±0.5σ. 'High' and 'Low' imply larger positive and negative deviations, respectively. Deviations of more than ±1σ are emphasized in bold. Abbreviations in the first column: IT (inter-temporal), GE (general equilibrium), PE (partial equilibrium) and RD (recursive dynamics). For carbon and energy intensity (CI and EI) and in the column on energy supply, only the highest deviation is indicated. An extended version of this table with values for each diagnostic indicator can be found in Supplementary Information B.1.

models have another commonality: they are all moderately sensitive to scenario assumptions. However, in many other ways, their projections notably differ. WITCH relies more on carbon capture in contrast to PROMETHEUS, which, in turn, has higher nuclear and bioenergy in their energy supply mix than any other model in this experiment. The latter is related to the high level of biofuel use in PROMETHEUS.

A commonality between WITCH, IMAGE and MESSAGEix can also be found; they all have rather high fossil use and carbon capture (the exception to this pattern seems to be TIAM-ECN, which does not project as much carbon capture). Interestingly, while high carbon capture intuitively would generate high abatement and carbon intensity reductions, this is not visible here, with, for example, MESSAGEix having +1.1σ carbon capture but −0.8σ carbon intensity reduction (Supplementary Information B.1).

In previous papers, scholars have tried to link model behaviour to the underlying model type in terms of solution approach (for example, partial or general equilibrium, recursive dynamic or inter-temporal). For example, Daioglou et al. (2020)[23] tried to link bioenergy use to model type, and Harmsen et al. (2021)[8] compared the output of six diagnostic indicators to the model types. However, in both studies, 'no direct relationship between model type and model behaviour' was found[8]. Even though this study contains more unique scenarios than in previous studies, we again find no clear link between model behaviour and model type (although the set considered here does include a smaller set of models). In Table 1, we see that inter-model differences are not only higher than inter-scenario differences in many of the indicators but also higher than inter-model type differences. Technology-rich energy system models can be among the most tax-sensitive models for abatement (Euro-Calliope) or among the lesser tax-sensitive ones (MESSAGEix)—that is, being technology rich does not imply a certain level of carbon tax sensitivity. Similar patterns can be observed for general versus partial equilibrium models and inter-temporal vs myopic models. It is important to emphasize that while having more scenarios than in previous studies allows for a more representative fingerprint of each model, a clear limitation for trying to attribute model behaviour to its type is the limited number of models (eight). Adding more models could help in observing such relations, but models are rarely fully independent from each other—which is a limitation in these model comparison exercises in general.

Discussion and conclusion

In this paper, we propose a unique framework to characterize energy model typology by quantifying their 'fingerprints' in scenario ensembles. We identified five key model dimensions along which models differ: model responsiveness, mitigation strategies, energy supply, energy demand and costs and effort split further into a total of 24 diagnostic indicators. A number of these indicators are re-used from existing literature[8,19] (Supplementary Information A.6), but we add more dimensions such as the energy supply mix, carbon capture, non-CO_2 and demand responses, which allows for a more complete interpretation of model behaviour. The indicators are calculated and visualized relative to the ensemble statistics of a set of tailored diagnostic scenarios, yielding a comprehensive insight into each model's fingerprint in relation to the other models.

The framework contextualizes results from individual models (or rather model versions, which may also evolve their behaviour[24,25]) by identifying typical model behaviour, which yields better interpretation and understanding of them. The latter, of course, is of vital importance to both researchers and policymakers[8,19]. Additionally, the substantial model differences, being commonly larger than inter-scenario differences, motivate caution on using individual model results without ensemble context in general. Specifically, we find highly tax-responsive models to also have the highest renewables, electrification and carbon intensity reduction. IMAGE is also relatively highly responsive but distinguishes itself with high fossil and carbon capture use. TIAM-ECN, having the lowest tax response, indeed also shows a low carbon intensity reduction. MESSAGEix shows a similar pattern (but more closely to the rest of the ensemble), and both of these models additionally show a low energy intensity reduction. Most models with medium tax-responsive models (WITCH, PROMETHEUS and PRIMES) show higher energy intensity reductions and moderate scenario responsiveness but vary in many other ways.

