-**Reviewer 1**

- This is an extremelly well written and well organized manuscript, very much in line with the scope of Environmental Modelling and Sofware. Emissions harmonization is a key issue for Integrated Assessment Models (IAMs), which most of the time have to deal with the difficulties of harmonizing historical emissions with model inputs (and sometimes even with model results). So, this paper is a very welcome addition to the literature, as it presents a new methodology for automated harmonization with a common historical data set of generic emissions trajectories of IAMs, proposing a framework and open-source software approach for that. As discussed in the paper, the methodology has proved to be extremely successfull in its objective of harmonizing emissions trajectories, when 2,000 emissions trajectories where tested for that.

The only two minor comments this reviewer would like to make are:

1. On page 15, lines 195-200, please justify, in the harmonization diagnostics, the proposed range used for the difference between harmonized and unharmonized trajectories (the 400% and 200% figures)

We have revised lines 195-200 to clarify that the values used to determine which trajectories are reported in the diagnostics user-definable and that values referenced in the paper are simply defaults that proved to be useful in initial usage and testing of the aneris package.

2. On page 19, lines 220-225, please provide the RCP that corresponds to SSP2-Ref

While SSP2-Ref does not correspond to a specific RCP (in fact, it lies between RCP-6 and RCP-8.5), we have added the following clarifying text:

> This SSP2 scenario lies between two RCPs, 6 and 8.5, with a radiative forcing level of approximately 6.5 Wm −2 in 2100.

Except for that, this reviewer is of the opinion that this manuscript can be accepted as is.

-**Reviewer 2**

-

General Comments:

The authors' aim in this paper is to develop a user friendly and common methodology to harmonize GHG emissions (model and historical) for modeling teams. They describe an algorithm in the main article and provide it in github for future users.

The writing in the paper is impeccable in terms of grammar and style.

This is a well-done article, but it needs to be revised to be clearer to readers that are not familiar with these IAM models. Authors need to explain data sources, assumptions, models either in an Appendix or in Footnotes. The paper needs to be self-contained and cannot be an attachment to other work being conducted with these models/scenarios.

We have striven to add additional information where possible (and even where slightly tangential). We have added a number of footnotes that provide additional explanation of topics core to the science provided by IAMs and have provided additional references for readers to investigate if they so desire.

The authors need to add the intuition behind harmonization process. Describing only methodology is not enough. The examples need to be increased and discussed better.

The authors agree that a better explanation of the intuition behind the harmonization process could be provided. Accordingly, we have added an additional subsection in our Methodology section denoted “The Conceptual Basis for Choosing Harmonization Methods”. We have placed this new section before a general discussion on available methods and our discussion of developing the method decision tree. We believe this will provide readers with a foundation for why various default methods are chosen at each branching point.

The authors need to add a section on Python code, explaining briefly the main steps. They need to explain the user-friendly features, and how others can use the code. Linking to github is not enough.

Section 2.4 (previously Section 2.3) has been largely reworked to include a significant discussion on the Python implementation. It now includes a new figure explaining the Python implementation, a discussion of that figure and the software components, a revised Figure 5 (now 6) complementing the new figure, and relevant code snippets. User-friendly features are highlighted.

Overall, this is a beneficial methodology and can help many analysts.

Specific Comments:

1. In keywords, please add *GHG*. *Climate Change* may not be the most appropriate keyword for this paper.

This has been updated as requested by the reviewer.

2. On Page 7, “Methodology & Implementation” section, please explain and provide Equation 1 before other Equations.

We have reordered the text in this section to highlight Equation 1 and clarified the importance of the convergence factor.

3. On Page 7, when presenting Table 1, explain why some options do not include Convergence. Main discussion in text is much later. All concepts introduced by authors need to be defined in a flow so as not to confuse reader.

The lack of a convergence year (i.e., a convergence year limit of infinity) has been clarified in the table caption and is discussed in the paragraph concerning the definition of the convergence factor.

