nature portfolio

Peer Review File

Electricity- and Hydrogen-Driven Energy System Sector-Coupling in Net-Zero CO2 Emission Pathways

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The manuscript focuses on a very interesting topic, that is the evaluation of electricity- and hydrogen-based energy system sector-coupling in Europe.

However, addressing the complexity and broad applicability of the proposed method, and contrasting it with existing models would elevate the paper's impact and accessibility.

Here are a few comments that I believe would add more info to the readers:

- I would suggest using impersonal statements rather than "we".

- Literature review in the introduction should be expanded, by including a deeper state-of-the-art.

- At the end of the introduction, a more detailed description of the research gap and the novelty of the paper would be beneficial.

- The methodology is sound, but I would suggest to generate a table with a more detailed description of the boundary conditions used in the simulations.

- The references are very few, and must be increased.

- Figure 9 needs a more comprehensive description.

Reviewer #2 (Remarks to the Author):

The paper is reasonable and can attract the readers. Many results are generated, but no explanation of how. In the section "Models and scenarios," no one equation, assimption, etc., is mentioned.

Reviewer #3 (Remarks to the Author):

This paper presents the results of a self-selected group of collaborating models results on electricity and hydrogen shares in final energy through to 2050 in Europe.

The paper has a number of issues.

1. The paper seems devoid of any discussion of other literature and modelling which suggests different and potentially higher roles for gas in final energy consumption, and hence provides no real sense of why the estimates are apparently low (as the paper suggests). A key role for hydrogen depends on its storability and ability to match demand in real time and on the relative price of techologies which favour hydrogen production, such as electrolysers, the price of CCS equipment and methane.

2. The paper nowhere discusses costs or the dependency of model results on assumptions about costs. Prices matter for the percentages of different technolgies contributions to FEC. No sense is provided on the extent to which prices matter, see above.

3. Models likely differ in key characteristics and how these relate shares of electricity and hydrogen. Indeed it seems that most models deliver very low percentages of hydrogen but some deliver significantly more, the paper needs to discuss why, again in the context of trying to understand what might be driving the model results.

4. The presentation of the range of estimates is not very satisfactory. Model results need to be compared directly as averages across models are meaningless on their own as model results cannot be combined to give averages (as these are not actually the outputs of any model). Some better attempt to present ranges of results from the different models would be more satisfactory.

5. Another unexplained point is what is supplying the rest of FEC if it is not hydrogen and electricity, if for instance this is synethetic fuel then this is part of the hydrogen economy.

6. A final question, is what is the paper actually trying to say? It reads like a partial summary of the outputs of one limited set of models with little insight into what might be driving the results. Much better papers summarising net zero model results exist which actually try and outline what the shape of net zero depends on. The simple presentation of quantity shares alone is not very

interesting to the reader. A paper with this title needs to do a comprehensive review of all relevant modelling (not just the 8 in the consortium) and then attempt to explain where model results might be coming from and explain differences between models in an insightful way.

Detailed comments:

Line 174-75, blue and green hydrogen compete with biomethane.

Line 243, hydrogen's role in synthetic fuel is what?

Line, 329, most models show zero or near zero hydrogen, but some show a lot more, why?

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The manuscript focuses on a very interesting topic, that is the evaluation of electricity- and hydrogen-based energy system sector-coupling in Europe.

Thanks a lot, and we much agree: this topic is very interesting, and also critical for understanding and influencing the current energy transition, in Europe and elsewhere.

However, addressing the complexity and broad applicability of the proposed method, and contrasting it with existing models would elevate the paper's impact and accessibility.

We agree here too, and we should have improved the paper in this respect. We have now done so, and hope that this reviewer appreciates the way in which we have addressed the complexity and broad applicability of the proposed method, and how we have contrasted it with other existing models and much more different literature on this subject matter.

Here are a few comments that I believe would add more info to the readers:

Thanks so much for having provided these comments – we agree that the way in which we have accounted for them adds more and relevant information to the readers.

- I would suggest using impersonal statements rather than "we".

This is more a matter of style, we think, and we prefer to use the more direct "we" for reasons of clarity. We hope the reviewer is OK with this.

- Literature review in the introduction should be expanded, by including a deeper state-of-the-art.

Very good point. We agree and have expanded and updated our literature review substantially with around a dozen of new references, thereby embedding our work much better in the broader literature and current state-of-the-art, as well as in the wider context of other approaches used to address and analyse this topic.

- At the end of the introduction, a more detailed description of the research gap and the novelty of the paper would be beneficial.

That makes lots of sense. In the revised introduction we have more sharply described the research gap that is being addressed in our manuscript, as well as the novelty of our approach and results in comparison to those presented in other publications using different research methods.