The framework also has limitations. First of all, the scenario projections are not statistical predictions—hence, the model-scenario ensemble cannot be used as such, and the ensemble median should not be interpreted as 'most appropriate'. Analogously, model behaviour that is distinct from the rest of the ensemble (outliers) should not be treated as 'least probable' and neither should the sorting of models by carbon tax response (in Table 1) reflect a form of 'ranking'. Whereas typical other multi-model comparison studies focus much more on robust messages across models, our aim is to identify and highlight their differences. The model differences presented here do not even necessarily illustrate a lack of consensus on future outcomes. Potentially, they also reveal that deep mitigation can be achieved in multiple ways (that is, a form of 'policy freedom'), albeit each conditioned in various respects, motivating the high dimensionality of this analysis. Hence, for the purpose of exploring the future, model differences can therefore be useful. This study aims only to map these differences; it does not identify its sources—which also requires the analysis of model inputs and parameters[26,27], for example associated with efficiency differences in final energy (R_5, Ed_{1-3} and C_3). A second important remark is that the fingerprint framework and Table 1 contain quantities that are relative to the ensemble. Therefore, the fingerprints dilute information on absolute differences—motivating the importance of Supplementary Table 4—which should always be taken into account when using this framework. The relative nature of the framework also means that there is no single 'absolute' fingerprint of any model; it is always subject to the ensemble the model is being compared to. Structural bias in the community (that is, across all included models) will therefore not be easily detected. Adding models to the presented diagnostic exercise or choosing a different regional scope will modify the model fingerprints (for example, Supplementary Information B.4 for global results). Still by choosing the diagnostic scenarios with a wide range of assumptions and a large variety of models, we do approximate a comprehensive overview of the behaviour of the eight models included (for the region of Europe). Finally, we also note that the scope of this research was Europe and deep mitigation scenarios. The model fingerprints will look different when applied to other regions or global results and to scenarios with little climate policy.

Even though we show only one application of the framework, it is intended as a general framework—both the methodology, code and data are flexible in nature. The model fingerprints for global output (rather than European) in Supplementary Information B.4 are an illustration of this, but it is also well suitable as a diagnostic tool for model development by determining the fingerprints for different versions of a single model. To quantify model fingerprints at a more detailed sectoral level, the dimensions, indicators and even the scenarios can be adjusted accordingly. On diagnostic indicators, earlier literature[8,19] stresses the importance of applicability to diverse models, beyond only the relevance, quantifiability and identification of heterogeneity. For the purpose of the latter, this framework with the current selection of diagnostic indicators is well suited whereas for the former, the framework may require translation—for example, for models that cover only the electricity sector. This way the framework can form the basis of future model intercomparison projects.

Methods

Diagnostic experiment

We conducted a diagnostic experiment in which we ran ten different scenarios using eight different models. The models used are shown in Table 2. The models intentionally vary in many respects: coverage (global or Europe), scope (IAMs and energy system models), type (simulations and optimizations) and more. Whereas the focus of this work is Europe, most models in fact have a global coverage. For global models, the basis of all scenarios is globally the trajectory associated with current policies implemented. Carbon prices are also issued globally, and whenever applicable, specific scenario assumptions are listed (in Table 3) for both the globe and Europe.

The scenarios in this diagnostic experiment are described in Table 3. A current-policies (DIAG-NPI) baseline scenario is included for reference. All other scenarios are based on DIAG-NPI but, in addition, have a predefined fixed linearly increasing carbon price profile and integrate additional constraints or assumptions (third column in Table 3). One scenario has the linearly increasing carbon price next to current national policies implemented (DIAG-C400-lin) without any additional assumptions. The carbon price profile $cp(t)$ is up to 2025 similar to each model's current policies ($cp_{NPI}(t)$, from DIAG-NPI), after which it becomes a linearly increasing tax as follows, expressed in US dollars (2010) and time t in years:

$$cp(t) = \begin{cases} cp_{NPI}(t) & \text{if } t < 2025 \\ 130 + 18 \times (t - 2025) & \text{if } t \geq 2025 \end{cases} \quad (1)$$

which results in a carbon price of US$580 $t^{-1}CO_2$ in 2050 (US$400 $t^{-1}CO_2$ in 2040). The eight variations of DIAG-C400-lin are explained in Table 3 and include one with a limit on the bioenergy potential (DIAG-C400-lin-LimBio) and one where electricity prices artificially kept low (DIAG-C400-lin-HighElectrification).

