

A multi-model analysis of hydrogen uptake in a climate-neutral Europe

Corresponding Author: Dr Amir Fattahi

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The paper applies an ensemble approach to scenario analysis to assess the dynamics of hydrogen in Europe under various decarbonization scenarios. The intention is to identify robust insights that can guide policymakers as they prioritize different options for achieving decarbonization goals. The novelty of the approach is to compare the results from 8 different integrated assessment models (IAMs). The authors “harmonize” the analyses by defining 5 consistent scenarios, which each model is then used to analyze. This approach has potential to generate significant new insights, but, as currently written, the paper is primarily a narrative description of the outcomes with very limited deeper analysis. As such, I cannot support publication in its current form.

-

Significance:

IAMs are a useful tool for exploring trade-offs in energy system evolution, and the collaboration across leading groups working on such models is positive and noteworthy. Since each model is configured differently, comparing the outcomes across an ensemble is a way of doing a “super” sensitivity analysis to uncover the most robust trends, as well as provide insight key trade-offs that are most sensitive to model definition. The robust outcomes where the models are in agreement is the primary focus of the paper. This agreement reinforces confidence in the conclusions that have already been reported for individual models (viz., that H2 adoption generally occurs later and in hard-to-decarbonize/hard-to-electrify sectors), but this alone does not rise to the level of significance needed for the journal.

The ensemble approach provides a path to address the question “where do the models give different answers and why?”. In my opinion, this is the more interesting and significant issue, precisely because it helps reduce the uncertainty by clarifying which trade-offs are most sensitive to modeling assumptions and hence might be most influenceable by policy or market impacts. This is flagged in observations 2 and 3 in the introduction (lines 113-124), but there is potential to more clearly and systematically assess the impacts of model configuration on divergent outcomes.

Methods and results:

1. The study provides an overview of the IAMs in Table 1. Given the goal of the study is to understand the outputs of the models, further discussion of the specific differences across the models (either in the Methods or as Supplementary Information) would be useful. When done over a range of the “dissenting” cases, this should give a clearer picture of what modeling elements and assumptions are really driving the results. This level of insight could be achieved with sensitivity analyses of each individual model, but the ensemble approach could uncover a higher level of sensitivity across modeling approaches (constraints, objective function definition, optimization methodology) that would offer a different and valuable insight beyond work with a single model.

Specifically,

- Geographic scope ... are there meaningful features that global models include (e.g., import export, as noted in Obs 3, lines 121-124) that EU models do not, and how might these impact the results?
- Approach ... how does the solver approach impact how the trade-offs are assessed?
- Sector resolution ... are there models that have higher or lower resolution in how H2 and other paths are treated in

specific sectors that might indicate their outcomes are more rigorous than other models that have lower resolution?

- Temporal resolution ... there seems to be an inconsistency in the description across models – please report both the time-step size and the time horizons for the models. Please also describe which models assume perfect foresight and which are myopic in converging on decarbonization targets.

2. Economic assumptions about the cost and cost trajectories of both H2 and its alternatives are key drivers in the trade-off analysis, it is necessary to also show and discuss these embedded assumptions. For example, a table that lists the assumptions for electrolyzer cost and future trajectory assumptions would be very valuable for a reader to quickly compare how this parameter is treated by each model. Similarly, a Table with cost assumptions about fuel cell vehicle and infrastructure costs versus battery electric vehicles would allow a reader to quickly assess the relative optimism and pessimism across the IAMs that can help explain why models show different levels of H2 adoption in transport. This is flagged in Lines 248-250, but not discussed.

In addition, given the current experience in Europe related to the buildout of distribution infrastructure, it is valuable to also summarize how the models handle hydrogen pipelines or grid/transmission lines in their assessments.

3. Table 2 lists the harmonized scenarios. It would be useful for a general readership to see how these parameters translate into actual inputs into the model. The cited references are useful, but since this is one of the key contributions of the paper, a more detailed description should be included.

The third bullet under the Net Zero entry is missing part of the sentence.

4. As noted above, the ensemble approach offers a significant opportunity to explore cases where one model deviates from the “consensus” of the others, and identify what aspect of the approach or core assumptions is responsible for the difference. This would add valuable insight into how the configuration of models impacts the insights. Specifically, areas of divergence (where models disagree) could point towards differences in underlying assumptions or model mechanics (objective function, solver approach, resolution) which can: (1) advance methodological efforts; (2) identify underlying assumptions where additional attention can help reduce uncertainty; and (3) flag areas where policy interventions might have the highest leverage in driving different outcomes.

Some specific areas from the presented results include:

- Figure 1 ... Is direct use in MEESA comparable to industrial use in other models? Is there any significance to the result that the model with highest temporal resolution (EURO-CALLIOPE) does not allocate any H2 for electricity? Is this a result of a constraint or the temporal resolution? H2 for transport only appears in meaningful levels in PRIMES and TEAM-ECN, is this because others favor EVs or synfuels? What are the underlying assumptions that drive these conclusions (cost of H2 FCV vs EVs, infrastructure, etc)?

- Figure 2 ... The results are useful, but more rigor is needed to interpret what is driving the different outcomes. Why does PRIMES predict significant H2 use for transport in NetZero, but the other models don't, and how might this be related to the underlying assumptions about EV vs H2 in contrast to Prometheus and Euro-Calliope? For H2 use in Industry, what assumptions in IMAGE and TIAM-ECN are favoring direct electrification and gases over H2, respectively?

- Figure 3 ... Please explain the grey “regression fit” in more detail. The figure shows several model results are outside the “95%” confidence interval, which can be confusing to a general reader.

- Figure 4 ... Could sector-level resolution versions of these plots be created? The results of the models indicate that H2 will tend to be deployed for hard-to-electrify sectors from 2040, so plots that show the uncertainty across models for industry and transport could be particularly helpful in visualizing this point.

- Figure SI4 is really interesting, and could be moved to the main text to provides the content for an extended discussion.

Additional comments:

1. Total system cost

The total system cost can also provide useful information beyond the relative shares of different technologies. For example, under the 3 constrained scenarios, do some models indicate a particular constraint is more impactful than others? Differences in the magnitude along the economic axis could generate additional insight from the ensemble approach that would be a significant contribution beyond the results from individual models.

Since these models optimize on an economic objective function, it should not be too difficult to include this information. Normalized costs for each scenario could be used if absolute values are not the most convenient basis.

2. Policy

The paper affirms previous insights obtained from individual IAM studies and other work (Lines 274-284). A deeper analysis, as suggested above, may produce further insight as to how policy makers might direct resources to alter the underlying dynamics, or confirm that the resources needed exceed policy ambitions to a degree that new expectation must be set.

Reviewer #2

(Remarks to the Author)

This manuscript presents the results of a model comparison study focussing on the role of hydrogen in Europe towards 2050. While many papers have been written on this particular topic, the novelty of the present manuscript is the use of a specially designed set of consistent scenarios across the different models in an effort to harmonize boundary conditions; the idea is that variations in results between the different models but within the same outer scenario (e.g. net zero emissions by 2050) are representative of uncertainty due to variations in costs, technological parameters as well as model formulation across the models.

The paper is relatively well-written (though some of the text, especially the discussion section, could be a bit more concise), and the analysis seems to be technically sound and correct; I think the conclusions are supported by the models and data that are used.

I have the impression that the manuscript does not exploit the available model data to a fully satisfactory level. The main impression I have after reading it is that there is a lot of variation in hydrogen deployment across different models --- this is, however, already clear from previous studies. The authors spend quite a bit of space on the conclusion that electrification is the number 1 strategy for decarbonisation in Europe, and that hydrogen uptake, especially direct use, will play only a relatively minor role, and that only well after 2030 (missing the EU target of 10 + 10 Mt by 2030). This conclusion, I would argue, is not particularly novel either, and basically well-known in the literature.

I was hoping for more insights that could have only been derived from this kind of harmonized comparison study. If there are large differences between models in particular sectors despite a similar "outer" scenario, why is that? Which particular differences in parameters, assumptions or modelling methodology are plausibly to blame? Moreover, the manuscript did not derive much or any insight from the full set of NetZero scenarios (including the ones with limited biomass, CCS and nuclear capacity), which is a shame as such scenarios are quite relevant.

In some specific instances, I think the manuscript overstates an alleged gap in previous research on European hydrogen, but I have to disclose that I have written on this specific topic myself.

** Introduction

The introduction is quite well-written, and sets up the context nicely. It is concise and to the point, which is great. I agree there is a lot of uncertainty and divergent perspectives around the future of hydrogen in Europe! Towards the end of the introduction, however, I think you can do a little bit of a better job making the research gap really concrete. One contention I have (on line 74) is that I think there is actually a pretty decent set of modelling studies in the literature that have full representation of hydrogen as an intermediate carrier with sector coupling etc.; several of the studies you cite do, and I could also refer you to the references in Table S1 in ref. 11 (full disclosure: I'm one of the authors of ref. 11, sorry for any self-references!) which point to a number of studies based on models such as JRC-EU-TIMES, MIRET-EU / Integrate Europe, Balmorel and Enertile. If I remember correctly, some of those have decent coverage of sector-coupling around hydrogen. Moreover, ref. 11 (sorry!), while being sector-coupled etc., also explores which factors (technological, policy-wise etc) are most significant in explaining the wide variation we see in cost-optimal scenarios.

In summary, as I see it, main novelty of the present manuscript, and the thing which I do think closes a research gap, is presenting a model comparison analysis with a focus on hydrogen and a harmonized set of scenarios. But be careful not to overstate the research gap.

** Results

For Figure 1, I'm not quite sure if a Sankey diagram is the right way of visualising the results, as the structure of the data is quite simple and doesn't call for this kind of diagram. An obvious alternative would be two panels (for inputs and outputs) with a set of stacked bars (one for each model) with the bars showing total hydrogen, broken down by category. It's not a critical flaw but you could consider this or other alternative presentations. (Also, I find it a little difficult to distinguish between the colours for direct uses, exports and losses.)

In the first subsection of the results section, I'm maybe missing a bit of explanation or insight. I understand that the results across the models vary a lot, maybe especially w.r.t. synthetic fuel production and "Industry" (it would be nice to briefly define what exactly is included in this sector). But what drives these variations, specifically? Why does PRIMES, for example, need a lot of hydrogen-derived synthetic fuel and the other models do not? How does this relate to other findings in the literature about what drives demand for hydrogen / synthetic fuel? Could this have to do with assumptions on the transportation sector, or is it something else? I think the reader would benefit from some more in-depth insights, even if only in reference to later sections.

I think the second subsection is mostly fine. I do wonder, however: could this subsection and the above be enriched by pulling in comparisons with the variations on NetZero scenarios that you haven't used yet so far? This could illuminate some of data. For example, I am left wondering "Why does PRIMES show such a high fraction of hydrogen in transportation" and, indeed, "why does MESSAGEix-GLOBIOM show such a high fraction of gas in transportation"?

In the third subsection, I appreciate seeing results from some of the different scenarios, although I think it's a shame that no conclusions are drawn from any comparison between LimBio, LimCC, LimNuc etc.; the reader is left to search the figure for a

pattern. Is there not much difference between these scenarios, or is the difference highly heterogeneous across models? That would be a finding worth stressing in itself! In that case, biomass, CCS & nuclear have been eliminated as unambiguous drivers of hydrogen uptake (though I doubt this, especially the former two).

For Fig 4, I fail to see the significance of the coloured heatmap interpolating CO₂ emissions. The text also doesn't bring this up at all. I guess in the second panel we can see a sort of correlation between electricity share and emissions, as expected. In the third panel, the heatmap only serves to show the sharp divide between NPI and net zero scenarios, which isn't new information. (It's obvious that the NetZero scenarios are going to have low emissions!) Moreover, I think the triangular interpolation is distracting at best; a more standard technique would be to fit some kind of smooth function to your data, like a regression, and plot that, to separate the pattern from the noise.

** Discussion

I'm sorry to bring this up again, but a number of the points you bring up in the first paragraph of this section have been addressed in ref. 11. There, we show explicitly how liberal use of CCS and/or biomass can entirely eliminate indirect uses of hydrogen; conversely, how limited alternatives can drive hydrogen production and demand upwards. That study also includes a global sensitivity analysis encompassing a variety of techno-economic parameters. More work could certainly be done in this area, but you should be more specific about the alleged gaps.

The link to the Zenodo repository with data and code doesn't seem to work, which is a problem w.r.t. to reproducibility.

** Methods

It would be nice to know a little bit about how the included models were selected; on which basis or selection criteria.

** Minor comments

- At least two citations are a little outdated (10, 11); the manuscript references preprints that have since been published (in ERL and Joule resp.)
- I could not accurately find the source for ref. 22.
- I'm not sure if I understand the "Heat" row under Buildings in Fig 2: presumably much of the electricity and gas demand in the buildings sector is already used for heating, so why is Heat also reported as a separate demand?
- For Fig 3, I somehow find it hard to believe that the grey region represents a 95% confidence interval if almost half the data points fall outside this range? Anyway, maybe this dataset is too small to derive a confidence interval like this?
- "not homogeneous" -> "heterogeneous".
- Line 266: "NetZero" -> "net zero" (describing the general policy, not the study-specific scenario?)
- Line 288: What does "policy momentum" mean? Anyway, one could argue that, whatever the official policy is, Europe as a whole is de facto following an electrification pathway already, so this is not much of a policy recommendation but rather a description of the current state of affairs.
- Ref. 35 is not properly cited.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

In my assessment, the response to reviewer comments and revised manuscript satisfactorily address the points raised in the initial review. I am in support of publication.