- The methodology is sound, but I would suggest to generate a table with a more detailed description of the boundary conditions used in the simulations.

Very good to read that this reviewer acknowledges the soundness of our methodology. We have expanded and improved the description of the scenarios i.e. simulations and the boundary conditions implied by these scenarios in the revised version of our manuscript, including by referring now to the publicly accessible Zenodo file (https://doi.org/10.5281/zenodo.6782373) in which a very detailed account is given of the assumptions and narratives underlying our scenarios and how these have been implemented in our models. As for the boundary conditions of our models beyond those imposed through our scenario requirements, we also explicitly refer to the

respective published model descriptions that serve that particular purpose, as well as to the publication by Dekker et al. (2023, Nature Energy), in which the carbon footprints have extensively been reported for all the models partaking in our analysis. Given the wealth of information already available in the public domain, we don't find it necessary or desirable to include another table in our paper, given that we have already a large set of tables and figures.

- The references are very few, and must be increased.

As mentioned: we much agree that this was a serious shortcoming of the earlier version of our manuscript. We have now corrected for this (also see our points above).

- Figure 9 needs a more comprehensive description.

Fair enough: we have now included a much more detailed and comprehensive description of Figure 9. We recognize that this Figure, and the way in which it presents some of our key findings, is quite novel, and may not be directly clear at first sight to all future readers of our paper. Hence Figure 9 deserves more explanation: we now do precisely this.

<u>Reviewer #2 (Remarks to the Author):</u>

The paper is reasonable and can attract the readers. Many results are generated, but no explanation of how. In the section "Models and scenarios," no one equation, assumption, etc., is mentioned.

We appreciate that the reviewer finds our work reasonable and sees potential to attract readers. This view suggests, and we agree, that the readership of our work could be large, possibly leading to many citations over the years to come. We do not really agree with the comment that no explanation is provided as to how the results were generated. The methodology section is very comprehensive, provides a clear description of the scenarios, and offers a thorough outline of how the different models were utilized. In an effort to still honor this reviewer's remark, in the new version of our paper we have revised the methodology section and introduction, and improved their clarity (see also our reply to reviewer 1, above). We also now better refer to the publications that provide lots of detail on the individual models and how they operate. Plus we now explicitly refer to the detailed scenario description which is publicly available online (at Zenodo), so that maximum detail and transparency is proffered with regards to how our results have been generated. As for the "absence" of equations, we find that providing the model formulation and objective function of all models employed in this study would damage its readability and broad interest, while not providing significant additional insight. Instead, we have provided extensive references to point the interested (more technical) readers to the right sources of information. These references contain all that these readers could possibly want to know about equations, objective functions, assumptions (both technical and socio-economic) and other relevant model parameters and design features. The feedback by reviewer 2 is surprisingly short, as mentioned already before and as expressed also to the associate editor handling this manuscript. Nevertheless, we entrust that this reply satisfies the reviewer, so that this paper on an important subject for the energy transition and climate change control can proceed in the review process – it would be a pity if the brevity of this remark imperiled the positive and constructive feedback that we received from reviewers 1 and 3.

Reviewer #3 (Remarks to the Author):

This paper presents the results of a self-selected group of collaborating models results on electricity and hydrogen shares in final energy through to 2050 in Europe.

We thank the reviewer for providing detailed feedback on our manuscript. Yet, the models and the modeling groups behind them are not exactly "self-selected": they are the models selected by the European Commission to initiate the European Climate and Energy Modelling Forum (ECEMF). Started as a 4-year EU-funded project, ECEMF is set out to become the prime European platform to host state-of-the art energy and climate modelling tools and results, and to enable rigorous model comparison exercises that yield robust policy and strategy insights. In the future, additional models are envisaged to become part of ECEMF, while as of the writing of this manuscript, the available models are those developed and employed by the ECEMF consortium members. The mandate and purpose of ECEMF are described in our manuscript and in the relevant references therein. Indeed, one of the many subjects that our consortium focuses on is the projection of electricity and hydrogen shares in final energy through 2050 in Europe, as the reviewer correctly observes. Of course, there are plenty of other topics that are being investigated as well in parallel.

The paper has a number of issues.

1. The paper seems devoid of any discussion of other literature and modelling which suggests different and potentially higher roles for gas in final energy consumption, and hence provides no real sense of why the estimates are apparently low (as the paper suggests). A key role for hydrogen depends on its storability and ability to match demand in real time and on the relative price of technologies which favour hydrogen production, such as electrolysers, the price of CCS equipment and methane.