Definitions of indicators

In this diagnostic exercise, we focus on Europe and 2050, but the framework can easily be adjusted for other regions or time frames. We consider only CO_2 emissions and no other Kyoto gases and only for the energy sector (the only exception being indicator M_4) because not all models from the ensemble include emissions from land use or other non-energy sectors in their scenario results. Another reason involves the focus of this study, which is on characterizing energy models rather than climate policy models in the broader sense. As shown to the bottom right of Fig. 1, some indicators are computed relative either to the current-policies scenario or to historical values of 2017–2021. The current-policies scenarios already differ substantially between models (Supplementary Information A.3), which can partially be explained by differences in regional aggregation (Supplementary Information A.1: for example, REMIND shows CO_2 emissions reductions by 74% (for EU28) while MESSAGEix shows increases by 33% (also including Turkey). Such differences in the current-policies scenario will affect the indicators, but for some (that is, R_1 and R_2), it is required to compare to a scenario (from the same model) with a lower carbon price. A more elaborate discussion on using the current-policies scenario or historical data (2017–2021) as a reference is given in Supplementary Information A.3. We emphasize that we build upon earlier work[8,19] in several of these indicators and that while some indicators discussed in this section are new definitions, the contribution of this paper lies in bringing many model dimensions together in one comprehensive model typology overview.

We refer to the first dimension as *Responsiveness* (green), which expresses the response of model output to various conditions into five indicators. The first indicator in this category is the sensitivity of the model's calculated emissions abatement to carbon pricing (R_1), which we quantify using the relative abatement index, that is, the relative reduction in emissions of carbon pricing scenarios with respect to the current-policies base scenario and emissions from energy and industrial processes for scenario s and year t denoted by $E(s, t)$:

$$R_1(s) = \frac{E(\text{NPI}, 2050) - E(s, 2050)}{E(\text{NPI}, 2050)} \quad (2)$$

We calculate R_1 by comparing the linearly increasing carbon price scenarios (Table 3) to the current-policies scenario, but we note that the term 'relative abatement index' is used previously for other pricing profiles as well: for example, Harmsen et al. uses the same terminology

Table 2 | Models used in the analysis

Abbreviation	Model	Coverage	Type	Institution
IMA	IMAGE 3.3	Global	IAM, simulation (RD-PE)	Netherlands Environmental Assessment Agency
TIA	TIAM-ECN 1.2	Global	IAM, optimization (IT-PE)	Netherlands Organization for Applied Scientific Research
MES	MESSAGEix-GLOBIOM 1.1	Global	IAM, optimization (IT-GE)	International Institute for Applied Systems Analysis
REM	REMIND 2.1	Global	IAM, optimization (IT-GE)	Potsdam Institute for Climate Impact Research
WIT	WITCH 5.0	Global	IAM, optimization (IT-GE)	Euro-Mediterranean Center for Climate Change
PRO	PROMETHEUS 1.2	Global	IAM, simulation (RD-PE)	E3-Modelling
PRI	PRIMES 2022	Europe	IAM, optimization (IT-PE)	E3-Modelling
EUR	Sector-coupled Euro-Calliope	Europe	ES, optimization (IT)	TU Delft

Model types (fourth column) are abbreviated: IAM (integrated assessment models), IT (inter-temporal), GE (general equilibrium), PE (partial equilibrium) and RD (recursive dynamics). More information can be found at https://www.iamcdocumentation.eu/index.php/IAMC_wiki. When referring to 'Europe', we mean the European Union together with the United Kingdom (that is, EU28), or the closest resemblance of this per model based on their regional aggregations.

Table 3 | Description of the diagnostic scenarios

Scenario name	Carbon pricing	Additional assumptions
DIAG-NPI	National policies implemented	–
DIAG-C400-lin	Linear, equation (1) (580 US$\$ t^{-1}CO_2$ in 2050)	–
DIAG-C400-lin-LimBio	Identical to DIAG-C400-lin	Global primary modern bioenergy supply limited to 100 EJ, which is on the lower end of the range given in the Sixth Assessment Report by IPCC[32]. Maximum European primary biomass use is limited to 7 EJ, which is close to current use[33]. Biomass imports of Europe set to 0.
DIAG-C400-lin-LimCCS	Identical to DIAG-C400-lin	CCS (including BECCS and DACCS) limited to 2 Gt CO_2 per year globally. Max CCS use in Europe limited to 250 Mt CO_2 per year.
DIAG-C400-lin-LimNuclear	Identical to DIAG-C400-lin	Nuclear power limited to today's levels and no new constructions are allowed except capacity already under construction.
DIAG-C400-lin-HighVRE	Identical to DIAG-C400-lin	Set direct LCOE for offshore wind at €30 MWh^{-1}, onshore wind at €20 MWh^{-1} and solar PV at €10 MWh^{-1} by 2050. Linear interpolation from current prices.
DIAG-C400-lin-HighElectrification	Identical to DIAG-C400-lin	Provide emissions-free electricity for demand sectors at €30 MWh^{-1} (excluding transmission and distribution).
DIAG-C400-lin-H2	Identical to DIAG-C400-lin	Provide emissions-free hydrogen for demand sectors at €45 MWh^{-1} (excluding transmission and distribution).
DIAG-C400-lin-ResidualFossil	Identical to DIAG-C400-lin	Set fossil fuel primary energy prices at €70 GJ^{-1} (before carbon pricing).
DIAG-C400-lin-HighEff	Identical to DIAG-C400-lin	By means of a variety of lifestyle change or high efficiency options (models are free to choose), reach final energy values of 26 EJ in Europe in 2050 including bunkers and non-energy.