Two minor points for author consideration:

1. Not all of the acronyms are defined at their first use (e.g., NPI), and it might be useful to do a pass through the manuscript to correct this.
2. The exclusion of hydrogen imports/exports from the modeling sidesteps an interesting set of questions around Europe as a net energy importer (viz., how might the results change if low cost H₂ were to become available from outside Europe?). I recognize this is clearly out of scope for the current manuscript, but given the authors general interest in modeling of how energy systems evolve, it might be worth exploring separately some aspects of this issue.

Reviewer #2

(Remarks to the Author)

Thank you for the work you put into the revisions. I feel like all my points have been addressed, most of them in a satisfactory way.

I still have a minor concern which I would like to raise.

Plainly, PRIMES is an outlier when it comes to hydrogen use in transportation and for synthetic fuel production. This is a surprising result (compared to the other models in the study) and pretty much the one data point that stands out the most in the first few figures. Both I and the other reviewer asked in our initial reviews for more discussion of the differences between models, and I think the revised version of the manuscript delivers on this to an extent. But after reading the revised manuscript, I'm still left wondering: "Why does PRIMES have such a high share of hydrogen in transportation, and why does

it produce so much synthetic fuel?". In the manuscript, we read: "In PRIMES, higher hydrogen shares reflect model-specific allocation to harder-to-electrify transport segments". To me, this doesn't really answer the question, it just states the surprising data point. It's the why that I'm interested in. Does PRIMES exogenously specify a high hydrogen share in certain transportation sectors? (That would weaken the result, but would be good to know.) Does it severely limit the use of BEVs? Does it somehow arrive at very low hydrogen prices and conclude that hydrogen in the transportation sector is more cost-effective than direct electrification? (That would be very surprising but interesting.) Does this rather reflect differences in scope in the transportation sector? Does PRIMES assume that a lot of shipping will run on hydrogen directly?

The reason for large-scale use of hydrogen for synthetic fuel production isn't discussed much in detail either as far as I can see.

The manuscript is already quite long and I wouldn't want you to expand the text with additional paragraphs upon paragraphs of lengthy explanation and investigation. But I think even brief references to the "why"s of differences in results would make the article much more interesting and satisfying to read.

I mention PRIMES specifically here since it's an obvious outlier, but the same could be said about other differences.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

All of my substantive concerns were satisfactorily addressed in the first revision.

The second revision addresses some minor points, and my view is that this manuscript is now suitable for publication.

Reviewer #2

(Remarks to the Author)

I appreciate my concerns being addressed w.r.t. a deeper explanation of the higher uptake of hydrogen in PRIMES specifically. I understand the dynamics a little better now (PRIMES having a higher "resolution" on the transportation sector with long haul freight etc.), though a full understanding would still require a deeper investigation. I realise, however, that this may be considered out of scope for the present article. The current version of the article is fit for publication.

Open Access This Peer Review File 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.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File 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 <https://creativecommons.org/licenses/by/4.0/>

Reviewer #1 (Remarks to the Author):

The paper applies an ensemble approach to scenario analysis to assess the dynamics of hydrogen in Europe under various decarbonization scenarios. The intention is to identify robust insights that can guide policymakers as they prioritize different options for achieving decarbonization goals. The novelty of the approach is to compare the results from 8 different integrated assessment models (IAMs). The authors “harmonize” the analyses by defining 5 consistent scenarios, which each model is then used to analyze. This approach has potential to generate significant new insights, but, as currently written, the paper is primarily a narrative description of the outcomes with very limited deeper analysis. As such, I cannot support publication in its current form.

Response: We thank the reviewer for the time and care invested in reviewing the manuscript. We appreciate the constructive and detailed nature of your comments, which helped us substantially improve both the analytical depth and the clarity of the paper. In the revised manuscript, we strengthened the contribution in three main ways. First, we expanded the cross-model comparison to more clearly separate robust signals from areas of divergence across sectors and time. Second, we conducted a structured survey across the participating modelling teams, reported the full results as a table in the Supplementary Information, and used these model-characteristics inputs explicitly to support interpretation in the Results and Discussion. Third, we improved transparency and readability by expanding and clarifying the Methods and Supplementary Information, including adding a new supplementary figure that further elaborates the main results, and moving one previously supplementary figure into the main text to improve clarity. Altogether, the manuscript has been substantially revised, particularly in the Results, Discussion, Methods, and Supplementary Information. Below, we respond point-by-point and indicate where changes were made (additions in purple; removed text struck through).

Significance:

IAMs are a useful tool for exploring trade-offs in energy system evolution, and the collaboration across leading groups working on such models is positive and noteworthy. Since each model is configured differently, comparing the outcomes across an ensemble is a way of doing a “super” sensitivity analysis to uncover the most robust trends, as well as provide insight key trade-offs that are most sensitive to model definition. The robust outcomes where the models are in agreement is the primary focus of the paper. This agreement reinforces confidence in the conclusions that have already been reported for individual models (viz., that H2 adoption generally occurs later and in hard-to-decarbonize/hard-to-electrify sectors), but this alone does not rise to the level of significance needed for the journal.

The ensemble approach provides a path to address the question “where do the models give different answers and why?”. In my opinion, this is the more interesting and significant issue, precisely because it helps reduce the uncertainty by clarifying which trade-offs are most sensitive to modeling assumptions and hence might be most influenceable by policy or market impacts. This is flagged in observations 2 and 3 in the introduction (lines 113-124), but there is potential to more clearly and systematically assess the impacts of model configuration on divergent outcomes.

Response: We agree with the reviewer that systematically examining where models provide different answers (and why) is one of the most interesting and policy-relevant outcomes of a multi-model ensemble, because it helps identify which trade-offs are most sensitive to modelling assumptions and structural choices. At the same time, the aim of our paper is to leverage the intercomparison to distinguish *more robust pathways* (where results converge across models) from *more fragile model deductions* (where results diverge), thereby clarifying which insights can be considered consistent across modelling approaches and which require greater caution in interpretation. In response to this comment, and to better exploit the ensemble nature of the dataset, we expanded the discussion of both robust and less-robust findings in multiple parts of the revised manuscript, with an increased focus on highlighting key areas of divergence and providing additional interpretation drivers behind them.

Methods and results:

1. The study provides an overview of the IAMs in Table 1. Given the goal of the study is to understand the outputs of the models, further discussion of the specific differences across the models (either in the Methods or as Supplementary Information) would be useful. When done over a range of the “dissenting” cases, this should give a clearer picture of what modeling elements and assumptions are really driving the results. This level of insight could be achieved with sensitivity analyses of each individual model, but the ensemble approach could uncover a higher level of sensitivity across modeling approaches (constraints, objective function definition, optimization methodology) that would offer a different and valuable insight beyond work with a single model.

Response: In response to this comment, we conducted a structured interview/questionnaire with all modelling teams participating in the intercomparison to systematically document key model characteristics and differences (including, where applicable, constraints, objective-function formulation, optimisation/solution approach, and sectoral/temporal/geographical resolution). The comprehensive results of this questionnaire are reported in the Supplementary Information. We also used these inputs to expand and sharpen the Results and Discussion in the revised manuscript, in order to better address the points raised below and to provide additional context for interpreting dissenting outcomes across models.

Specifically,

- Geographic scope ... are there meaningful features that global models include (e.g., import export, as noted in Obs 3, lines 121-124) that EU models do not, and how might these impact the results?
- Approach ... how does the solver approach impact how the trade-offs are assessed?
- Sector resolution ... are there models that have higher or lower resolution in how H2 and other paths are treated in specific sectors that might indicate their outcomes are more rigorous than other models that have lower resolution?
- Temporal resolution ... there seems to be an inconsistency in the description across models – please report both the time-step size and the time horizons for the models. Please also describe which models assume perfect foresight and which are myopic in converging on decarbonization targets.

Response: The model survey (reported in the Supplementary Information, Table SI 1) enabled a more systematic comparison of model structure and key assumptions, which we now use explicitly to interpret dissenting outcomes.

- *Geographic scope.* MEESA and Euro-Calliope are European models (EU27/UK/EEA), while PROMETHEUS, IMAGE, MESSAGEix-GLOBIOM, TIAM-ECN and REMIND are global; PRIMES has Europe/EU focus (Table 1). EU spatial detail differs: Euro-Calliope resolves countries; MEESA uses multi-country regions (~1–4 EU countries/region); and global models typically represent the EU as 2–3 subregions (or aggregated multi-country regions). This affects whether extra-European market feedbacks and cross-region interactions (including hydrogen trade opportunities) are endogenous (global models) or reflected through fixed boundary conditions (European-only/EU-only models).

- *Approach/solver.* MEESA, MESSAGEix-GLOBIOM, TIAM-ECN and Euro-Calliope are optimisation models (investment/operation). PROMETHEUS is a recursive simulation model; IMAGE is an integrated assessment simulation model; REMIND is general equilibrium; PRIMES is partial equilibrium (Table 1). These paradigms influence trade-offs (least-cost optimisation vs simulated dynamics and market adjustment), and we therefore clarified these categories in Table 1/Supplementary Information and use them in the Discussion of divergent outcomes.

- *Sector resolution.* The survey clarifies differences in sectoral coverage and hydrogen end-use representation. Several models include a broad set of end-use sectors (buildings, industry incl. feedstocks, transport incl. aviation/navigation, power and heat). Euro-Calliope (in the version used here) does not represent the disaggregation of the industry sectors, but it aggregates the demands according to four different commodities: methane, methanol, diesel, and kerosene. Each energy carrier can be produced via fossil fuels, biomass, or hydrogen (e-fuels). In this regard, the hydrogen used to produce e-fuels is subject to cost minimization and carbon constraints in the model. Hydrogen CHP technology is considered for combined heat and power provision. On top of the endogenous demand for hydrogen, Euro-Calliope model does consider an exogenous demand of hydrogen that is estimated according to the levels of hydrogen usage excluding refinery needs. MEESA also treats a substantial share of hydrogen demand exogenously, based on harmonised ECEMF sectoral inputs, and reports this as aggregate direct use rather than a directly comparable industrial category. We now refer to these differences more explicitly when interpreting sectoral divergence, including PRIMES' high transport hydrogen share, Euro-Calliope's limited hydrogen use in power, and cross-model differences in industry.

- *Temporal resolution and foresight.* The survey clarifies both modelling horizon and time-step size, as well as planning foresight and sub-annual detail. MEESA, PROMETHEUS and Euro-Calliope run to ~2050, while IMAGE, MESSAGEix-GLOBIOM, TIAM-ECN and REMIND extend to ~2100. Time steps range from yearly (PROMETHEUS, IMAGE) to 2–5-yearly (MEESA, MESSAGEix-GLOBIOM, TIAM-ECN, REMIND). Euro-Calliope uses high temporal resolution, i.e., 3-hourly operation for a given scenario year (Table 1). We report these characteristics more consistently in Table 1 and the Supplementary Information, and use them to clarify how operational flexibility, storage needs, and power-sector balancing can influence hydrogen uptake. For example, Euro-

Calliope's limited use of hydrogen in the power sector reflects the availability of lower-cost flexibility options, such as batteries, flexible electric-vehicle charging, and dispatchable biomass-, waste-, and biogas-based generation, rather than temporal resolution alone.

Manuscript text changes (revised text in purple; strike-through indicates removed phrasing in the revised text):

Revised text: "The third observation is the limited representation of hydrogen trade in these models. Only IMAGE and REMIND report non-zero imports in 2050 (≈ 0.2 EJ, ≈ 1.4 Mth_{H₂}, HHV). This is consistent with model scope and trade implementation: IMAGE can endogenously draw on global trade options, while REMIND includes a stylised import option that becomes attractive when domestic production is relatively costly. In contrast, European-only/EU-focused models (Euro-Calliope, PRIMES, MEESA) do not endogenously represent extra-European market feedbacks, and several other models do not report explicit hydrogen trade flows in the harmonised outputs. ~~Only IMAGE and REMIND incorporate hydrogen imports, both reporting low trade volumes of approximately 0.2 EJ, equivalent to an estimated 1.4 Mt hydrogen (higher heating value).~~ Furthermore, none of the models project hydrogen exports."

Revised text: "To represent this diversity, IAMs generally employ aggregated conversion efficiencies and exogenous cost inputs, whereas sector-coupled and multi-vector optimisation frameworks use detailed process-level simulations with varying spatio-temporal resolution, technology scope and infrastructure representation, which affects hydrogen's role in power-system balancing (via operational constraints) and sectoral allocation when demand is endogenous vs. exogenous. For example, Euro-Calliope represents multiple sectors and broad cross-sector competition, but with a more aggregated representation of industry and with part of hydrogen demand specified exogenously. In this framework, hydrogen is used endogenously for synthetic-fuel production and transport choices, but its role in the power sector remains limited because lower-cost flexibility options are available. These include battery storage, dispatchable generation technologies such as CHP plants fuelled by waste, biofuels, and biogas, and demand-side flexibility from electric vehicles. By contrast, PRIMES' more detailed end-use representation allows hydrogen to compete more directly with electrification and synthetic fuels within sectors, contributing to its higher hydrogen share in transport."

2. Economic assumptions about the cost and cost trajectories of both H₂ and its alternatives are key drivers in the trade-off analysis, it is necessary to also show and discuss these embedded assumptions. For example, a table that lists the assumptions for electrolyzer cost and future trajectory assumptions would be very valuable for a reader to quickly compare how this parameter is treated by each model. Similarly, a Table with cost assumptions about fuel cell vehicle and infrastructure costs versus battery electric vehicles would allow a reader to quickly assess the relative optimism and pessimism across the IAMs that can help explain why models show different levels of H₂ adoption in transport. This is flagged in Lines 248-250, but not discussed.