Good point. We have expanded and updated our literature review accordingly, and better placed our analysis within the recent literature (see also our replies to reviewer 1, above). These added references include those that see higher potentials for gases in final energy consumption that can assist in mitigating climate change. Note that the "key role" for hydrogen that reviewer 3 refers to (storage and competitiveness with other technologies) is properly reflected in the energy system models utilized in our study, including all the options that reviewer 3 refers to. This point is better emphasized in the revised version of our manuscript. Indeed, in all our models, all technologies, energy carriers and storage options compete so as to jointly deliver on and meet energy demand in all sectors for all services. Of course, different models possess different time and space resolutions: we now better compare our model types with those that represent different resolutions.

2. The paper nowhere discusses costs or the dependency of model results on assumptions about costs. Prices matter for the percentages of different technologies contributions to FEC. No sense is provided on the extent to which prices matter, see above.

Of course, prices and technology cost assumptions matter a lot for the outcomes of any models, certainly also ours – we fully agree. In most energy system and integrated assessment models (such as used by us) a variety of prices (e.g. energy carrier prices) are endogenous and highly depend on the techno-economic assumptions adopted. On the other hand, technology assumptions, including on costs, usually are exogenous and strongly depend on available external data and on the modelers' assessments regarding how these costs may reduce over time, e.g. through learning phenomena or economies-of-scale. One key added value of multi-model analyses, such as in our study, is that they rely on a range of costs and model formulations, so as to encompass many possible realizations of the energy system. Indeed, in our field, deterministic forecasting (as

opposed to scenario-based projections) is virtually impossible, as has been demonstrated by many examples in the past. The value of our type of cross-model comparison studies (in Model Intercomparison Projects, or MIPs, like ECEMF) is that they scan the solution space of many scenarios that are possible – this is thus exactly what the added value of our current paper is: we leave to individual teams to decide on what they think the best technology assumptions are, so that the diversity in the models that we use together yields a good representation of the entire realistic range in which the feasible energy futures may lie. In this sense, we are confident that the competitiveness of hydrogen in the overall energy system – and, importantly its uncertainty – is well represented in the combination of our models and consequently our results. Note, as we write in our paper, that in our assessment we only consider direct use of hydrogen for energy purposes, hence we do not assess hydrogen penetration via synthetic fuels and the use of hydrogen as feedstock to produce materials (e.g. fertilizers). These two routes could significantly increase hydrogen penetration in the economy. This point is better emphasized in the revised version of our manuscript, and we leave this important subject for future research, by ourselves and/or others.

3. Models likely differ in key characteristics and how these relate shares of electricity and hydrogen. Indeed it seems that most models deliver very low percentages of hydrogen but some deliver significantly more, the paper needs to discuss why, again in the context of trying to understand what might be driving the model results.

We fully acknowledge this point (see also our reply on point 2, above). In the revised manuscript we improved the discussion of the differences in hydrogen penetration, their possible causes and their significance in the EU energy policy context. Note that some models deliver higher shares of hydrogen only in 2050, only for one specific scenario (and not all models do so even there), and only in some specific sectors, while other sectors display a more even spread in hydrogen diffusion outcomes across models. All of this we now better and more accurately formulate in the revised version of our paper (see notably our expanded and more refined text for Figures 8, 9 and 10).

4. The presentation of the range of estimates is not very satisfactory. Model results need to be compared directly as averages across models are meaningless on their own as model results cannot be combined to give averages (as these are not actually the outputs of any model). Some better attempt to present ranges of results from the different models would be more satisfactory.

We present our results mostly in terms of confidence intervals and box-plots, hence fully reporting the ranges in the distribution of results ensuing from the various models, which is the stated purpose of our study (see also our replies above). The central trends are reported as a guide to the eye, and in order to provide an aggregate summary of our findings, in line with common practice in this type of multi-model analyses and as found in many renowned studies, including notably in all IPCC reports. Yet, the considerations by reviewer 3 are now more explicitly discussed in our revised manuscript. We also refer the reviewer to the comprehensive analysis of model results, ranges and characteristics that we recently published in Nature Energy (Dekker et al., 2023) and which we cite in our manuscript. Furthermore, note that in several of the Figures in our paper we do compare the results of different models directly with each other, in the way the reviewer suggests (e.g. Figure 9).

5. Another unexplained point is what is supplying the rest of FEC if it is not hydrogen and electricity, if for instance this is synthetic fuel then this is part of the hydrogen economy.

Good point. The rest of FEC is depicted in Figure 3, as now more explicitly highlighted in our revised manuscript. We also added some text around Figure 10 to more clearly describe the possible role of synthetic fuels. As mentioned, we intend to research in more depth the role of hydrogen in synthetic fuel production in follow-up research.