4. Figure 1 and discussion surrounding Figure 1 does not consider sectoral and technological specifities enough. These decision branches are very modelling oriented. It is not enough linked to the complex topic of climate change.

The decision branches are oriented towards both modeling concerns as well as sectoral as well as overall scenario design. We have introduced a new section entitled “The Conceptual Basis for Choosing

Harmonization Methods" which is designed to address this point. We go into more detail in that section; however, we reproduce the opening sentence below to provide a notion of the section framing to the reviewer.

> *The goals of any harmonization exercise is threefold: aligning model results in the harmonization year to a common historical data source, faithfully representing the original IAM's internal consistency between the driver of emissions (e.g. energy use) and emissions, and maintaining critical parameters from the original scenario design.*

5. In Figure 1, explain that you have years in parentheses, and why.

We have clarified the Figure caption and added explanatory text to the discussion of the figure in the text body.

6. Page 10, add more examples of exceptions where the "offset" option will be used. Link this to issues surrounding to sector/technology ang GHG emissions. This paper should not be only modelling focused. The authors need to put their work in a context.

We explain in the text that the only case in which an offset method is chosen is when the model does not report a given sector/gas combination. Offsets are only applied because a ratio method is not applicable to a model value of 0. These instances rarely, if ever, occur; however, we wish to provide a robust product which must take into account such occasions (we have adapted the text to include this comment).

7. Page 10, the authors write "Models can additionally report negative emissions in certain contexts", please add discussion with examples on *which* contexts.

We have further clarified that negative emissions of CO2 are possible due to specific mitigation technologies such as carbon capture and sequestration (CCS) among others.

8. Figure 2, please explain what these Graphs are based on? An example/an illustration can go a long way.

We have expanded the introductory discussion of Figure 2 in section 2.2 to try to better explain that these trajectories are illustrative examples (not model results) of the types of situations that can arise when attempting to choose which harmonization method to apply to a given emissions trajectory.

9. Page 12, discussion on computation of threshold for cv is incomplete. For readers not familiar with these models and scenarios, the paper and its contents have to be self-contained. Please add details on computation: time period, which regions? Is this global? Which IAM models?

We agree with the reviewer and have expanded the discussion of the cv parameter. We have explicitly listed the IAMs included in the analysis and highlighted that Figure 3 includes realizations of cv values for individual model regions, emissions species, and sectors. We have explicitly noted that the computation is only concerned with the historical time series (not model results) and includes derivative information of the entire time series. We highlight in the discussion of Figure 3 that aggregation to model regions is necessary to provide a best-default for tau_cv in the python module because harmonization occurs on native model regions (not necessarily the spatial level of the historical dataset).

10. Move Equation 5 up to when it is first mentioned

Eqn. 5 has been moved, thank you for pointing this out.

11. Page 19, “Results” section, the authors use a scenario with negative emissions as an illustrative case. They need to explain more why this is a good example, what are we learning, how is this relevant?

We have expanded the introductory discussion in this section to discuss how key indicators derived from model results are affected by negative emissions trajectories. We have furthermore provided explanation as to why these two scenarios are important as case studies because they represent two relative extreme cases in the use of an automated harmonization approach and thus provide a context and frame of reference as to its general applicability.

12. Figure 7, make it more visible that this is global, add the year for 0

We have updated the xaxis labels and have added a title clarifying that these are global emissions.

13. Page 23, line 261, check for grammar

Fixed (also is representative the trends -> also is representative of the trends)

14. Figure 9 is illustrative but too crowded. It is difficult to follow with all regions included. Consider simplifying or dividing.

Response to 14 is combined with response to 15 below.