Response: We agree that the cost and cost trajectories of hydrogen technologies (e.g., electrolyzers) and key competing options (e.g., BEVs vs. FCVs and related infrastructure) are important drivers of the trade-offs in our models and can help explain differences in hydrogen uptake across the ensemble. However, providing a single cross-model table of techno-economic assumptions that is both comprehensive and comparable is beyond the scope of this paper for three reasons: (a) a meaningful comparison would need to cover the full

set of competing technologies and energy carriers in each sector (for transport alone this includes, at minimum, electricity/BEVs, hydrogen/FCVs, fossil liquids, biofuels, and e-fuels, each with vehicles and infrastructure), not only hydrogen production; (b) models differ in technology definitions and granularity (technology variants, cost components, and learning formulations), so “like-for-like” cost entries are often not directly comparable; and (c) outcomes are not driven by costs alone, but also by non-cost constraints and parameters (e.g., diffusion limits, resource potentials, and binding scenario constraints), which can dominate technology choices even under similar cost assumptions. We therefore clarified the text around Lines 248–250 to explicitly acknowledge this and to explain that each model determines hydrogen uptake through its endogenous system-wide optimisation/market-equilibrium mechanism using internally consistent techno-economic databases and constraints. Detailed technology-cost assumptions are documented in the model/ECEMF references already cited in the manuscript (including studies that report electrolyser cost trajectories across models), and we refer interested readers to those sources for model-specific input values.

Revised text: “Hydrogen demand as a means to achieve NetZero targets varies widely, from negligible uptake to over 11 EJ (~77 MtH₂) in 2050. This spread reflects differences in (i) key techno-economic assumptions (e.g., electrolyser/CCS costs and learning, and BEV-FCV cost competitiveness), (ii) model structure (sector/process detail, substitution options, and spatio-temporal resolution), and (iii) non-cost constraints and parameters (e.g., diffusion limits and resource potentials) that can bind even when costs are similar. This heterogeneity is most pronounced in industry and transport, where differences in process-level technology representation can materially affect substitution pathways and hydrogen’s competitiveness versus electrification or synthetic fuels. ~~This variability reflects differences in model assumptions and the diversity of substitution pathways, particularly in the industrial and transport sectors where the representation of process-level technologies remains limited or not homogeneous across models. This divergence is expected to diminish as our models further develop.~~ We recommend harmonizing input parameters in future studies to enhance comparability and improving the granularity of industrial sector modelling to better capture technology-specific substitution options. Moreover, pronounced modelling uncertainty persists regarding the magnitude and timing of hydrogen deployment. Across models, early decarbonisation is consistently driven by electrification, with hydrogen expansion occurring mainly after 2040 as hard-to-abate sectors come into focus. Some models—particularly those envisaging limited CCS, constrained biomass, or optimistic technology cost reductions—demonstrate only moderate hydrogen uptake. This suggests that low hydrogen deployment can be offset by alternative mitigation strategies, such as greater reliance on CCS, BECCS, or sustained technological learning. Others, such as PRIMES, project relatively high hydrogen shares in transport, yet these are exceptions rather than the rule.”

In addition, given the current experience in Europe related to the buildout of distribution infrastructure, it is valuable to also summarize how the models handle hydrogen pipelines or grid/transmission lines in their assessments.

Response: We agree and have clarified how infrastructure is represented across the participating models, based on the model survey and Table 1. Hydrogen transport and distribution (T&D) is treated heterogeneously: (i) some models include explicit pipeline options (e.g., cross-border and/or distribution pipelines in PROMETHEUS and MESSAGEix-GLOBIOM); (ii) others represent hydrogen T&D via

aggregated cost proxies rather than explicit network expansion (e.g., IMAGE uses within-region distribution costs and distance-dependent inter-regional transport costs; REMIND builds a stylised T&D network with parameterised costs); and (iii) some models do not model hydrogen networks explicitly in the version used here (e.g., MEESA, TIAM-ECN, Euro-Calliope), implying that hydrogen uptake reflects production/end-use competition under simplified or omitted infrastructure representations. Electricity grids are likewise represented with varying detail: Euro-Calliope has high spatial (country-level) and temporal (3-hourly) resolution and explicitly captures electricity system operation and interconnection constraints, while most global IAMs use more aggregated regional representations. We now refer to these differences when interpreting cross-model outcomes, particularly for hydrogen imports/exports and the extent to which hydrogen provides power-system flexibility.

Added text in Method: “Infrastructure representation differs across models and can influence both hydrogen trade and the relative attractiveness of hydrogen versus electrification. Hydrogen transport and distribution (T&D) is treated heterogeneously: (i) some models include explicit pipeline options (e.g., cross-border and/or distribution pipelines in PROMETHEUS and MESSAGEix-GLOBIOM); (ii) others represent hydrogen T&D via aggregated cost proxies rather than explicit network expansion (e.g., IMAGE uses within-region distribution costs and distance-dependent inter-regional transport costs; REMIND builds a stylised T&D network with parameterised costs); and (iii) some models do not model hydrogen networks explicitly in the version used here (e.g., MEESA, TIAM-ECN, Euro-Calliope), implying that hydrogen uptake reflects production/end-use competition under simplified or omitted infrastructure representations. Electricity grids are likewise represented with varying detail: Euro-Calliope has high spatial (country-level) and temporal (3-hourly) resolution and explicitly captures electricity system operation and interconnection constraints, while most global IAMs use more aggregated regional representations. We refer to these differences when interpreting cross-model outcomes, particularly for hydrogen imports/exports and the extent to which hydrogen provides power-system flexibility.”

3. Table 2 lists the harmonized scenarios. It would be useful for a general readership to see how these parameters translate into actual inputs into the model. The cited references are useful, but since this is one of the key contributions of the paper, a more detailed description should be included.

Response: We agree that, for a general readership, it is helpful to translate the harmonised scenario descriptors in Table 2 into the concrete policy levers and constraints through which they are implemented in the participating models. In response to this comment, we have strengthened the main text by adding a more explicit paragraph immediately before Table 2, clarifying how the NPI and NetZero scenarios are operationalised across the model ensemble. In particular, we now explain that, while the scenarios share a common harmonised logic, their implementation differs by model structure, e.g., through explicit policy protocols in IMAGE, unconstrained or weak-policy baseline settings in TIAM-ECN and REMIND for NPI, and through combinations of carbon-price trajectories, emissions constraints, and policy reinforcement in models such as IMAGE, REMIND, OSeMBE, TIAM-ECN, and WITCH under NetZero. This provides readers with a clearer sense of how the harmonised scenarios translate into actual model inputs, while keeping the main text

concise. For a more comprehensive model-by-model discussion of scenario implementation, readers are referred to the companion paper by Pietzcker et al., currently under review.

Revised text in SI:

“In this study, we analyse five thoroughly vetted scenarios developed within the ECEMF consortium, which are grouped into two main categories: the National Policies Implemented (NPI) scenario and four NetZero scenario variants. The NPI scenario represents a baseline that reflects the continuation of currently enacted national policies, excluding planned initiatives such as the EU Green Deal. In contrast, the NetZero scenarios are designed to achieve net-zero greenhouse gas emissions in Europe by 2050 and, for global models, constrain cumulative emissions to limit global warming to 1.5°C. To assess the impact of contested technologies, we examine alternative variants of the NetZero scenarios that limit the deployment of biomass (LimBio), nuclear energy (LimNuc), or carbon capture and storage (LimCCS). In practice, these harmonised scenarios are implemented through model-specific combinations of policy settings and system constraints. For example, in IMAGE, the NPI scenario follows an explicit protocol of implemented climate and energy policies from major economies, while the NetZero scenario combines this baseline with an increasing carbon price calibrated to achieve both the global cumulative CO₂ budget and the EU27 net-zero greenhouse-gas target. In REMIND and TIAM-ECN, the NPI scenario reflects the weaker-policy environment of roughly 2018 to 2021, before the EU Green Deal, with no additional binding climate ambition beyond that baseline. Under NetZero, REMIND applies EU climate targets for 2030 and 2050 together with a global cumulative CO₂ constraint, using iteratively adjusted carbon-price trajectories, while TIAM-ECN combines a globally uniform rising carbon tax with an explicit cumulative greenhouse-gas emissions budget. In OSeMBE, NetZero is represented through an exogenously specified and gradually increasing carbon price from 2025 onward, which drives the cost-optimal transition towards net-zero by 2050. In WITCH, the NetZero pathway is implemented through an emissions-cap trajectory with an implicit carbon price that rises after 2030, while maintaining existing policies at their current level. Thus, **Error! Reference source not found.** reports the common scenario logic, while the specific translation into model inputs reflects each framework’s representation of policy and system dynamics. The detailed parametrisation of all scenarios is provided in **Error! Reference source not found.**, with CO₂_emissions trajectories for each scenario depicted in **Error! Reference source not found.**. Unless otherwise specified, figures throughout this study present comparative analyses of NPI and NetZero scenario outcomes. For a more comprehensive description of scenario implementation across models, please refer to Pietzcker et al.²², currently under review.”

The third bullet under the Net Zero entry is missing part of the sentence.

Response: Thank you for spotting this. We corrected the incomplete bullet in Table 2 (NetZero scenario description) to restore the missing text.

Added text in Table 2: “The EU climate target can be achieved either preferably through an exogenous CO₂ price trajectory or an emission constraint. Try to achieve this target by combining a smooth carbon price trajectory and carbon constraints.”

4. As noted above, the ensemble approach offers a significant opportunity to explore cases where one model deviates from the “consensus” of the others, and identify what aspect of the approach or core assumptions is responsible for the difference. This would add valuable insight into how the configuration of models impacts the insights. Specifically, areas of divergence (where models disagree) could point towards differences in underlying assumptions or model mechanics (objective function, solver approach, resolution) which can: (1) advance methodological efforts; (2) identify underlying assumptions where additional attention can help reduce uncertainty; and (3) flag areas where policy interventions might have the highest leverage in driving different outcomes.

Response: We agree. To better leverage the ensemble, we now explicitly flag key areas of divergence across models and relate them to documented model characteristics (approach and objective function, foresight, and spatial and temporal resolution). We use the model questionnaire (reported in the Supplementary Information) to provide context for why certain outcomes differ, and we then address the main divergence patterns figure-by-figure in the sub-comments below.

Some specific areas from the presented results include:

- Figure 1 ... Is direct use in MEESA comparable to industrial use in other models? Is there any significance to the result that the model with highest temporal resolution (EURO-CALLIOPE) does not allocate any H2 for electricity? Is this a result of a constraint or the temporal resolution? H2 for transport only appears in meaningful levels in PRIMES and TEAM-ECN, is this because others favor EVs or synfuels? What are the underlying assumptions that drive these conclusions (cost of H2 FCV vs EVs, infrastructure, etc)?

Response: Thank you, we agree that Fig. 1 can be used more explicitly to interpret divergence. We therefore clarify in the text that (i) ‘direct use’ in MEESA is not comparable to ‘industry’ in other models because MEESA, as an energy supply and power-system oriented model, implements a substantial part of hydrogen demand exogenously based on harmonised ECEMF sectoral inputs, rather than endogenously allocating it across detailed end-use sectors. More specifically, MEESA receives an aggregated hydrogen final-demand input for industry, residential, and transport sectors, to which endogenous hydrogen demand for electricity generation is added. Hydrogen is therefore reported as aggregate ‘direct use’ rather than as a directly comparable industrial category; (ii) Euro-Calliope’s high temporal resolution alone does not imply hydrogen use for power, and in our application Euro-Calliope represents multiple sectors and allows broad cross-sector competition, but with a more aggregated representation of industry and with some hydrogen demands specified exogenously. In the scenarios analysed here, the limited role of hydrogen in the power sector primarily reflects the availability of lower-cost flexibility alternatives, including battery storage, flexible electric-vehicle charging, and dispatchable generation based on waste, biofuels, and biogas, rather than temporal resolution alone; and (iii) the higher transport hydrogen shares in PRIMES and TIAM-ECN are discussed more explicitly in Fig. 2, we clarify that PRIMES allocates more hydrogen to harder-to-electrify transport segments, while other models, including Euro-Calliope, more often favour direct electrification and, in some cases, synthetic fuels depending on sectoral structure and cost assumptions. We fully agree that tracing outcomes back to detailed techno-economic assumptions would be valuable, but a model-by-model attribution of underlying cost and behavioural assumptions is beyond the scope of this study; instead, we focus on ensemble behaviour and on messages that are robust across models despite their differences.

Revised text around Fig.1:

“Hydrogen fulfils multiple functions in energy systems—as an energy carrier, storage medium, feedstock and fuel. Hydrogen Production routes encompass electrolysis; steam-methane reforming with or without carbon capture and storage ($\pm$ CCS); and biomass gasification ($\pm$ CCS). It may be stored as compressed gas, cryogenic liquid, or in chemical carriers, and then blended with natural gas or reconverted to power, heat, or chemicals via fuel cells, combustion turbines, direct-reduction processes or synthetic-fuel synthesis. To represent this diversity, IAMs generally employ aggregated conversion efficiencies and exogenous cost inputs, whereas sector-coupled and multi-vector optimisation frameworks use detailed process-level simulations with varying spatio-temporal resolution, technology scope and infrastructure representation.

To illustrate this heterogeneity, Fig 1 presents a collection of Sankey diagrams depicting the hydrogen transformation chain, as projected by our models for 2050 in the NetZero scenario. The diagrams show how hydrogen is supplied to the European energy system and how it is subsequently utilised in the various demand sectors.

A first key observation, is the order-of-magnitude differences in absolute hydrogen deployment across models (Fig 1: Orange bars). In some cases, hydrogen emerges as a major energy carrier; in others, its contribution is peripheral. These differences are also evident in the relative hydrogen uptake shown in Fig SI 2 and Fig SI 4. These disparities primarily reflect differing assumptions regarding technology costs, resource potentials, electricity prices, and the competitiveness of alternative decarbonisation strategies, such as electrification or the use of fossil fuels with CCS.