6. A final question, is what is the paper actually trying to say? It reads like a partial summary of the outputs of one limited set of models with little insight into what might be driving the results. Much better papers summarising net zero model results exist which actually try and outline what the shape of net zero depends on. The simple presentation of quantity shares alone is not very interesting to the reader. A paper with this title needs to do a comprehensive review of all relevant modelling (not just the 8 in the consortium) and then attempt to explain where model results might be coming from and explain differences between models in an insightful way.

As clearly discussed in the introduction of our manuscript, we are analyzing H2-based sector coupling in a decarbonized Europe. We do this on the basis of the energy system and integrated assessment models that have been selected to form the founding basis of ECEMF. We include in our manuscript a comprehensive review of the most relevant literature on the subject, in order to provide adequate context for our study and identify the research gap that we are trying to bridge. However, this is not a review article that attempts at reporting all modelling results available on the subject (as the reviewer seems to hint): the selected models represent the state-of-the-art when it comes to energy modelling in Europe, as per the ambition and mandate of ECEMF. It is impossible to include all models available for studies like ours: even the IPCC (which has explicit task to summarize research results of available models comprehensively) does not and is unable to do this. Yet, to better explain this, we thoroughly revised the introduction of our manuscript and expanded the literature review. We remain confident that the results and their discussion provide many novel insights and adequately address where the differences in model outcomes may come from, as well as what their significance is within the European energy and climate policy context.

Detailed comments:

Line 174-75, blue and green hydrogen compete with biomethane.

Absolutely. Blue and green hydrogen competing with biomethane is exactly what all of our models simulate, and the outcomes that we present is precisely the resultant of this competitive behaviour. We don't exactly understand why this reviewer raises this issue for lines 174-175, but we entrust that this is clear now.

Line 243, hydrogen's role in synthetic fuel is what?

As we explained above and in the text of our manuscript, in our paper we do not investigate the role of hydrogen in synthetic fuel production, as we only analyse its direct use in the sectors mentioned. As we clearly specify, once one starts including this additional potential usage of hydrogen, one is likely to find substantially higher shares of hydrogen in final energy consumption than we currently report. This we leave for future more detailed analysis.

Line, 329, most models show zero or near zero hydrogen, but some show a lot more, why?

Note that all models show zero to near zero hydrogen use in the early years of our simulation (2020 and 2030), so there is full consistency there. Only in later years, and 2050 in particular, one sees some relatively small differences emerging in the levels of hydrogen usage between models, as indicated in Figure 8. We have added some text here, so as to make sure that Figure 8 is understood the way it should be. Note that the higher hydrogen percentages only occur in one specific scenario, for only three models, and then only in some specific sectors. When we report a share in FEC for hydrogen in 2050, we are talking about the average over sectors, as explained.

REVIEWER COMMENTS

Reviewer #3 (Remarks to the Author):

The authors have generally done a good job responding to my comments.

However the abstract and the introduction need to make clear that while H2 has a small modelled share in final consumption, it may have a significantly larger share in intermediate consumption. The hydrogen economy is not just about final consumption, but about the use of H2 as inputs to synthetic fuels and ammonia etc.

I made the point about the lack of economic data comparing the models, this has not been incorporated into the current version of the paper. I would like to see some basic comparisons of assumptions about methane and carbon prices in the eight models, as this might be driving some of the differences in the models.

Three of the models explicitly dont allow a role for hydrogen in at least one major sector of the economy as per Table 1. There is an issue as to whether these models are biased against hydrogen and should be dropped from the comparison.

REVIEWER COMMENTS

Reviewer #3 (Remarks to the Author):

The authors have generally done a good job responding to my comments.

Many thanks, this is very good to hear. Great to read that reviewer 3 thinks we have generally done a good job in responding to his comments.

However the abstract and the introduction need to make clear that while H2 has a small modelled share in final consumption, it may have a significantly larger share in intermediate consumption. The hydrogen economy is not just about final consumption, but about the use of H2 as inputs to synthetic fuels and ammonia etc.

Faire enough. We agree and have made clearer at several places in the manuscript that, while H2 has a small modelled share in final energy consumption, it may have a significantly larger share in intermediate energy consumption. We had already specified this in the previous version of our text, but agree that this is an important fact, hence needs to be iterated several times, which we now do. It is exactly as the reviewer writes: Indeed, the hydrogen economy is not just about final energy consumption – even while that is the focus of our current paper (as now more clearly announced in Abstract and Introduction) – but about the use of H2 as inputs to a broad range of activities, including the production of e.g. synthetic fuels and ammonia. We have thus added more nuance to our claims, and added several specific pieces texts in both the Abstract and Introduction, as this reviewer suggests. Note also that in the text accompanying Figure 9 we had already extensively described, and tentatively quantified, the potential indirect role of hydrogen, hence in intermediate energy consumption, precisely as this reviewer has in mind.