To align with earlier diagnostic scenario runs[8], 2010 US dollar is used as currency for the carbon pricing. All other prices and monetary variables are in 2020 Euro. More details can be found in the model-scenario protocol, publicly available[32]. Because of the variety in model characteristics, not all models were able to run all prescribed scenarios. Still over 90% of all theoretically possible model-scenario combinations are included in the database (Supplementary Information A.5).

for an exponentially increasing carbon price[8], while we use a linearly increasing one (equation (1)). The second indicator (R_2) is the timescale of carbon price response, which is defined as the number of years until the scenario has 66% lower CO_2 emissions than the current-policies scenario has in the same year. A related (Supplementary Information A.2) third indicator is the maximum mitigation speed (R_3), which is the maximum emissions reduction over any five-year increment, yielding insights in how fast the model can mitigate in short periods. The fourth (R_4) and fifth (R_5) indicators focus on scenario sensitivity for each model, resulting in a single value (per model) rather than a scenario range, indicating how the primary energy mix and the final energy demand across sectors vary among its scenarios. We write the fraction of an energy carrier c in the primary energy mix in scenario s as $f_{c,s}$, its average across all scenarios as $\bar{f_c}$ and the total number of unique carriers considered as n_c. Then, R_4 is computed as follows:

$$R_4 = \frac{1}{n_c}\sum_c \mathrm{var}\left(\frac{f_{c,s}}{\bar{f_c}}\right) \quad (3)$$

In other words, it represents the average of the inter-scenario variances per primary energy carrier fraction. R_5 is computed similarly but uses final energy use across the industry, transportation and buildings sectors (instead of primary energy carrier fractions). If R_4 and R_5 are high, there is much inter-scenario variation in the primary energy mix or final energy demand in the model's output, while if these values are low, the scenarios are more or less similar and the model is relatively static with respect to the scenario assumptions.

The second dimension to express model typology concerns the *mitigation strategies* (yellow in Fig. 1). Whereas to some extent all models rely on all mitigation strategies we categorize here, it is useful to quantify which are more prominent in model output relative to others in the ensemble. The first indicator in this dimension is the reduction in carbon intensity (M_1, emissions divided by the final energy) compared to the baseline scenario. Similarly, the second mitigation strategy (M_2) involves reduction in energy intensity, that is, final energy divided by the gross domestic product (purchasing power parity). These reductions are compared to the average historical values between 2017

and 2021, by taking an ensemble average across this period. The third mitigation indicator (M_3) is expressed in the amount of carbon captured (including direct air capture) in 2050, focusing on the energy sector. The fourth (M_4) is the quotient between non-CO_2 emissions reduction and CO_2 emissions reduction (using 100-year global warming potential values analogous to IPCC Sixth Assessment Report). Because only a few models reported non-CO_2 emissions specifically from the energy sector, we are calculating this quotient based on non-CO_2 and CO_2 emissions from all sources.

The third dimension involves *energy supply*, expressed in the primary energy mix (blue). The associated indicators are primary energy consumption per carrier calculated as fractions of the total primary energy in 2050: coal (Es_1), oil (Es_2), gas (Es_3), solar (Es_4), wind (Es_5), biomass (Es_6) and nuclear (Es_7).

The fourth dimension focuses on the *energy demand* (red). We have chosen to split up the three major end-use sectors transport, industry and buildings when looking at their respective electrification (Ed_1, Ed_2 and Ed_3). The fourth indicator quantifies the CO_2 emissions in the electricity sector (Ed_4), which is partially redundant when already knowing the primary energy mix (shown in the blue indicators), but we include it to provide a quick overview of specifically the electricity sector. To address the high inter-model variation on hydrogen[28], Ed_5 quantifies the fraction of hydrogen in the final energy mix across scenarios.