At the same time, part of the variation in Fig. 1 reflects differences in model scope and in how hydrogen demand is represented and reported. In MEESA (an energy supply and power-system oriented optimisation model), hydrogen demand is implemented to a substantial extent through harmonised ECEMF sectoral demand inputs and is therefore not endogenously allocated across detailed end-use sectors, which is why hydrogen is reported as aggregate ‘direct use’ rather than, e.g., ‘industry’. Euro-Calliope provides high temporal resolution for electricity-system operation (chronological multi-hourly), but it also represents multiple sectors and broad cross-sector competition. However, the industrial sector is represented more aggregately and some hydrogen-related demands are exogenous. Its limited hydrogen use in the power sector therefore reflects the availability of lower-cost flexibility alternatives, rather than temporal resolution alone. The notable direct hydrogen shares in PRIMES and TIAM-ECN reflect model-specific sectoral allocation choices, especially in harder-to-electrify transport uses. In Fig. 2, we therefore disaggregate final energy use to investigate the resulting sectoral energy mix in more detail.”

- Figure 2 ... The results are useful, but more rigor is needed to interpret what is driving the different outcomes. Why does PRIMES predict significant H2 use for transport in NetZero, but the other models don't, and how might this be related to the underlying assumptions about EV vs H2 in contrast to Prometheus and Euro-Calliope? For H2 use in Industry, what assumptions in IMAGE and TIAM-ECN are favoring direct electrification and gases over H2, respectively?

Response: We agree that further interpretation of the divergent outcomes in Fig. 2 is valuable. We therefore strengthened the revised manuscript with more explicit model-specific interpretation. For transport, we now

clarify that PRIMES' higher hydrogen share is driven by model-specific allocation within harder-to-electrify transport segments, rather than by a simple one-to-one competition between hydrogen, electricity, and synthetic fuels across all transport modes. By contrast, Euro-Calliope represents a broad set of transport subsectors and allows competition between battery-electric, fuel-cell, and combustion-based options, but in the scenarios analysed here hydrogen is not economically competitive, partly because vehicle and refuelling infrastructure costs favour a stronger shift towards battery-electric transport. PROMETHEUS also retains a larger role for liquid fuels, illustrating that transport outcomes depend not only on hydrogen versus electrification costs, but also on broader model structure and sectoral substitution possibilities.

For industry, we added a more specific explanation for IMAGE. IMAGE combines a more detailed technology representation for some industrial subsectors with a more aggregated carrier-based representation for the remaining industrial energy demand. Across both parts, technologies and energy carriers compete on cost, while electrification, hydrogen, gaseous fuels, biomass, and CCS remain subject to different resource and feasibility constraints. In this framework, hydrogen competitiveness is influenced not by a single assumption, but by the broader cost-based competition among decarbonisation options, including electrification, gaseous fuels, biomass, and CCS. Biomass availability is endogenously constrained, while CCS deployment is limited by both regional storage potential and feasibility constraints.

We also clarify that a detailed, model-by-model attribution to all underlying techno-economic input assumptions, e.g., BEV versus FCV costs, infrastructure costs, and diffusion constraints, is not feasible within the scope of this ensemble study. The participating models differ fundamentally in modelling paradigm, sectoral detail, and cross-sector coupling, and the observed outcomes emerge from a large interacting set of assumptions across multiple sectors rather than a small number of parameters. Accordingly, we focus on documenting and interpreting ensemble-level patterns using the harmonised outputs and the model-characteristics survey reported in the Supplementary Information.

Revised text around Fig. 2: “Transport transitions are the most heterogeneous: REMIND, TIAM-ECN, and IMAGE project over 78% electricity, whereas other models linger around 40%; liquid fuels dominate in PROMETHEUS (59%) and Euro-Calliope (54%), and PRIMES uniquely allocates 38% to hydrogen, raising the sector mean across models to 7%. This sectoral mean is strongly influenced by the outlier behaviour of PRIMES, as other models allocate less than 6% to hydrogen, ~~reflecting divergent assumptions on decarbonizing heavy duty vehicles and domestic air and water transport.~~ Interpreting this divergence requires acknowledging that models differ in paradigm, sectoral detail, transport representation, and cross-sector coupling. In PRIMES, higher hydrogen shares reflect model-specific allocation to harder-to-electrify transport segments. In Euro-Calliope, by contrast, road transport can be met by battery-electric, fuel-cell, and combustion technologies, but hydrogen remains economically uncompetitive once vehicle and infrastructure costs are considered, leading to a predominantly battery-electric pathway. PROMETHEUS retains a comparatively high role for liquid fuels, indicating a different balance between electrification, hydrogen, and liquid-fuel substitution pathways. In industry, cross-model differences similarly reflect differences in process detail and substitution structure. In IMAGE, industrial outcomes emerge from competition among electrification, hydrogen, gaseous fuels, biomass, and CCS under cost, resource, and feasibility constraints, rather than from a single dominant substitution rule.”

- Figure 3 ... Please explain the grey “regression fit” in more detail. The figure shows several model results are outside the “95%” confidence interval, which can be confusing to a general reader.

Response: Thank you for highlighting this potential confusion. We now clarify in the figure caption and accompanying text that the grey line is a simple linear regression fit across the pooled model-year points (intended as an illustrative summary of the overall association rather than a predictive model for each individual IAM). In particular, the regression fit is used to graphically summarise where models tend to agree versus disagree on average across the ensemble. The grey shaded region denotes the 95% confidence interval around the fitted mean relationship. Individual points can fall outside this band, especially in small and heterogeneous datasets such as ours, because the band reflects uncertainty in the estimated mean trend rather than a 95% prediction interval for individual observations. To avoid misinterpretation, we revised the caption to explicitly state that the shading is a confidence interval for the mean fit, and we additionally report that a prediction interval would be wider.

Revised caption: “Scatter plots illustrate the association between CO₂ price and the share of electricity (panel a) and hydrogen (panel b) in final energy in the NetZero scenario. Hydrogen includes both direct use and indirect use via hydrogen-derived liquids or gases (e.g., biofuels, synfuels). Colours denote models and shapes denote years (see legend). Black lines show regression fits, with shaded grey regions indicating 95% confidence intervals. the 95% confidence interval around the fitted mean relationship (i.e., the uncertainty in the estimated regression line, not a 95% prediction interval for individual model-year points). The dashed red line indicates the average ETS price in the first half of 2025. All monetary units are expressed in €2020. Euro-Calliope does not report CO₂ prices for 2040 year. MEESA is excluded as it does not report CO₂ prices. Models differ in their CO₂-pricing approach: some prescribe an exogenous carbon tax, while in others the CO₂ price emerges endogenously from imposed emission constraints (see Methods–Scenarios).”

- Figure 4 ... Could sector-level resolution versions of these plots be created? The results of the models indicate that H₂ will tend to be deployed for hard-to-electrify sectors from 2040, so plots that show the uncertainty across models for industry and transport could be particularly helpful in visualizing this point.

Response: In response to your comment, we made the sectoral disaggregation and added the new plot in the Supplementary Information. Moreover, we replaced triangulation-based interpolation with a smooth 2D regression surface to highlight the underlying relationship and reduce visual noise.

Revised text around Fig 5: ~~“By 2040, a clear separation emerges between the NPI and NetZero scenarios at a 45% electrification rate, establishing this as the minimum required for deep decarbonization across models. In contrast, deep decarbonization is achieved regardless of hydrogen uptake. The net-zero emissions frontier (dashed green line, Fig. 4, 2050) demarcates the boundary between combinations of electricity and hydrogen shares that are sufficient to achieve carbon neutrality and those that are not. Combinations lying above or to the right of this frontier correspond to net-negative or near-zero CO₂ emissions, while those below and to the left remain above zero. This frontier illustrates that net-zero is not uniquely determined by a single decarbonisation pathway: multiple combinations of electrification and hydrogen penetration can, in principle, achieve the same emissions outcome. The frontier is crossed by a substantial number of model–scenario combinations in 2050, confirming that net-zero is technically achievable across a diverse range of pathways. However, most NetZero-variant data points cluster near or just above the frontier, with median electricity shares around 55% and median hydrogen shares around 8% across the NetZero scenario variants. Only two model–scenario combinations exceed a 9% hydrogen share in 2050, i.e., PRIMES under NetZero and TIAM-ECN under LimCCS, suggesting that high hydrogen penetration, while feasible, represents a less common~~

decarbonisation pathway among the models assessed. A sectoral disaggregation of this figure is provided in Fig SI 6.”

Added text and figure in SI: “

CO₂ emission envelope per final energy sectors

Across the three sectoral panels of **Fig SI 1**, the uncertainty envelopes reflect the distinct decarbonisation dynamics of Buildings, Industry, and Transport, driven by differences in electrification potential and the role of hydrogen across sectors. Together they decompose the aggregate transition documented in **Fig SI**, where the NetZero electricity share spans 45-65% and hydrogen remains below 9% in 2050 at the total final energy level.

In Buildings (**Fig SI 1**, top row), electrification is the dominant and early decarbonisation lever, consistent with the aggregate finding in **Fig SI** that electricity carries the initial decarbonisation burden. The electricity share under NetZero rises from a median of 46% in 2030 to 64% in 2040 (+18 percentage points), with hydrogen remaining negligible (0-2% in 2040, 0-3% in 2050). TIAM-ECN consistently occupies the high-electrification extreme, reaching 80-93% by 2040-2050. Hydrogen adoption in buildings is minimal across all models and scenarios, confirming that direct electrification (e.g., heat pumps) is the primary pathway for this sector.

In Industry (**Fig SI 1**, middle row), hydrogen plays a structurally larger role than in Buildings, and its contribution grows over time. This sector is the primary driver of the post-2040 hydrogen acceleration observed at the aggregate level in **Fig SI**. Under NetZero, the hydrogen share spans 1-8% in 2030, widening to 1-10% by 2040 and 2-20% by 2050, with a median of 7% in 2050. Electrification is more constrained here, plateauing at a median of approximately 54% in 2050, reflecting hard-to-electrify industrial processes. The highest hydrogen shares (up to 24%) appear in PROMETHEUS under LimBio in 2050, flagging hydrogen as a biomass substitute in this model. Industry thus displays the clearest evidence of the two-step transition described for **Fig SI**, i.e., electrification carries the initial burden, followed by increasing hydrogen penetration as first-mover electrification opportunities are exhausted.

In Transport (**Fig SI 1**, bottom row), the transition is the most heterogeneous across models. Electricity shares are low in 2030 (NetZero median 14%, range 6–20%) but surge between 2030 and 2040 (median +23 percentage points, the largest sectoral increase in **Fig SI 1**), reaching 36% by 2040, contributing substantially to the sharp aggregate electricity rise seen in **Fig SI** over the same period. Hydrogen in transport is generally modest (median 1% in 2050 under NetZero), but PRIMES projects an extreme value of 38% hydrogen share in 2050 under NetZero, driven by hydrogen uptake in heavy-duty and shipping segments. This single model substantially widens the upper bound of the sector's hydrogen uncertainty envelope in **Fig SI 1** and partly underlies the PRIMES outlier behaviour noted in **Fig SI**.

By 2040 (**Fig SI 1**, middle column), the NPI-NetZero separation in electricity share is sharpest in Transport, where the NPI range (5-29%) and NetZero range (28-72%) barely overlap (1.7 percentage points), compared to Industry (12 pp overlap) and Buildings (17 pp overlap). This confirms Transport electrification as the strongest early sectoral discriminator between mitigation and reference pathways, underpinning the aggregate 45% electrification threshold identified in **Fig**

SI. Across all sectors of **Fig SI 1**, hydrogen adoption remains largely a post-2040 phenomenon under NetZero, consistent with its role in decarbonising residual hard-to-abate processes rather than leading the transition.