I made the point about the lack of economic data comparing the models, this has not been incorporated into the current version of the paper. I would like to see some basic comparisons of assumptions about methane and carbon prices in the eight models, as this might be driving some of the differences in the models.

Yes indeed, we have exchanged on this subject also in the previous review round. Many thanks for coming back to it, so that we can further clarify. Apologies for perhaps not having been sufficiently complete in our previous answers and remarks.

With regards to the carbon prices in the eight models, please note that these are exogenous (at least in the two carbon price scenarios described in our paper), so these are by definition the same for all models. We have checked once more our manuscript to make sure that this is entirely clear to any (type of) reader. Hence no further comparisons are needed here, since the carbon prices across models are identical. In fact, the very purpose of this scenario exercise was to see what the differences are as a result of identical carbon prices across a set of representative and well-known integrated assessment models in our field of work.

Hence let us here further elaborate on methane prices, the other point that this reviewer refers to when mentioning that (s)he would like to see basic comparisons of assumptions across the eight models, as these might be driving some of the differences in the models. With respect to the latter point, yes of course: our models are all cost-driven, hence costs and prices indeed drive many (but not all) of the differences between models.

At first sight, we thought that this reviewer refers to the price of methane (CH4) as a greenhouse gas (GHG), the emissions of which contribute, of course, to climate change (global temperature

rise). In the three presented scenarios, however, we only apply prices to CO2, and we do so in all regions of the world (not just for Europe, as reported in our paper). We do not levy taxes or induce prices on CH4 emissions; therefore, we cannot say anything about CH4 emission prices in the presented scenarios, nor of the effect that such CH4 prices may potentially have. Rather, our eight models make different, mostly exogenous, assumptions on how emissions of CH4 decline over time, so as to eventually lead to their contributions to climate neutrality. Hence, CH4 emissions as GHG driver is not subject of study or active dynamics in our models, and thus not of our analysis. For that reason, CH4 GHG prices are not reported in our scenario database / explorer, even while some models may generate them.

At closer inspection and after having discussed internally, however, we came to the conclusion that this reviewer probably means something quite different here: methane prices in that interpretation, as probably intended by him/her, could refer to those of methane as an energy resource, for instance for the production of H2 through steam methane reforming (SMR, in principle with our without CCS, generating blue or grey H2, respectively). For this interpretation our reply goes as follows (entire next paragraph, including three bullet points).

When talking of the prices of methane as fuel, in addition to the comment this reviewer makes in this review round, we went back to the comments (s)he made in the previous review round, and based on the combination of those, we deduct that with methane (s)he probably refers to natural gas as an input for the SMR process. (S)he appears concerned about the low share of hydrogen in final energy consumption (see our replies to the previous comment, which we have accounted for in the text, in yellow) and asks for price assumptions regarding green and blue hydrogen production (hence referring in the first review round to "electrolysers" and "the price of CCS equipment and methane"). With this interpretation, the following replies come to our mind:

o The choice of hydrogen or other alternatives in final energy consumption does not necessarily follow the price of hydrogen producing processes (also other factors are at play, e.g. potentials, that affect deployabality and diffusion speeds). To compare the costs, one would need to compare the cost assumptions of all main energy carriers in the final energy mix, which goes beyond the scope of our paper. For instance, for the transport sector, we would need to compare prices of electricity, diesel, hydrogen, kerosene, synthetic diesel, biodiesel, synthetic kerosene, bio kerosene, ammonia, etc. This is precisely why we have our models, to undertake exactly this task, i.e. make such detailed and elaborate comparison through the principle of cost-minimization. While one can do such comparison in a smaller scope, for instance within a limited set of only a few technologies, the purpose of our research is to do so across all final energy carriers. Each model in this ensemble has its own cost (and potential etc.) assumptions that can vary in multiople respects. For instance, see the price assumptions regarding electrolysers that are provided in Table 3 of this paper: https://doi.org/10.12688%2Fopenreseurope.15590.2 (which we refer to in our own paper). In other words, given the sophistication and high technological detail of all our models, making the cost comparison that we think this reviewer suggests, is not only very laborious and in practice hardly achievable (the number of technological options is simply too large, the simulated energy system too complex, and the sets of technologies often not necessarily directly inter-comparable, because different models possess varying degrees of technological detail), it also does not really add much, because our cross-model comparison exercise itself is the tool precisely to do such a comparison of costs across all technologies adopted in the respective models.

o Moreover, not all choices in our models are driven purely by costs, as they can be limited by constraints as well. For instance, in all models in this ensemble, the CCS potential in Europe is limited to 80 and 500 Mt by 2030 and 2050, respectively. Therefore, the choice of using CCS is not only driven by the CO2 price, but also the assumed CO2 storage potential. Hence, even if one were to do the nearly impossible task of comparing all input technology input costs across models (since different models also possess differences in terms of the richness in technologies covered), then still

there would be assumptions on e.g. technology potentials (and other technology characteristics and specifications) that would overrule cost assumptions.
o Different models adopt differences in terms of exogenous natural gas price developments, and these are not reported in our (public) scenario database. In our TIAM-ECN model, for example, the natural gas price follows a business-as-usual scenario with a slight increase over time.