15. Figure 10, add information on which scenarios are lines and which scenarios are dashed.

We have revised the original manuscript’s Figures 9 and 10 (now Figures 10 and 11) to include one set of data for each region in each panel. We have further unified the colors and style of each line and added a consistent legend across panels in each figure. Finally, we have streamlined the choice of regions shown in each figure to better highlight the key points being made in each section of the case study rather than focus on super-regional trends as was done in the original manuscript.

-**Reviewer 3**

-

The paper is well written, however I had great difficulty understanding it. My area of expertise does not overlap at all with this paper which is probably most of the problem. I’m a software person and a model developer but our models and software are very different to what is being described here. I’m sure there will be many readers of EM&S that will also have similar difficulties understanding this paper.
While I roughly understand what you’re trying to do, I didn’t follow much of the Results section. The entire paper and in particular the results section, suffers from too many acronyms, most of which I’ve never heard of before. For example: MESSAGE-GLOBIOM, SSP, SSP2, SSP2-Ref, SSP2-4.5, RCP, LUC, CEDS, WMO,

CMIP6, IAMC, SRES, LULUC, CPA – Ref, PAS – Ref, SAS-Ref, R5ASIA-45, SAS. One way to overcome this is to not use acronyms for all of these but instead spell it out each time, at least for the less commonly used ones. Overall the results section needs to be written for scientists / software people from other disciplines. I would encourage you to do more 'story telling' (more story and less detail). The results section seems to be a dump of a lot of results without the necessary description to say why you're presenting them. Also, many of the graphs have many lines on them using markers, line types and colors to distinguish between different 'treatments'. They are very complex and hinder your message. Think about ways to simply them.

We recognize that many of the acronyms used, while well understood in the IAM literature, are not likely to be as well understood to a broader audience. To ameliorate this, we have added an acronym table at the end of the paper and have taken care to explicitly spell out all acronyms upon first usage.

I wonder whether the results section should be turned into a case study that describes the problem you're trying to solve and then outlines what you did (with results) to solve it i.e. a more narrative approach. The discussion section could then talk about what worked and what didn't in the case study. This would move the paper away from being a methodology paper but would help greatly with its readability by non-experts like me.

As a modeler and software person, I would also be interested in some of the Python implementation details which weren't discussed at all. Perhaps a short section on this would be good.
More comments / suggestions in the attached, annotated PDF.

Regarding the Python implementation, we refer Reviewer 3 to our response to Reviewer 2's similar comment.

Comment - what is MESSAGE-GLOBIOM

We have clarified that MESSAGE-GLOBIOM is an IAM (i.e., "the IAM MESSAGE-GLOBIOM")

Comment - myriad ways

Here we use myriad in its adjective form (and not its noun form)

Comment - 200/400 %

We refer Reviewer 3 to our response to Reviewer 1's similar comment.

Comment - clarify null trajectories

We have changed the term "null" to "0-valued"

Comment - explain WMO

This has been clarified and added to the acronym information.

Comment - 2 scenarios

We have clairified in the text why the two scenarios are chosen (notably that they provide bounding examples of emissions growth and emissions reductions). We have further clarified in footnotes where readers can find additional information on the full SSP scenario design.

Comment - explain marker scenario

We have added a footnote explaining the term marker scenario with references for further reading on the topic.

Comment - explain radiative forcing

We have added a footnote explaining the term radiative forcing with references for further reading on the topic.

Comment - calculation of 1940

1940 was arrived at by multiplying 970 (trajectories per scenario) and 2 (number of scenarios).

Comment - explain MAGICC

We have added a footnote explaining the what MAGICC is/does with references for further reading on the topic.

Comment - explain model horizon

We have changed the term “model horizon” to “modeled time period” to better clarify the sentence.

Comment - explain Python implementation

Regarding the Python implementation, we refer Reviewer 3 to our response to Reviewer 2’s similar comment.

Comment - fix grammer in line 1474

The sentence has been updated to read: “Of the remaining trajectories, harmonization method overrides were applied, and a discussion was provided detailing why overrides were deemed necessary.”