The fifth and final dimension of model typology concerns mitigation *costs and effort* (purple). Policy costs in energy system models can be calculated in different ways, such as consumption loss and the area under the marginal abatement cost curve—we use whatever is available for each model. The first indicator in this dimension concerns the costs per marginal abatement value (C_1), which captures the ratio between cumulative additional policy costs (being the scenario costs minus that in the current policies) and the value of emissions reduction; if this number is high, the abatement was expensive. Note that this metric is related to (a cumulative version of) the cost per abatement value metric in earlier work[8,19] and is computed as follows:

$$C_1(s) = \frac{\text{Cumulative additional policy costs between 2020 and 2050}}{\overline{\text{CP}} \cdot (E(s,2020) - E(s,2050))} \quad (4)$$

where $\overline{\text{CP}}$ is the average carbon price over the period 2020–2050, and $E(s,t)$ is the CO_2 emissions in scenario s and time t. The second indicator is the transformation index (C_2), first proposed by Kriegler et al. (2015)[19], which is the sum of the absolute changes in the fractions of the primary energy carriers in 2050 (compared to the values in the period 2017–2021). Similarly, C_3 quantifies the changes in final energy demand across sectors.

Model typology table

Table 1 contains a quantification of typical model behaviour in terms of how it deviates from the ensemble. This 'deviation metric' is calculated as follows, for each indicator i and model m:

$$\text{Deviation metric}\,(i,m) = \frac{\eta_m - \langle\eta\rangle}{\langle\sigma\rangle} \quad (5)$$

where η denotes the median and brackets (<>) indicate the full ensemble. In other words, we compute the difference of the model's median with the ensemble's median and express this in the number of standard deviations. In a few columns of Table 1, we combine indicators for readability (Supplementary Table 5 provides the more detailed version of this table). For scenario responsiveness, we average the metric above for R_4 and R_5. Similarly, for electrification, we average it for Ed_1, Ed_2 and Ed_3. In the column on energy and carbon intensity and the one on energy supply, simply the highest deviation is reported.

Data availability

Processed data and a dataset containing all indicator values are publicly available on Zenodo[29] and are maintained on Github (https://github.com/MarkMDekker/IAMfingerprints). The scenario output data are available directly from the IIASA ECEMF database (https://data.ece.iiasa.ac.at/ecemf/) and on Zenodo[30].

Code availability

All analysis and plotting code is publicly available on Zenodo[29] and is maintained on Github (https://github.com/MarkMDekker/IAMfingerprints). The model comparison protocol can be accessed on Zenodo[31].

References

1. European Commission *European Green Deal* (European Union, 2019).
2. Wang, M. et al. Breaking down barriers on PV trade will facilitate global carbon mitigation. *Nat. Commun.* **12**, 6820 (2021).
3. Byers, E. et al. *AR6 Scenarios Database* (Zenodo, 2022); https://doi.org/10.5281/ZENODO.5886911
4. van Beek, L., Hajer, M., Pelzer, P., van Vuuren, D. & Cassen, C. Anticipating futures through models: the rise of integrated assessment modelling in the climate science-policy interface since 1970. *Glob. Environ. Change* **65**, 102191 (2020).
5. Dekker, M. et al. On the consensus in climate policy scenarios. Preprint at *Research Square* https://doi.org/10.21203/rs.3.rs-2073170/v1 (2022).
6. Sognnaes, I. et al. A multi-model analysis of long-term emissions and warming implications of current mitigation efforts. *Nat. Clim. Change* **11**, 1055–1062 (2021).
7. Luderer, G. et al. Impact of declining renewable energy costs on electrification in low-emission scenarios. *Nat. Energy* **7**, 32–42 (2022).
8. Harmsen, M. et al. Integrated assessment model diagnostics: key indicators and model evolution. *Environ. Res. Lett.* **16**, 054046 (2021).
9. Smith, S. J. et al. Long history of IAM comparisons. *Nat. Clim. Change* **5**, 391 (2015).
10. Blanford, G. J., Kriegler, E. & Tavoni, M. Harmonization vs. fragmentation: overview of climate policy scenarios in EMF27. *Clim. Change* **123**, 383–396 (2014).
11. Clarke, L. et al. International climate policy architectures: overview of the EMF 22 International Scenarios. *Energy Econ.* **31**, S64–S81 (2009).
12. van Vuuren, D. P. et al. The costs of achieving climate targets and the sources of uncertainty. *Nat. Clim. Change* **10**, 329–334 (2020).
13. Luderer, G. et al. Assessment of wind and solar power in global low-carbon energy scenarios: an introduction. *Energy Econ.* **64**, 542–551 (2017).
14. Tavoni, M. et al. Post-2020 climate agreements in the major economies assessed in the light of global models. *Nat. Clim. Change* **5**, 119–126 (2015).
15. Luderer, G. et al. Residual fossil CO_2 emissions in 1.5–2 °C pathways. *Nat. Clim. Change* **8**, 626–633 (2018).
16. Flato, G. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) 741–866 (Cambridge Univ. Press, 2013).
17. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
18. Andrews, T., Gregory, J. M., Webb, M. J. & Taylor, K. E. Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.* **39**, L09712 (2012).
19. Kriegler, E. et al. Diagnostic indicators for integrated assessment models of climate policy. *Technol. Forecast. Soc. Change* **90**, 45–61 (2015).
20. PBL Netherlands Environmental Assessment Agency *IMAGE Documentation* (2022); https://models.pbl.nl/image/index.php/Welcome_to_IMAGE_3.2_Documentation