In conclusion, we think that we provided all the relevant economic data that are needed to compare the outcomes of our eight models, and we think there is certainly no lack of economic data in describing our models more generally, since these have been extensively reported in the several publications from our ECEMF project (that we explicitly and several times refer to; see e.g. Dekker et al., 2023) as well as in the publications specifying the nature and features of each of the eight models that we use (which we extensively refer to as well in our present paper). Hence we believe that these valuable remarks by the reviewer have been incorporated into the current revised version of our paper.

Three of the models explicitly don't allow a role for hydrogen in at least one major sector of the economy as per Table 1. There is an issue as to whether these models are biased against hydrogen and should be dropped from the comparison.

True, and good observation: this is indeed what Table 1 shows. In fact, in each plot we only include those models that actually report the specific variable(s) being plotted, i.e. models that do not allow direct penetration of H2 in a certain sector are automatically excluded from the corresponding plots. So we do tell the complete story here. Whether some models should or should not be expanded to include H2 in more sectors is a subject for ongoing discussion; we think it is too early to conclusively confirm whether these models are biased against hydrogen, since we know as of yet still too little about where and in what sectors broad H2 diffusion will actually materialize. We explain this in the text, and in the new version of our manuscript we ascertain that it is well emphasized. Note that in the final plots (e.g. the contour plot) we exclude from the comparison those models that do not have direct H2 penetration, exactly for the reason highlighted by the reviewer (since there may perhaps be a potential bias in these models against hydrogen usage), though we report several full plots (hence with all models represented) in the appendix for completeness.

REVIEWER COMMENTS

Reviewer #3 (Remarks to the Author):

I appreciate the effort the authors have made to respond to my comments and thank them for the extensive response.

However, they raise issues in their response which strengthen two of the points that I made that the paper itself should be explicit about.

First, the extensive response on assumptions lying behind the models is revealing. What you are saying is that there are substantial underlying differences in price assumptions. While I take the general point that there are a lot of assumptions in any model and it is not possible to document all of them in short paper, I am asking the simpler question which is can you separate out key model assumptions which might drive model results. In my view this is a key value of comparative modelling such as yours, otherwise you are just presenting a range of results without going into the obvious question, which is: what is driving the differences? Since you draw high level policy conclusions based on your results, it is perfectly reasonable of the reader (me) to expect you to give straightforward insights into what drives model results differences (and not expect me to read a load of other papers!).

To address this you can simply add a table which says what basic assumption is made about the price of methane (in 2050) and what is assumed about whether carbon prices are applied to methane. I am aware of recent modelling which goes precisely into the fact that assumptions about methane chain leakage and carbon pricing drive both quantity and source of methane production. Indeed it is useful exercise, for your models, more generally to provide a single table which details key pricing assumptions, as this would seem to be revealing and of course gives real colour to what the models might be saying about shares in particular futures. Collectively, you might be able to think of a short list of variables which you believe to be key. I think methane price and effective carbon price (base price times coverage) are two key ones.

Second, I appreciate the point about noting that the models include hydrogen and also that the group of authors may not agree on this issue. However, since you do collectively draw a conclusion on a limited role for hydrogen on basis of leaving it out of some sectors and of not modelling its intermediate role, I think you do need to explicitly say in the conclusions and the abstract that models which allow hydrogen in every major sector do find more use of hydrogen(?), otherwise a key headline conclusion that you currently draw is not actually in line with what the modelling suggests.

REVIEWER COMMENTS

<u>Reviewer #3 (Remarks to the Author):</u>

Our replies are in bold italics.

I appreciate the effort the authors have made to respond to my comments and thank them for the extensive response.

Many thanks. We have profited a lot from this reviewer's comments, and thank him/her for this constructive and insightful exchange, which our manuscript has benefitted enormously from.

However, they raise issues in their response which strengthen two of the points that I made that the paper itself should be explicit about.

We hereby make another attempt to further clarify, in the hope that we come to a mutual understanding of what our analysis is about, and what types of things our models are projecting and capable of doing, while being unfit, at least for the moment and at the current state-of-play, for inspecting certain other dimensions of this subject.