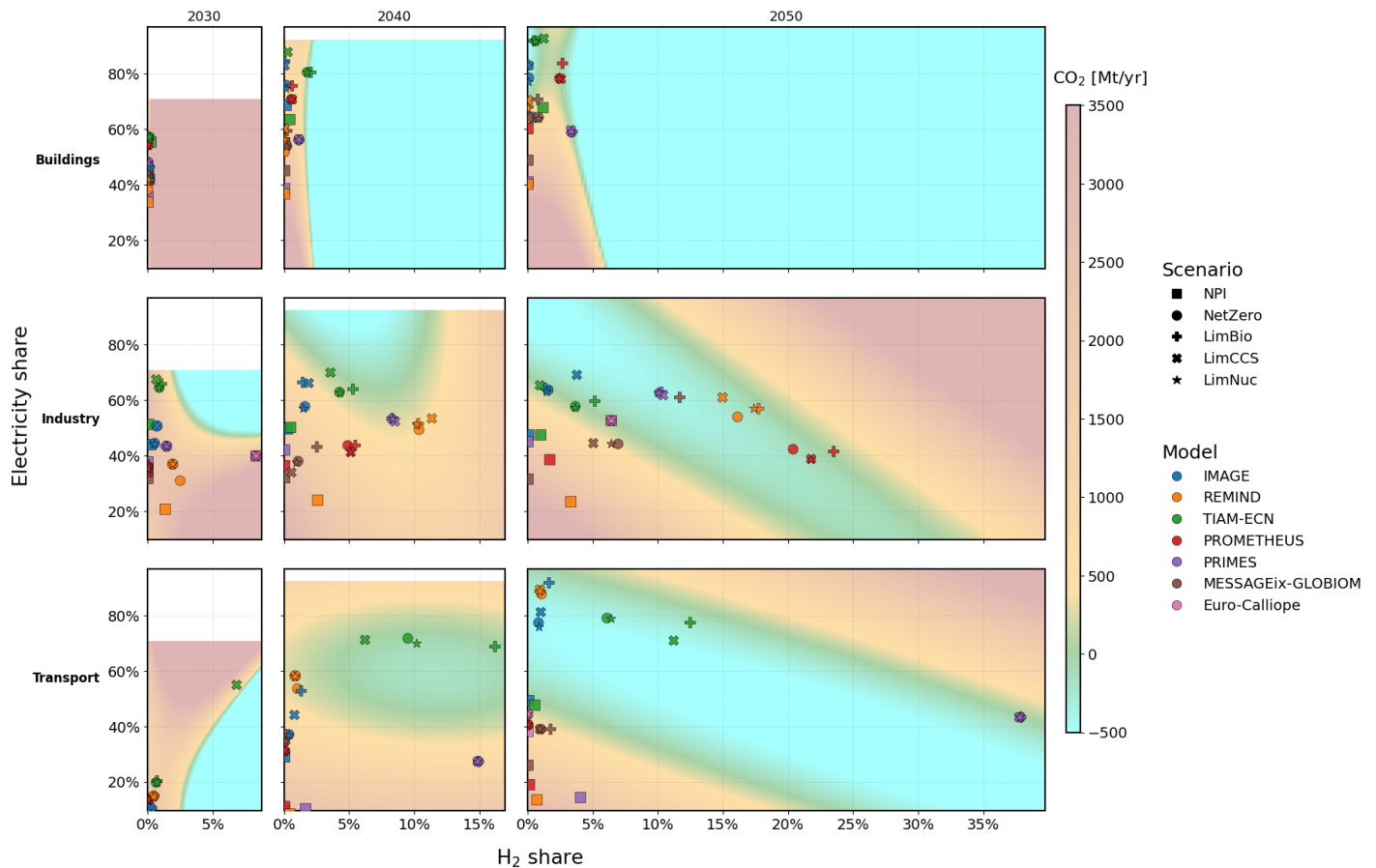


Fig SI 1. Envelope of CO₂ emissions and electricity/hydrogen shares in sectoral final energy
Scatter plots show the association between CO₂ emission levels and the shares of electricity and hydrogen in sectoral final energy use for Buildings (top row), Industry (middle row), and Transport (bottom row), in 2030 (left column), 2040 (middle column), and 2050 (right column). Each point represents an individual scenario, with colours indicating models and marker shapes indicating scenarios (see legend). The background shading shows a smooth quadratic fitted emission surface (regression-based), used to highlight the underlying pattern in the data. Colours range from net-negative emissions (cyan), to near-zero emissions (green), to higher emissions (dark orange/brown), as indicated by the colour bar.

”

- Figure SI4 is really interesting, and could be moved to the main text to provides the content for an extended discussion.

Response: In response to your comment, we have moved Fig SI4 into the main text, labelled as Fig 2. We have also revised the Discussion section to incorporate the key points illustrated by this figure.

Additional comments:

1. Total system cost

The total system cost can also provide useful information beyond the relative shares of different technologies. For example, under the 3 constrained scenarios, do some models indicate a particular constraint

is more impactful than others? Differences in the magnitude along the economic axis could generate additional insight from the ensemble approach that would be a significant contribution beyond the results from individual models. Since these models optimize on an economic objective function, it should not be too difficult to include this information. Normalized costs for each scenario could be used if absolute values are not the most convenient basis.

Response: We agree that normalised cost comparisons are a useful way to extract additional insight from the ensemble, and we note in the revised text that system cost metrics can help further explain inter-model differences beyond technology shares. However, harmonising “total system cost” across models is not straightforward because the definition of the system boundary differs substantially across the modelling frameworks in this intercomparison, for example some models represent primarily the European power system, others the European energy system, and several represent the global energy system with Europe as a region. These differences affect which cost components are included (and how they are accounted for), making like-for-like aggregation challenging. Nevertheless, following the reviewer’s suggestion, we arranged to provide indicative cost comparisons using normalisation and additional harmonisation assumptions. To keep the scope of the present paper focused on hydrogen dynamics, the detailed cost analyses are provided in a companion paper by Pietzcker et al, and we will include a link to that paper in the manuscript once it is published.

Revised text in Discussion: “Hydrogen demand as a means to achieve NetZero targets varies widely, from negligible uptake to over 11 EJ (~77 MtH₂) in 2050. This variability reflects differences in model assumptions and the diversity of substitution pathways, particularly in the industrial and transport sectors where the representation of process-level technologies remains limited or not homogeneous across models. A complementary lens for interpreting such divergence is total system cost, but we do not report a harmonised cost comparison here because ‘system cost’ is not defined consistently across model types and formulations, and system boundaries differ substantially (e.g., power-system versus whole-energy-system versus global-system representations). This divergence is expected to diminish as our models further develop. We recommend harmonizing input parameters in future studies to enhance comparability and improving the granularity of industrial sector modelling to better capture technology-specific substitution options. Moreover, pronounced modelling uncertainty persists regarding the magnitude and timing of hydrogen deployment. Across models, early decarbonisation is consistently driven by electrification, with hydrogen expansion occurring mainly after 2040 as hard-to-abate sectors come into focus. Some models—particularly those envisaging limited CCS, constrained biomass, or optimistic technology cost reductions—demonstrate only moderate hydrogen uptake. This suggests that low hydrogen deployment can be offset by alternative mitigation strategies, such as greater reliance on CCS, BECCS, or sustained technological learning. Others, such as PRIMES, project relatively high hydrogen shares in transport, yet these are exceptions rather than the rule. Given this high uncertainty, we advocate the use of global sensitivity and uncertainty analyses to elucidate robust hydrogen pathways”

2. Policy

The paper affirms previous insights obtained from individual IAM studies and other work (Lines 274-284). A deeper analysis, as suggested above, may produce further insight as to how policy makers might direct resources to alter the underlying dynamics, or confirm that the resources needed exceed policy ambitions to a degree that new expectation must be set.

Response: We agree that parts of our analysis corroborate insights from previous IAM studies. The added value of our work lies in demonstrating that these conclusions remain robust across a harmonised multi-model ensemble, including NetZero variants with constrained biomass, CCS, and nuclear. In response to your comment, we have clarified the Discussion to more explicitly articulate the policy implication of this finding, namely whether targeted interventions could realistically shift near-term hydrogen deployment, or whether the consistent shortfall across models indicates that current ambitions exceed feasible resource and cost constraints.

Revised text in Discussion: “None of the models consistently achieve the European Commission’s REPowerEU target of 10 Mt domestic renewable hydrogen production and 10 Mt of imports by 2030. Across models, projected domestic supply in the NetZero scenario averages ~0.35 EJ (~2.5 MtH₂) in 2030, far below the target, while hydrogen imports remain limited because most models either omit trade or include it only through exogenous assumptions (e.g., REMIND). Projected domestic hydrogen supply in the NetZero scenario averages ~0.35 EJ (about 2.5 MtH₂) across models in 2030, far below the target. Most models either omit hydrogen trade or include it only through exogenous assumptions (e.g., REMIND), resulting in modelled hydrogen imports that also fall well short of 10 Mt. The persistence of this shortfall across diverse model types and formulations raises concerns about the feasibility of current policy ambitions, particularly in light of rising cost estimates for both domestic and imported renewable hydrogen^{30,31} and a widening implementation gap. This gap raises concerns about the achievability of current policy ambitions, especially in light of recent increases in cost estimates for both domestic and imported renewable hydrogen^{30,31} and a widening implementation gap³². Taken together, these results suggest that, under prevailing assumptions on technology costs, infrastructure readiness, and renewable availability in Europe, the 10 Mt + 10 Mt target for 2030 is unlikely to be met without either substantial technological breakthroughs or highly coordinated and targeted policy interventions. These findings indicate that, given the prevailing high capital costs for electrolyzers, immature infrastructure, and limited access to low-cost renewables in Europe, the 10 Mt + 10 Mt target for 2030 is likely unattainable without significant technological breakthroughs or coordinated policy interventions.”

Reviewer #2 (Remarks to the Author):

This manuscript presents the results of a model comparison study focussing on the role of hydrogen in Europe towards 2050. While many papers have been written on this particular topic, the novelty of the present manuscript is the use of a specially designed set of consistent scenarios across the different models in an effort to harmonize boundary conditions; the idea is that variations in results between the different models but within the same outer scenario (e.g. net zero emissions by 2050) are representative of uncertainty due to variations in costs, technological parameters as well as model formulation across the models.

Response: We thank the reviewer for the time and dedication invested in reviewing the manuscript, and for the thoughtful and in-depth comments provided throughout the review. We found these comments extremely helpful, and they guided substantial improvements to the manuscript. In the revised version, we conducted a structured survey across the participating models, reported the full results as a table in the Supplementary Information, and explicitly used these inputs to inform the interpretation of both robust patterns and areas of divergence in the main text. We also expanded the Supplementary Information with an additional figure that further elaborates Fig. 5, and moved one previously supplementary figure into the main text to improve clarity and analytical depth. Altogether, the manuscript has been substantially revised, particularly in the Results, Discussion, Methods, and Supplementary Information, to better exploit the ensemble nature of the dataset and to clarify the added value of the harmonised multi-model comparison. We hope that these revisions satisfactorily address the reviewer's comments. Below, we respond point-by-point and indicate where changes were made (additions in purple; removed text struck through).

The paper is relatively well-written (though some of the text, especially the discussion section, could be a bit more concise), and the analysis seems to be technically sound and correct; I think the conclusions are supported by the models and data that are used.

Response: We thank the reviewer for this positive assessment. We have further streamlined the Discussion to improve conciseness, while maintaining the technical robustness of the analysis and the strength of the conclusions.

I have the impression that the manuscript does not exploit the available model data to a fully satisfactory level. The main impression I have after reading it is that there is a lot of variation in hydrogen deployment across different models --- this is, however, already clear from previous studies. The authors spend quite a bit of space on the conclusion that electrification is the number 1 strategy for decarbonisation in Europe, and that hydrogen uptake, especially direct use, will play only a relatively minor role, and that only well after 2030 (missing the EU target of 10 + 10 Mt by 2030). This conclusion, I would argue, is not particularly novel either, and basically well-known in the literature.

I was hoping for more insights that could have only been derived from this kind of harmonized comparison study. If there are large differences between models in particular sectors despite a similar "outer" scenario, why is that? Which particular differences in parameters, assumptions or

modelling methodology are plausibly to blame? Moreover, the manuscript did not derive much or any insight from the full set of NetZero scenarios (including the ones with limited biomass, CCS and nuclear capacity), which is a shame as such scenarios are quite relevant.

Response: We thank the reviewer for this important comment. We agree that simply documenting inter-model spread in hydrogen deployment, or reiterating electrification as a near-term decarbonisation lever, is not sufficient on its own, as these messages are broadly recognised in the literature. In the revised manuscript, we therefore strengthened the analysis to extract insights that are specific to a harmonised multi-model comparison and to make fuller use of the scenario set and the available ensemble outputs. Concretely, we conducted a structured survey of model characteristics, reported the full results as a table in the Supplementary Information, and used these inputs explicitly to interpret where and why models diverge, e.g., by distinguishing differences in scope, modelling paradigm, sectoral detail, and temporal resolution. We also expanded the Results and Discussion to more clearly separate robust signals from areas of divergence across sectors and time, and we strengthened the use of the full NetZero scenario set by adding additional scenario- and sector-resolved evidence in the Supplementary Information, including a new figure that elaborates the main results and moving one previously supplementary figure into the main text to improve clarity.

Regarding attribution, we agree that a key promise of intercomparison is to help explain why models disagree within the same “outer” scenario. At the same time, a full parameter-level attribution across models is not straightforward in a study of this scope, because models differ in structure, technology definitions, and endogenous decision mechanisms, and outcomes reflect interacting assumptions across multiple sectors rather than a small set of comparable inputs. We therefore provide a structured interpretation that links divergence patterns to documented model characteristics (from the survey), rather than attempting a like-for-like comparison of detailed input databases that are not directly comparable across frameworks.

We also note that the full underlying dataset used for this study will be made publicly available once the accompanying paper led by Robert Pietzcker et al. is published. This will enable other researchers to interrogate the ensemble results and explore additional variables and relationships beyond the scope of the present manuscript. Finally, we recognise the reviewer’s point that cost and system-wide metrics can provide additional insight beyond technology shares. However, harmonising system-cost definitions across models is challenging because system boundaries differ substantially, e.g., power-system, whole-energy-system, and global-system representations. We therefore keep the focus of this paper on hydrogen dynamics, including direct and indirect uses, while the companion paper (currently under review in Nature Energy) provides the broader cost- and system-wide perspective, and we will include a link to that paper in the manuscript once it is published.

In some specific instances, I think the manuscript overstates an alleged gap in previous research on European hydrogen, but I have to disclose that I have written on this specific topic myself.

**** Introduction**

The introduction is quite well-written, and sets up the context nicely. It is concise and to the point,

which is great. I agree there is a lot of uncertainty and divergent perspectives around the future of hydrogen in Europe! Towards the end of the introduction, however, I think you can do a little bit of a better job making the research gap really concrete. One contention I have (on line 74) is that I think there is actually a pretty decent set of modelling studies in the literature that have full representation of hydrogen as an intermediate carrier with sector coupling etc.; several of the studies you cite do, and I could also refer you to the references in Table S1 in ref. 11 (full disclosure: I'm one of the authors of ref. 11, sorry for any self-references!) which point to a number of studies based on models such as JRC-EU-TIMES, MIRET-EU / Integrate Europe, Balmorel and Enertile. If I remember correctly, some of those have decent coverage of sector-coupling around hydrogen. Moreover, ref. 11 (sorry!), while being sector-coupled etc., also explores which factors (technological, policy-wise etc) are most significant in explaining the wide variation we see in cost-optimal scenarios.