21. Baumstark, L. et al. REMIND2.1: transformation and innovation dynamics of the energy-economic system within climate and sustainability limits. *Geosci. Model Dev.* **14**, 6571–6603 (2021).
22. E3Modelling *PRIMES Documentation* (2018); https://e3modelling.com/modelling-tools/primes
23. Daioglou, V. et al. Bioenergy technologies in long-run climate change mitigation: results from the EMF-33 study. *Clim. Change* **163**, 1603–1620 (2020).
24. Kann, A. & Weyant, J. P. Approaches for performing uncertainty analysis in large-scale energy/economic policy models. *Environ. Model. Assess.* **5**, 29–46 (2000).
25. Lempert, R. J. A new decision sciences for complex systems. *Proc. Natl Acad. Sci. USA* **99**, 7309–7313 (2002).
26. Usher, W., Barnes, T., Moksnes, N. & Niet, T. Global sensitivity analysis to enhance the transparency and rigour of energy system optimisation modelling. Preprint at https://open-research-europe.ec.europa.eu/articles/3-30 (2023).
27. Krey, V. et al. Looking under the hood: a comparison of techno-economic assumptions across national and global integrated assessment models. *Energy* **172**, 1254–1267 (2019).
28. Henke, H. et al. The mutual benefits of comparing energy system models and integrated assessment models. Preprint at https://doi.org/10.12688/openreseurope.15590.1 (2023).
29. Dekker, M. M. Code for: identifying energy model fingerprints in mitigation scenarios. Zenodo https://doi.org/10.5281/zenodo.8220166 (2023).
30. Pietzcker, R. et al. ECEMF diagnostic scenarios. Zenodo https://doi.org/10.5281/zenodo.7634844 (2023).
31. ECEMF. Model comparison protocol (2.2). Zenodo https://doi.org/10.5281/zenodo.6811317 (2022).
32. IPCC *Climate Change 2022: Mitigation of Climate Change* (eds Shukla, P. R. et al.) (Cambridge Univ. Press, 2022).
33. Mandley, S. J., Daioglou, V., Junginger, H. M., van Vuuren, D. P. & Wicke, B. EU bioenergy development to 2050. *Renew. Sustain. Energy Rev.* **127**, 109858 (2020).

Acknowledgements

This work was supported by the European Climate and Energy Modelling Forum (ECEMF, H2020 grant agreement number 101022622). We acknowledge and thank H. Henke, E. Fejzic, M. Lewarski and I. Tatarewicz for their input throughout the project and useful comments on the manuscript. This work was also supported by the Exploring National and Global Actions to reduce Greenhouse gas Emissions (ENGAGE, H2020 grant agreement number 821471).

Author contributions

M.M.D., V.D., R.P., R.R., H.-S.d.B., F.D.L., L.D., J.E., A.F., T.F., P.F., O.F., E.G., M.H., D.H., M.K., V.K., F.L., G.L., S.P., I.T., B.Z., B.v.d.Z., W.U. and D.v.V. conceived the study and diagnostic experiment and contributed to the scenario runs and to the writing of the manuscript. M.M.D., V.D., D.v.V., R.P. and R.R. analysed the model output. M.M.D. devised the framework, performed the analysis, generated the figures and wrote the first draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41560-023-01399-1.

Correspondence and requests for materials should be addressed to Mark M. Dekker.

Peer review information *Nature Energy* thanks Alaa Al Khourdajie, Yang Ou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

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

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

© The Author(s) 2023