First, the extensive response on assumptions lying behind the models is revealing. What you are saying is that there are substantial underlying differences in price assumptions. While I take the general point that there are a lot of assumptions in any model and it is not possible to document all of them in short paper, I am asking the simpler question which is can you separate out key model assumptions which might drive model results. In my view this is a key value of comparative modelling such as yours, otherwise you are just presenting a range of results without going into the obvious question, which is: what is driving the differences? Since you draw high level policy conclusions based on your results, it is perfectly reasonable of the reader (me) to expect you to give straightforward insights into what drives model results differences (and not expect me to read a load of other papers!).

Yes, indeed: all of this is of course entirely reasonable and we much agree. As we explained before, and below here again, there are substantial underlying differences in price assumptions, especially if they are determined endogenously, for instance for methane, as we explain here again (next point and our reply thereto). It does not apply to carbon prices, however, since these are determined exogenously, and are adopted uniformly across all models (see also below), by the design of the current study. The answer to the other question in this comment is, yes, we can separate out key model assumptions which might drive model results. Obvious examples are population and GDP growth projections until 2050. Since these are key to any of the type of projections and scenario runs that we make, we have made an effort to harmonize them according to SSP2 (that is, the second Shared Socioeconomic Pathway, for which we refer to Fricko et al., 2017). Hereby we achieve providing identical drivers for energy demand in all models. All of this is explicitly mentioned and described in the text of our manuscript. We much agree that evaluating key model assumptions that drive results is a key value of comparative modelling such as ours: indeed, otherwise we are just presenting a range of results without going into the obvious question: what is driving the differences? All of this is at the core of our daily inter-model comparison activities in a project like ECEMF, and the current paper is the fruit of such activities and collaboration between a representative set of institutes with modelers like ourselves. In the new version of our paper, we have now substantially reformulated and expanded upon our high level policy conclusions (see also below), and we hope that our new text and the present set of replies explain better where we have harmonized between models and at what levels differences can be expected between model results and the reasons for these differences (e.g. sector-coverage for H_2

use and differences therein across models, and direct versus indirect use of hydrogen, the latter still being under-represented in the current versions of our models). We also hope that the reviewer can now appreciate these matters better, without the need to 'read a load of other papers'!

To address this you can simply add a table which says what basic assumption is made about the price of methane (in 2050) and what is assumed about whether carbon prices are applied to methane. I am aware of recent modelling which goes precisely into the fact that assumptions about methane chain leakage and carbon pricing drive both quantity and source of methane production. Indeed it is useful exercise, for your models, more generally to provide a single table which details key pricing assumptions, as this would seem to be revealing and of course gives real colour to what the models might be saying about shares in particular futures. Collectively, you might be able to think of a short list of variables which you believe to be key. I think methane price and effective carbon price (base price times coverage) are two key ones.

Fair enough, in principle, and we have been extensively considering to add such a Table along the lines suggested, in order to specify the price of methane and what is assumed about whether carbon prices are applied to methane. However, for each model the "price" of methane is not a single number, but will vary by region and time-period. In general, this "price" is in fact a "shadow price" – hence calculated endogenously – which strongly depends on many techno-economic assumptions and scenario constraints in each model. Among the many assumptions are the costs of the extraction of fossil fuels, which for non-global models might include (amongst many others) exogenous assumptions on the production costs or trading prices of methane (which, again, would vary by region and time-period, as well as across models and scenarios). The very same observations apply to all fossil fuels, hence not just methane, and they apply in general to essentially all primary energy carriers in our eight models. For these reasons we reason in the following way: (i) a Table with the price of methane in each model would be pages long, and, due to its high disaggregation, would not provide much insight; (ii) compiling such Table would require months, as most models currently do not have these data ready to report; (iii) focusing exclusively on methane would trigger a significantly biased interpretation of our model results, since it begs the question "how about the role of other primary carriers?". For these reasons, we conclude that reporting methane prices would not add much value to the manuscript, and it would also be out of scope, both content- and time-wise, under the current state-of-affairs.

With regard to carbon prices: for our study this price should be interpreted as an exogenous tax on CO_2 emissions. Also, we are not assigning or monitoring prices of other GHGs. All of this is clearly stated in the manuscript, and we have ascertained once more that we do so correctly. Indeed, we report the exogenous CO_2 prices in detail, and explain how they increase from 0 to 80 \$/tCO_2 before 2040, and increase with a 5%/yr rate after 2040, in the C0-80 scenario; grow linearly from 130 to 580 \$/tCO_2 between 2025 and 2050 in the C400-lin scenario; while remaining zero in the NPI scenario. To make this entirely clear, we have added in the caption to Table 2: "The main features of each of our three scenarios are specified, where applicable, with the corresponding exogenous CO_2 price trajectory over time (in \$/tCO_2)".