In summary, as I see it, main novelty of the present manuscript, and the thing which I do think closes a research gap, is presenting a model comparison analysis with a focus on hydrogen and a harmonized set of scenarios. But be careful not to overstate the research gap.

Response: Thank you for this important comment, and for pointing out where our introduction could be more precise. We agree that we should not overstate the research gap. Indeed, there is a substantial body of European modelling work that represents hydrogen as an intermediate carrier with sector coupling, and we have revised the introduction accordingly to better reflect this literature. In particular, we now frame the main novelty more clearly as the harmonised multi-model intercomparison focused on hydrogen, rather than implying that full conversion-chain representations are generally missing. We also updated Ref. 11, which was cited as a preprint in the submitted version, to its published version in the revised manuscript, and we refer interested readers to Table S1 in Ref. 11 for additional relevant studies (e.g., sector-coupled European models). Finally, we revised the wording around the research gap to emphasise that what remains limited is not hydrogen representation per se, but systematic cross-model comparison under a harmonised scenario design, with consistent reporting of direct and indirect hydrogen uses to clarify robust insights versus model-dependent outcomes.

Revised text in main: “Previous work, such as by van der Zwaan et al.¹³, has highlighted the importance of sector coupling for achieving net-zero CO₂ emissions, by primarily focusing on electricity and direct hydrogen consumption by 2050. ~~Comprehensive analysis of hydrogen’s full conversion chain, especially indirect uses such as synthetic fuel production, remains limited~~ A growing body of European modelling studies represents hydrogen as an intermediate carrier within sector-coupled energy system frameworks, including analyses that explore hydrogen’s role across multiple carriers and sectors (see also Table S1 in Ref. 11 for additional examples). At the same time, evidence is more limited on how robust conclusions about hydrogen’s full conversion chain, including indirect uses such as synthetic fuel production, remain when assessed consistently across different model families and formulations under a harmonised scenario design. ~~despite evidence that indirect uses can exceed direct consumption by a factor of 1.5 to 6.9 in prior studies~~¹⁴ Correspondingly, there are relatively few systematic model-comparison studies using harmonised scenarios with a dedicated focus on hydrogen and consistent reporting of both direct and indirect uses to clarify hydrogen’s infrastructure and system-wide impacts. ~~In addition, the field~~

~~lacks systematic model comparison studies using harmonized scenarios to clarify hydrogen's infrastructure and system-wide impacts."~~

**** Results**

For Figure 1, I'm not quite sure if a Sankey diagram is the right way of visualising the results, as the structure of the data is quite simple and doesn't call for this kind of diagram. An obvious alternative would be two panels (for inputs and outputs) with a set of stacked bars (one for each model) with the bars showing total hydrogen, broken down by category. It's not a critical flaw but you could consider this or other alternative presentations. (Also, I find it a little difficult to distinguish between the colours for direct uses, exports and losses.)

Response: Thank you for this helpful suggestion. In response, we created the requested alternative visualisation as a supplementary figure, showing hydrogen inputs and outputs as stacked bars for each model, which addresses the readability concern and supports quick cross-model comparison. At the same time, we prefer to keep the Sankey diagrams in the main text because they more directly illustrate the hydrogen conversion chain as a flow system, including intermediate transformations and end uses, which is central to the purpose of Fig. 1. The stacked-bar representation necessarily compresses the information, and it can obscure smaller flow components in models with low total hydrogen volumes, i.e., the relative structure of the flows becomes harder to interpret when the overall bar is small (e.g., IMAGE). We also refer the readers to the SI to look at the bar chart representation.

Revised text: "To illustrate this heterogeneity, Fig 1 presents a collection of Sankey diagrams depicting the hydrogen transformation chain, as projected by our models for 2050 in the NetZero scenario. The diagrams show how hydrogen is supplied to the European energy system and how it is subsequently utilised in the various demand sectors. A more compact representation is provided in the form of bar charts in Fig SI 2."

Added section in SI:

Hydrogen conversion chain demonstrated through bar chart

In the main text, hydrogen conversion pathways are presented using Sankey diagrams (**Error! Reference source not found.**). For a more compact cross-model comparison, this supplementary figure shows the same underlying data as stacked bar-chart visualizations, with NetZero in the top row and NPI in the bottom row (supply on the left, demand on the right).

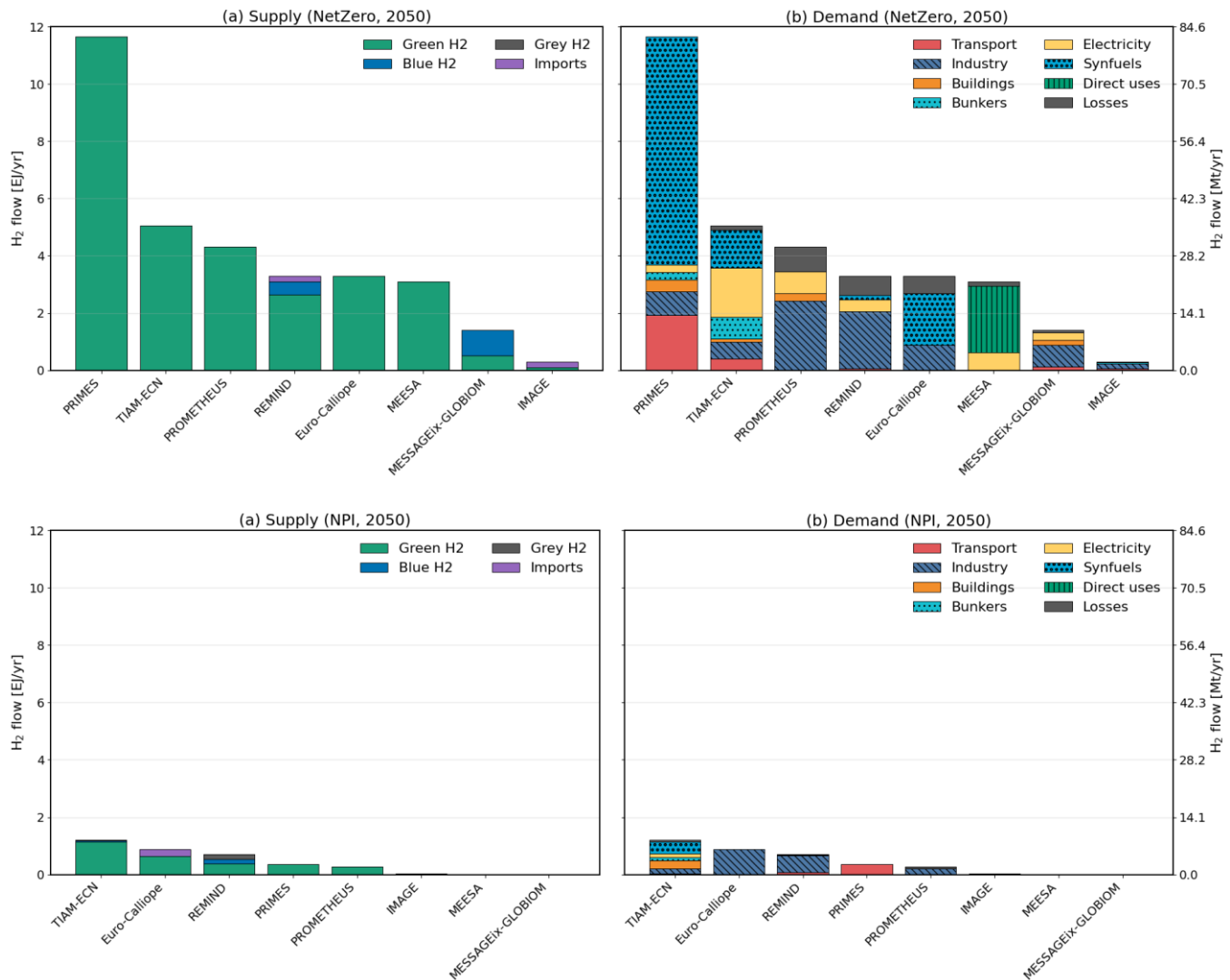


Fig SI 2. Modelled hydrogen supply and demand composition across models in 2050 under NetZero (top) and NPI (bottom)

Stacked bar charts show hydrogen supply (left panels) and demand (right panels) by model, with models ordered by total hydrogen supply. Supply is disaggregated into Green, Blue, Grey, and Imports; demand is disaggregated into Transport, Industry, Buildings, Bunkers, Electricity, Synfuels, Direct uses, and Losses. Bars are in EJ/yr (left axis), and the right axis shows equivalent Mt/yr (HHV conversion).

In the first subsection of the results section, I'm maybe missing a bit of explanation or insight. I understand that the results across the models vary a lot, maybe especially w.r.t. synthetic fuel production and "Industry" (it would be nice to briefly define what exactly is included in this sector). But what drives these variations, specifically? Why does PRIMES, for example, need a lot of hydrogen-derived synthetic fuel and the other models do not? How does this relate to other findings in the literature about what drives demand for hydrogen / synthetic fuel? Could this have to do with assumptions on the transportation sector, or is it something else? I think the reader would benefit from some more in-depth insights, even if only in reference to later sections.

Response: We thank the reviewer for this important comment. This concern is closely aligned with the points raised by Reviewer #1, and in response we further revised the manuscript to provide

more explicit interpretation of the observed cross-model variation in the first Results subsection, rather than only describing it. To support this, we also performed a structured survey across the participating modelling teams, the results of which are reported in the Supplementary Information and used explicitly to interpret divergent outcomes across models.

In the revised text, we now clarify that the variation in hydrogen and synthetic-fuel deployment reflects differences in model scope, sectoral detail, transport representation, infrastructure treatment, and in whether hydrogen demand is represented endogenously or partly exogenously. We also make clearer what is included in the broader sector categories. In particular, 'industry' refers to the industrial energy demand category as harmonised across the ECEMF outputs, though the level of subsector and process detail represented within that category differs substantially across models.

We further added model-specific interpretation where possible. For example, PRIMES' comparatively high hydrogen-derived synthetic-fuel use is linked to model-specific allocation choices in harder-to-electrify transport segments, whereas Euro-Calliope tends to favour battery-electric transport and lower-cost flexibility options, while PROMETHEUS retains a larger role for liquid fuels. We therefore now point readers more explicitly to the later sectoral analysis in Fig. 2, where these differences are unpacked in greater detail. More generally, the revised text now explains that hydrogen and synthetic-fuel demand are not driven by a single parameter, but emerge from the interaction of technology costs, infrastructure assumptions, substitution possibilities, and sector-specific representation, which is consistent with broader findings in the literature.

Revised text: Please refer to the tracked changed document.

I think the second subsection is mostly fine. I do wonder, however: could this subsection and the above be enriched by pulling in comparisons with the variations on NetZero scenarios that you haven't used yet so far? This could illuminate some of data. For example, I am left wondering "_Why_ does PRIMES show such a high fraction of hydrogen in transportation" and, indeed, "why does MESSAGEix-GLOBIOM show such a high fraction of gas in transportation"?

Response: We thank the reviewer for this suggestion. This point is closely aligned with the comments raised by Reviewer #1 regarding the need for more explicit interpretation of divergent outcomes across models. In response, we further revised the manuscript to clarify the main observations and to make better use of the available cross-model information. We also conducted a structured survey across the participating modelling teams, reported in the Supplementary Information, and use these model-characteristics inputs explicitly to interpret the main outliers and sources of divergence.

We agree that comparisons across the NetZero variants, i.e., constrained biomass, CCS, and nuclear cases, can help illuminate model behaviour. At the same time, we chose to keep the core narrative in the main text centred on the NetZero and NPI scenarios, to preserve clarity and avoid overloading the main Results section with multiple scenario dimensions at once. We therefore use the NetZero variants primarily in the later Results subsection and the Supplementary Information,

where they help test the robustness of the main findings and identify where specific constraints materially shift electricity and hydrogen outcomes.

Regarding the specific transport outliers, we expanded the text in this subsection and the subsequent discussion to provide additional clarification. In particular, we now explain that PRIMES' high hydrogen share in transport reflects model-specific allocation to harder-to-electrify transport segments, rather than a simple one-to-one competition between hydrogen, electricity, and synthetic fuels across all transport modes. By contrast, MESSAGEix-GLOBIOM's relatively high role for gaseous fuels in transport reflects its stylised transport representation and reporting aggregation: the reported gas category includes multiple gaseous carriers, i.e., natural gas, biogas, and hydrogen mixed in the gas grid, and also captures shipping-related fuel use. Its broader regional aggregation within Europe also contributes to this outcome. The MESSAGEix-GLOBIOM result should therefore be interpreted as a broader gaseous-fuel and shipping pathway, rather than as fossil natural gas use in road transport alone.

More generally, these differences reinforce our main interpretation that the observed divergence is driven not by a single parameter, but by interacting differences in model scope, sectoral detail, transport representation, substitution possibilities, and infrastructure treatment. We now make these points more explicit in the revised Results and Discussion sections.

Added text: “Interpreting this divergence requires acknowledging that models differ in paradigm, sectoral detail, transport representation, and cross-sector coupling. In PRIMES, higher hydrogen shares reflect model-specific allocation to harder-to-electrify transport segments. In Euro-Calliope, by contrast, road transport can be met by battery-electric, fuel-cell, and combustion technologies, but hydrogen remains economically uncompetitive once vehicle and infrastructure costs are considered, leading to a predominantly battery-electric pathway. PROMETHEUS retains a comparatively high role for liquid fuels, indicating a different balance between electrification, hydrogen, and liquid-fuel substitution pathways. MESSAGEix-GLOBIOM's comparatively high gas share in transport should also be interpreted with care, because its reported gas category aggregates multiple gaseous carriers and includes shipping-related fuel use, rather than representing fossil natural gas use in road transport alone. In industry, cross-model differences similarly reflect differences in process detail and substitution structure. In IMAGE, industrial outcomes emerge from competition among electrification, hydrogen, gaseous fuels, biomass, and CCS under cost, resource, and feasibility constraints, rather than from a single dominant substitution rule.”