Second, I appreciate the point about noting that the models include hydrogen and also that the group of authors may not agree on this issue. However, since you do collectively draw a conclusion on a limited role for hydrogen on basis of leaving it out of some sectors and of not modelling its intermediate role, I think you do need to explicitly say in the conclusions and the abstract that models which allow hydrogen in every major sector do find more use of hydrogen(?), otherwise a key headline conclusion that you currently draw is not actually in line with what the modelling suggests.

Yes, this is a good point. This is true of course, and we may indeed not have been sufficiently clear about this matter in the previous version(s) of our manuscript. We appreciate the reviewer's

suggestion to try and tease out the key features that are likely driving the H_2 penetration levels in each model. But note that much of this has already been investigated in Dekker et al. (2023), specifically with the ED5 indicator in the "model fingerprints" framework. We refer to this article several times in our own present paper. More in detail, Table SI5 in Dekker's article shows that low shares of direct H_2 penetration are typically associated with high shares of fossil fuels in primary energy, while high H_2 penetration is usually accompanied with high shares of renewables in primary energy. However, one can also observe several exceptions to this overall trend in that Table, hence we conclude that there is no specific set of model features that clearly correlates with the level of direct H_2 penetration. It is the complex ensemble of assumptions embedded in the model topology, its techno-economic database, its scenario parametrization and its objective function optimization scheme that ultimately determines the outcomes, with regards to H_2 penetration as well as any other result. As correctly observed by the reviewer, if the model topology excludes direct H_2 penetration routes in certain sectors (i.e. for Euro-Calliope, IMAGE and WITCH, as shown in Table 1 of our manuscript), this may create a bias towards lower levels of H_2 use. This holds true if H_2 use is accounted for in terms of direct penetration in Final Energy Consumption (as done in our present study), but may well prove not to be the case if secondary uses of H_2 are accounted for (as we clearly state at several occasions in the manuscript).

This leads to a few important observations. 1) A follow-up study is needed in which indirect H_2 penetration is thoroughly accounted for. Note that, due to the diversity in modelling H_2 conversion routes across models, this proves to be a complex task, which deserves a separate, dedicated piece of analysis. We have just embarked on such research, but we expect to be busy with that for at least another year, if not more. 2) The diversity in approaches to modelling H_2 reflects an overall uncertainty among modelers and analysts on the role this energy carrier is expected to play in future energy systems. Possibly, we will see this uncertainty gradually being reduced during the years to come, as more scientific knowledge about the use of H_2 becomes available, and as real-life insights are gained based on the practical implementation of the use of H_2 and the lessons are internalized that will accompany H_2 diffusion. We have now more clearly transformed this observation into one of our central conclusions of our article. See notably our rewritten Abstract, which we now end by stating precisely this: "We conclude that the variance observed across results from different models reflects the uncertainties that abound in the shape of deep decarbonisation pathways, in particular with regard to the role of H_2". 3) The modelling community may benefit from a set of streamlined rules for properly accounting H_2 penetration in the energy system. This is a topic that the modelling community is currently discussing. Before this discussion has been settled amongst the energy system and integrated assessment modelers, however, we may be nearing the stocktaking for AR7 of the IPCC.

Back to the reviewer's first remark on this matter: it is not necessarily always the case that our group of authors does not agree on the issue of how to include hydrogen in our models. Rather, the variance that we observe is partly at least also a reflection of the fact that we haven't yet spent the time and efforts needed to fully understand the subject and draw the necessary conclusions so as to update all models in our field to a unique set of minimal criteria with respect to the simulation of H_2 in the main sectors where currently analysts think that its use may become substantial. In other words, the topic is still relatively new, so that not all ramifications of the possible uses of H_2 have settled into the practice of our modelling community. In the Abstract we have thus made the necessary change, as this reviewer suggests. Also in the Discussion section of our paper we have introduced several changes so as to account for these remarks, justified as they are. In the Discussion, we now explicitly write, in new pieces of text: "After all, while at present we collectively draw the conclusion that there may possibly remain a limited role for the direct use of hydrogen to satisfy final energy consumption, we acknowledge that this conclusion is based on the fact that some models leave it out in certain sectors, as a result of which hydrogen may be somewhat under-represented. If indirect or intermediate uses of hydrogen are accounted for in all models – with still

unknown effects on its direct usage – its aggregate role is likely perceived larger. Projecting the mid-century extent of aggregate hydrogen usage is a subject that deserves concerted attention by researchers and modelers like ourselves, as well as the broader policy making and private sector communities". We entrust that this addition reflects that we largely agree with the opinion expressed by the reviewer under this remark.