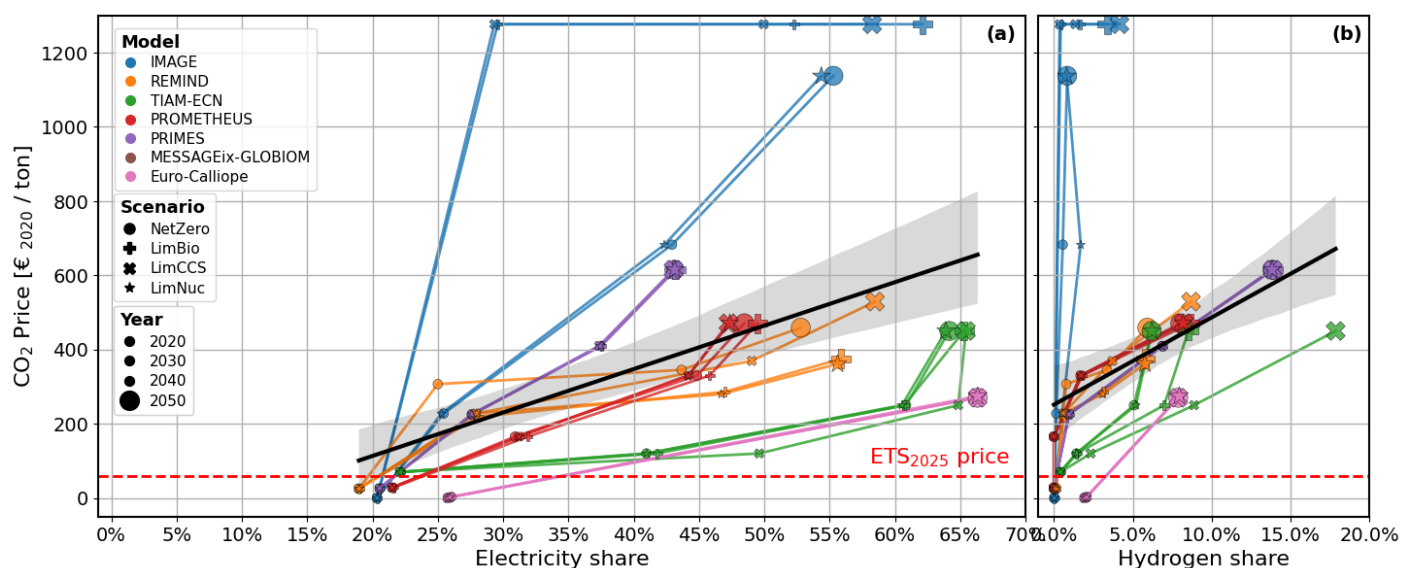
In the third subsection, I appreciate seeing results from some of the different scenarios, although I think it's a shame that no conclusions are drawn from any comparison between LimBio, LimCC, LimNuc etc.; the reader is left to search the figure for a pattern. Is there not much difference between these scenarios, or is the difference highly heterogeneous across models? That would be a finding worth stressing in itself! In that case, biomass, CCS & nuclear have been eliminated as unambiguous drivers of hydrogen uptake (though I doubt this, especially the former two).

Response: We thank the reviewer for this observation and agree that the comparison across

NetZero variants warrants explicit interpretation rather than leaving the reader to infer patterns from the figure alone. In response, we revised the figure to include all NetZero variants (NetZero, LimBio, LimCCS, and LimNuc) and added a dedicated paragraph in the Results section that explicitly discusses cross-scenario similarities and differences.

Added text: “Comparing the four NetZero variants, NetZero, LimBio, LimCCS, and LimNuc, the main patterns remain robust, i.e., a positive CO₂ price, penetration relationship, a much steeper response for hydrogen than electricity, and larger inter model than inter scenario variation. LimNuc consistently shows the smallest deviations from NetZero across models. By contrast, LimCCS and LimBio show the strongest impacts on both electricity and hydrogen shares, indicating stronger substitution effects under tighter supply side constraints. In LimCCS, TIAM-ECN becomes a clear hydrogen outlier at roughly 18% by 2050, and REMIND also increases both electricity and hydrogen shares. IMAGE shows a similar signal, under LimBio and LimCCS, the CO₂ price surges, with comparatively higher electricity and hydrogen shares. PRIMES, notably, shows little to no change in hydrogen uptake in LimCCS and LimBio, suggesting its hydrogen pathway is not strongly competing with CCS or biofuel substitution. PROMETHEUS shows only slight share changes across scenario variants, as hydrogen uptake emerges mostly in specific industrial sectors (like steelmaking), which are not massively impacted by limitations in nuclear, biomass, and CCS.”

Revised figure:



For Fig 4, I fail to see the significance of the coloured heatmap interpolating CO₂ emissions. The text also doesn't bring this up at all. I guess in the second panel we can see a sort of correlation between electricity share and emissions, as expected. In the third panel, the heatmap only serves to show the sharp divide between NPI and net zero scenarios, which isn't new information. (It's obvious that the NetZero scenarios are going to have low emissions!) Moreover, I think the triangular interpolation is distracting at best; a more standard technique would be to fit some kind of smooth function to your data, like a regression, and plot that, to separate the pattern from the noise.

Response: We thank the reviewer for this comment. We agree that the original heatmap did not sufficiently communicate its purpose, and we carefully reconsidered how best to present these results. In the revised manuscript, we believe the current version better supports our conclusions and provides a more accurate representation of the underlying model outputs. In particular, it preserves the raw structure of the ensemble more directly, including how the emissions envelope defined by the model spread evolves over time, i.e., shifting upward and widening mainly along the electrification axis.

At the same time, in response to the reviewer's suggestion, we also produced a smooth quadratic-fit version of the figure and provide it in the Supplementary Information for readers who may find this representation more intuitive. However, after consideration, we decided not to adopt this version in the main text. While the quadratic surface offers a smoother visual summary, it also introduces several drawbacks. First, it obscures some features that are visible in the original representation, especially the changing envelope of the most extreme model outcomes across time. Second, because the dataset is relatively small and heterogeneous, a regression-based surface may give a stronger impression of statistical structure than is warranted. Third, adopting such a fit would raise additional methodological questions, e.g., whether the sample is sufficient to support this type of fitted analysis.

For these reasons, we retained the revised main-text figure in a form that we consider more transparent and faithful to the ensemble outputs, while also making the alternative quadratic-fit version available in the Supplementary Information for comparison. We hope this addresses the reviewer's concern while allowing readers to assess both visualisations.

Revised text: “Across all panels (Fig 4), the CO₂ emission envelope expands to reflect the wide uncertainty in projected electrification and hydrogen uptake among models and scenarios. Within the NetZero scenario group in 2050, hydrogen penetration is more constrained, varying from 1% to 9% (with outliers at 14% and 18%). In contrast, the range of electricity shares spans from 45% to more than 65%. Between 2030 and 2040, the electricity share increases most sharply, defining the primary trajectory of emissions reduction in the early NetZero transitions (Fig 4, 2040). By contrast, hydrogen adoption accelerates mainly after 2040, consistent with a two-step transition in which electrification carries the initial decarbonisation burden, followed by broader hydrogen deployment as more challenging sectors decarbonise. At the sectoral level (Fig SI 1), this electrification is most pronounced in Transport, where NPI and NetZero electricity ranges barely overlap, confirming Transport electrification as the strongest early discriminator between mitigation and reference pathways.

Model-specific characteristics strongly influence the breadth of the envelopes. The PRIMES and TIAM-ECN models drive much of the spread in projected hydrogen shares across all periods. The wide envelope of TIAM-ECN primarily reflects its LimCCS scenario, which combines high electricity (67% in Fig 4, 2040 and 2050) and high hydrogen (18% in Fig 4, 2050) shares with sustained net-zero emissions. In PRIMES, all NetZero scenarios except NPI converge on similar outcomes, demonstrating the model's relative insensitivity to varying policy or technology constraints. Further exploration of emission reduction pathways can be found in Pietzcker et al²².

By 2040, a clear separation emerges between the NPI and NetZero scenarios at a 45% electrification rate, establishing this as the minimum required for deep decarbonization across models. In contrast, deep decarbonization is achieved regardless of hydrogen uptake.

Beyond these anchor models, the ensemble reveals two broader decarbonisation archetypes by 2050. One group, led by TIAM-ECN and Euro-Calliope, pursues deep electrification combined with moderate-to-high hydrogen penetration. A second, most clearly represented by PRIMES, achieves comparable emissions reductions through a high-hydrogen, lower-electricity pathway. The remaining models converge on a central pathway of moderate electrification and low-to-moderate hydrogen, forming the modal trajectory of the ensemble.

Notably, even under the NPI scenario, where climate policy ambition is limited, several models still project small but non-negligible hydrogen uptake in 2050, suggesting that some models treat hydrogen as economically competitive even at low carbon prices, reflecting structural cost assumptions rather than purely policy-driven deployment."

Added figure and caption in SI:

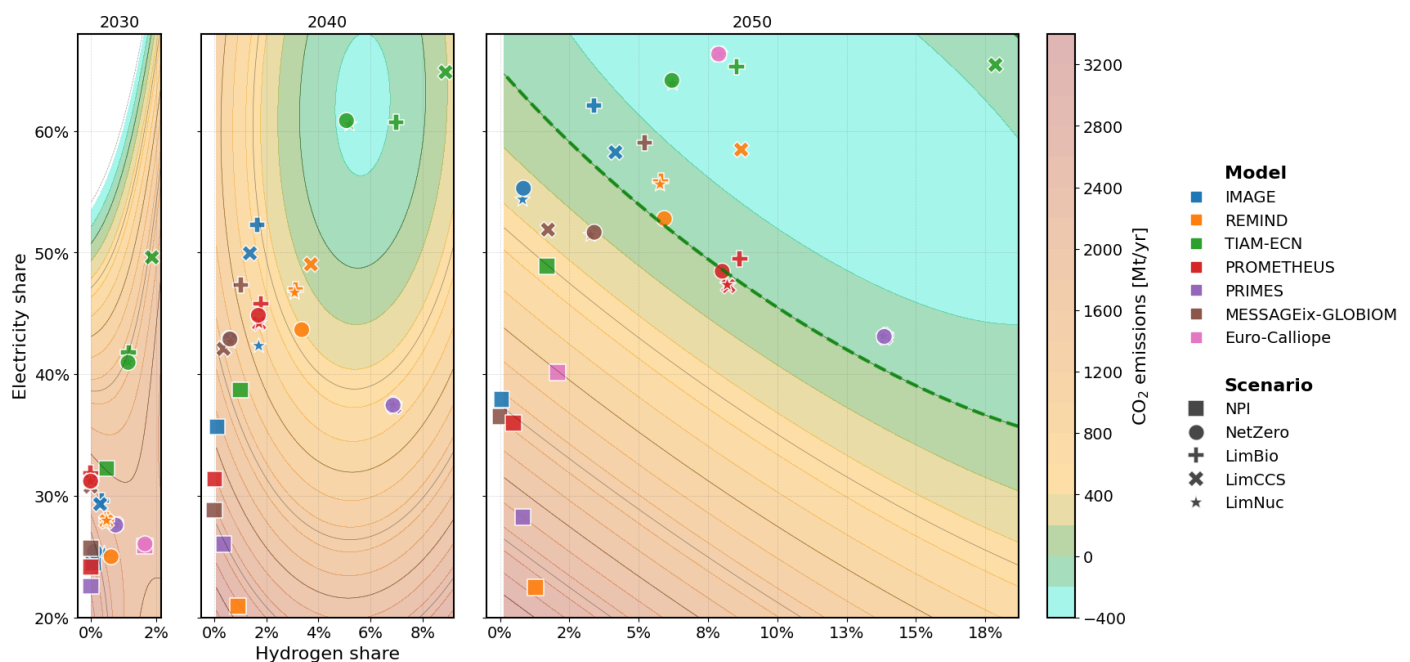


Fig SI 7. Envelope of CO₂ emissions and electricity/hydrogen shares in final energy.

Scatter plots show the association between CO₂ emission levels and the shares of electricity and hydrogen in final energy use for the years 2030 (left), 2040 (middle), and 2050 (right). Each point represents an individual scenario, colour-coded by model (see legend). The background shading shows a smooth quadratic fitted emission surface (regression-based), used to highlight the underlying pattern in the data. Colours range from net-negative emissions (cyan), to near-zero emissions (green), to higher emissions (dark orange/brown), as indicated by the colour bar. The dashed green line in the 2050 panel marks the net-zero emissions frontier (fitted CO₂ = 0 Mt/yr), separating combinations of electricity and hydrogen shares associated with net-negative emissions (above/right) from those with residual positive emissions (below/left).

** Discussion

I'm sorry to bring this up again, but a number of the points you bring up in the first paragraph of this

section have been addressed in ref. 11. There, we show explicitly how liberal use of CCS and/or biomass can entirely eliminate indirect uses of hydrogen; conversely, how limited alternatives can drive hydrogen production and demand upwards. That study also includes a global sensitivity analysis encompassing a variety of techno-economic parameters. More work could certainly be done in this area, but you should be more specific about the alleged gaps.

Response: We thank the reviewer for returning to this point and agree that several mechanisms discussed in the first paragraph of the Discussion, in particular the substitution between hydrogen, CCS, and biomass, have been addressed in previous work, including Ref. 11. We have revised this paragraph to explicitly acknowledge these studies and to avoid overstating any missing gaps in the literature. The revised text now positions our contribution more clearly as a harmonised multi-model intercomparison that tests the robustness of such insights across different model families and formulations, using consistent reporting of both direct and indirect hydrogen uses.

At the same time, while we fully recognise the contribution of Ref. 11, we note that the application of global sensitivity analysis to integrated assessment models remains relatively uncommon in practice, especially in large multi-model exercises. One intention of this work is therefore to motivate and normalise the complementary use of harmonised intercomparison and sensitivity-analysis approaches in IAM-based assessments, rather than to replace existing single-model studies. We believe that the revised Discussion now better reflects this positioning and clarifies how our analysis builds on, and complements, earlier work.

Revised text: “Hydrogen demand to achieve NetZero varies widely, from negligible uptake to over 11 EJ (~77 Mth₂) in 2050, reflecting differences in (i) key techno-economic assumptions, e.g., electrolyser and CCS costs and learning, and BEV–FCV cost competitiveness, (ii) model structure, i.e., sectoral and process detail, substitution options, and spatio-temporal resolution, and (iii) non-cost constraints and parameters, e.g., diffusion limits and resource potentials that can bind even when costs are similar, especially in industry and transport where process-level technology representation is still uneven. **Several of these mechanisms, e.g., substitution between hydrogen, CCS and biomass, have been examined in previous sector-coupled European studies**¹¹. Fig 2 shows that this spread is concentrated in a small number of model-sector combinations, i.e., under NPI hydrogen remains mostly industrial with only marginal transport uptake in most models, while broader uptake appears only in selected NetZero outcomes. The added value of our analysis is to test how robust such conclusions remain across different model families and formulations under a harmonised scenario design, using consistent reporting of both direct and indirect hydrogen uses. A complementary lens is total system cost, but we do not report a harmonised comparison because system cost definitions and boundaries differ across model types, e.g., power system, whole energy system, and global system frameworks. Across models, early decarbonisation is led by electrification, while hydrogen expands mainly after 2040 as hard-to-abate sectors become more prominent. In several models, low hydrogen uptake is offset by alternative mitigation routes, e.g., CCS, BECCS, or stronger technological learning, while high transport hydrogen in PRIMES remains an exception rather than a central trend. We therefore recommend harmonizing key input assumptions and improving industrial sector granularity in future model comparisons, and **complementing multi-model comparisons with global sensitivity**¹¹

and uncertainty to identify robust hydrogen pathways.”

The link to the Zenodo repository with data and code doesn't seem to work, which is a problem w.r.t. to reproducibility.

Response: Thank you for pointing this out. At the current stage, we cannot yet make the full scenario dataset publicly available because the underlying scenario data are part of an accompanying paper led by Pietzcker et al., which is currently under review at Nature Energy. To avoid prejudicing that review process, the data have therefore not yet been uploaded to Zenodo. Once the companion paper is published, we will upload the full set of scenarios and associated data to Zenodo. We have clarified this in the manuscript to ensure transparency regarding the timing of data availability and to address reproducibility concerns.

**** Methods**

It would be nice to know a little bit about how the included models were selected; on which basis or selection criteria.

Response: We thank the reviewer for this suggestion and agree that greater transparency on model selection is valuable. In response, we added text to the Methods section clarifying the basis for model inclusion. Specifically, the models were selected from those participating in the European Climate and Energy Modelling Forum (ECEMF) Horizon Europe project consortium (<https://www.ecmf.eu/>), which brings together a set of well-established European and global energy system and integrated assessment models that are actively used to inform European climate and energy policy. We agree that more systematic and explicit criteria for model selection merit further attention, and note that a dedicated methodological paper on model selection for intercomparison studies is currently being developed by an ECEMF partner, focusing explicitly on this issue.

Added text in Method: “The models were selected from those participating in the European Climate and Energy Modelling Forum (ECEMF) Horizon Europe project consortium, which brings together a set of well-established European and global modelling frameworks that are actively used to inform European climate and energy policy. The selection reflects the ECEMF objective of comparing a diverse yet policy-relevant ensemble of models rather than applying strict ex-ante screening criteria. We agree that model selection in intercomparison studies warrants careful consideration, and note that a dedicated methodological paper on model selection for intercomparison exercises is currently being developed by an ECEMF partner.”

**** Minor comments**

- At least two citations are a little outdated (10, 11); the manuscript references preprints that have since been published (in ERL and Joule resp.)

Response: Thanks you. We have updated the references.

- I could not accurately find the source for ref. 22.

Response: Indeed. This is the companion paper that is under review in Nature Energy. Since the corresponding author did not opt for a preprint, the correct reference will be updated the moment

the paper is published.

- I'm not sure if I understand the "Heat" row under Buildings in Fig 2: presumably much of the electricity and gas demand in the buildings sector is already used for heating, so why is Heat also reported as a separate demand?

Response: This is based on the harmonised parameter definitions in the ECEMF project. The heat here refers to direct heat supply (e.g., district heat, recovered process heat, solar heating and warm water).

- For Fig 3, I somehow find it hard to believe that the grey region represents a 95% confidence interval if almost half the data points fall outside this range? Anyway, maybe this dataset is too small to derive a confidence interval like this?

Response: This was raised by the first reviewer. The grey area refers to the uncertainty around the fitted regression line. We have added an extra note in the caption to clarify this: "Black lines show regression fits, with shaded grey regions indicating the 95% confidence interval around the fitted mean regression line (i.e., the uncertainty in the estimated regression line, not a 95% prediction interval for individual model-year points)."

- "not homogeneous" -> "heterogeneous".

Response: We have removed this in the revised text.

- Line 266: "NetZero" -> "net zero" (describing the general policy, not the study-specific scenario?)

Response: You are right indeed. We have removed this in the revised text.

- Line 288: What does "policy momentum" mean? Anyway, one could argue that, whatever the official policy is, Europe as a whole is de facto following an electrification pathway already, so this is not much of a policy recommendation but rather a description of the current state of affairs.

Response: Thank you for pointing this out. To further clarify our message, we have revised this text: "In the short term, policy should prioritise electrification, supported by rapid renewable build-out and targeted, strategic hydrogen investments in key industrial use cases, rather than aiming for full realisation of the 10 Mt + 10 Mt hydrogen target by 2030. The ensemble results suggest that a more pragmatic pathway would sequence large-scale hydrogen deployment predominantly into the 2040s and beyond, aligning ambition with demonstrated progress in technology costs, infrastructure readiness, and the emissions price signals required to support uptake. However, such sequencing warrants caution for hard-to-abate sectors, particularly iron and steel, where a substantial share of existing assets is expected to retire around 2035. Ensuring that replacement capacity is hydrogen-ready is therefore essential to avoid carbon-intensive lock-in and unnecessary reinvestment. In parallel, prioritising new or relocated production capacity in regions with abundant renewable resources could reduce future hydrogen transport costs and facilitate a smoother scale-up once technological maturity is achieved."

- Ref. 35 is not properly cited.

Response: The reference is updated, particularly its doi which was referring to the paper print, is now referring to the pdf.

Response to Reviewer Comments, Revision 2

We thank the editor and reviewers for their careful assessment of the revised manuscript. We are grateful that both reviewers found the first-round revisions largely satisfactory and appreciate the remaining comments, which helped us further improve clarity, transparency, and interpretation. Below, we respond point-by-point to the second-round comments. Reviewer comments are reproduced in black, and our responses are provided in blue. Manuscript changes will be indicated in the same format as before, with added or revised text shown separately where needed.

Reviewer #1 (Remarks to the Author), Revision 2:

In my assessment, the response to reviewer comments and revised manuscript satisfactorily address the points raised in the initial review. I am in support of publication.

Two minor points for author consideration:

1. Not all of the acronyms are defined at their first use (e.g., NPI), and it might be useful to do a pass through the manuscript to correct this.

Response: We thank the reviewer for pointing this out. We have performed a pass through the manuscript to ensure that acronyms are defined at first use. In particular, we now define National Policies Implemented (NPI) at its first occurrence in the Introduction, and we have also clarified other acronyms where needed, including CCS, BEV, FCV, and BECCS.

2. The exclusion of hydrogen imports/exports from the modeling sidesteps an interesting set of questions around Europe as a net energy importer (viz., how might the results change if low cost H₂ were to become available from outside Europe?). I recognize this is clearly out of scope for the current manuscript, but given the authors general interest in modeling of how energy systems evolve, it might be worth exploring separately some aspects of this issue.

Response: We thank the reviewer for this thoughtful suggestion. We agree that low-cost hydrogen imports could influence Europe's transition pathway, especially by changing the balance between domestic electrification, domestic hydrogen production and hydrogen-derived fuels in hard-to-electrify sectors. In the present study, however, hydrogen trade is represented heterogeneously across the model ensemble, which prevents a systematic assessment of this question. We have therefore clarified how hydrogen trade is treated in the models where this is relevant, and added a short discussion noting that a dedicated harmonised trade sensitivity would be needed to assess the effect of abundant low-cost hydrogen imports.

Manuscript changes: We revised the paragraph discussing hydrogen trade after Fig. 1 to clarify that hydrogen trade is represented heterogeneously across the model ensemble. We now explain, in brief, how trade is treated in Euro-Calliope, IMAGE, MEESA, PRIMES, REMIND and TIAM-ECN, and clarify that the ensemble does not provide a systematic assessment of abundant low-cost hydrogen imports. We also added two sentences to the Discussion explaining how low-cost extra-European hydrogen could affect the balance between domestic electrification, domestic hydrogen

production and hydrogen-derived fuels, and why this would require a dedicated harmonised trade sensitivity.

Reviewer #2 (Remarks to the Author), Revision 2:

Thank you for the work you put into the revisions. I feel like all my points have been addressed, most of them in a satisfactory way.

I still have a minor concern which I would like to raise.

Plainly, PRIMES is an outlier when it comes to hydrogen use in transportation and for synthetic fuel production. This is a surprising result (compared to the other models in the study) and pretty much the one data point that stands out the most in the first few figures. Both I and the other reviewer asked in our initial reviews for more discussion of the differences between models, and I think the revised version of the manuscript delivers on this to an extent. But after reading the revised manuscript, I'm still left wondering: "Why does PRIMES have such a high share of hydrogen in transportation, and why does it produce so much synthetic fuel?". In the manuscript, we read: "In PRIMES, higher hydrogen shares reflect model-specific allocation to harder-to-electrify transport segments". To me, this doesn't really answer the question, it just states the surprising data point. It's the why that I'm interested in. Does PRIMES exogenously specify a high hydrogen share in certain transportation sectors? (That would weaken the result, but would be good to know.) Does it severely limit the use of BEVs? Does it somehow arrive at very low hydrogen prices and conclude that hydrogen in the transportation sector is more cost-effective than direct electrification? (That would be very surprising but interesting.) Does this rather reflect differences in scope in the transportation sector? Does PRIMES assume that a lot of shipping will run on hydrogen directly? The reason for large-scale use of hydrogen for synthetic fuel production isn't discussed much in detail either as far as I can see.

The manuscript is already quite long and I wouldn't want you to expand the text with additional paragraphs upon paragraphs of lengthy explanation and investigation. But I think even brief references to the "why"s of differences in results would make the article much more interesting and satisfying to read.

I mention PRIMES specifically here since it's an obvious outlier, but the same could be said about other differences.

Response: We thank the reviewer for this important comment. We agree that the previous wording did not sufficiently explain the reason behind the PRIMES outlier behaviour in transport and synthetic fuel production. We therefore consulted the PRIMES modelling team and clarified this point in the manuscript.

The high hydrogen use in transport in PRIMES is not the result of an exogenously imposed hydrogen share, nor does it arise from an explicit restriction on battery-electric vehicle deployment. Rather, it reflects the model's relatively detailed transport-sector representation. PRIMES distinguishes a broad set of transport modes, vehicle classes and user groups, including several freight and specialised heavy-duty applications. In this more granular representation, hydrogen and

hydrogen-derived fuels can become attractive in applications where direct electrification is more challenging, particularly long-haul freight and specialised heavy-duty segments, where operational requirements, range constraints, charging infrastructure needs and high utilisation rates can make exclusive reliance on battery-electric options less favourable.

The higher production of synthetic fuels is linked to the same transport representation and to the way PRIMES captures fuel decarbonisation policies during the transition. Synthetic fuels provide a route to reduce emissions from the remaining internal-combustion vehicle fleet and from transport segments where direct electrification is more difficult. Thus, the PRIMES result should be interpreted as a consequence of its more granular transport-sector structure and policy representation, rather than as an imposed hydrogen pathway or a restriction on electrification. We have added a concise clarification to the manuscript, while avoiding a lengthy additional discussion.

Manuscript changes: We revised the sentence discussing PRIMES in the Results section to clarify why it reports higher transport hydrogen use and synthetic fuel production. The revised text now states that these outcomes are not exogenously imposed, but arise from PRIMES' detailed representation of transport modes, vehicle classes and user heterogeneity, which allows hydrogen and hydrogen-derived fuels to compete in long-haul freight and specialised heavy-duty segments. We also clarified that synthetic fuel deployment is supported by fuel decarbonisation policies and the continued operation of internal-combustion fleets during the transition.

Response to Reviewers

Manuscript NCOMMS-25-91045B

We thank the reviewers for their careful assessment and for confirming that their concerns have been addressed. No additional scientific revisions were requested in this round. We have completed the editor's production checklist and updated the manuscript package accordingly.

Reviewer 1

“All of my substantive concerns were satisfactorily addressed in the first revision... this manuscript is now suitable for publication.”

Response: We thank the reviewer for confirming that the substantive and minor concerns were satisfactorily addressed and that the manuscript is suitable for publication.

Reviewer 2

“I appreciate my concerns being addressed... The current version of the article is fit for publication.”

Response: We thank the reviewer for acknowledging the expanded explanation of hydrogen uptake in PRIMES and for confirming that the current manuscript is fit for publication. No further scientific changes were requested.