#Topic editor

The reviewers appreciate the progress made, but also point to needs for further improvement.

[Response] We thank all reviewers for the comments. They are helpful for improving the quality of the manuscript.

In particular, the justification for the use of time-independent MACs is not yet sufficient. For instance, the manuscript does not mention capital stock inertia as a key reason for time-dependence in abatement potential. Given the goal to apply the emIAM for long-term climate analysis, the authors should consider deriving separate MACs for short-term (for, e.g., 2030, when existing fossil capital stocks limit abatement potentials) and long-term (from 2050 onwards, beyond the lifetime of the bulk of current capital stocks).

[Response] We thank the editor for pointing out this issue. In light of the comment from the editor, we have tested time-dependent MAC curves using the data from AIM. We chose AIM because AIM gives markedly different data points in early decades compared to those in later decades, especially in scenarios with low carbon budgets. REMIND is used in the rest of the manuscript as an illustrative case, but we found that REMIND is not suitable for this exploration because the outliers mainly originate from peak-budget scenarios, which may have caused by constraints associated with the net-zero target.

Thus, for AIM, we introduced time-dependent MAC curves for CO_2, CH_4, and N_2O before 2050, when the data points shift with time. We kept the original MAC curves after 2050 because the data points generally follow the same line after 2050. Specifically, we introduced a new parameter e and an additional term into the MAC curves before 2050, as described by the following equation:

$$f(x) = \begin{cases} a * x^b + c * x^d, \ t \geq 2050 \\ a * \left(x * (e * (t - 2050)^2)\right)^b + c * (x * (e * (t - 2050)^2))^d, \ t < 2050 \end{cases} \quad \text{(R1)}$$

We kept the values of the parameters a, b, c and d, which have already been estimated. We estimated the parameter e by considering all data points from 2020 to 2045, with no data points being excluded (e is 0.00124 for CO_2, 0.00098 for CH_4, and 0.00076 for N_2O, respectively). This revised equation better captures both the near-term time-varying relationship and the longer-term stable relationship in the data points (Figure R1).

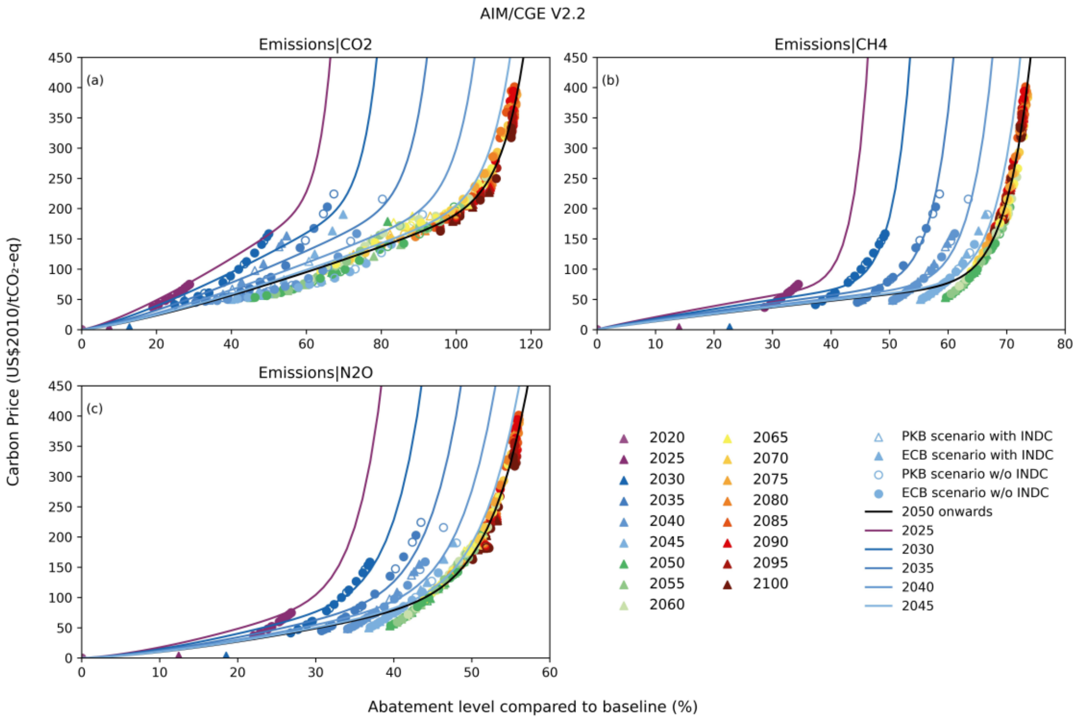
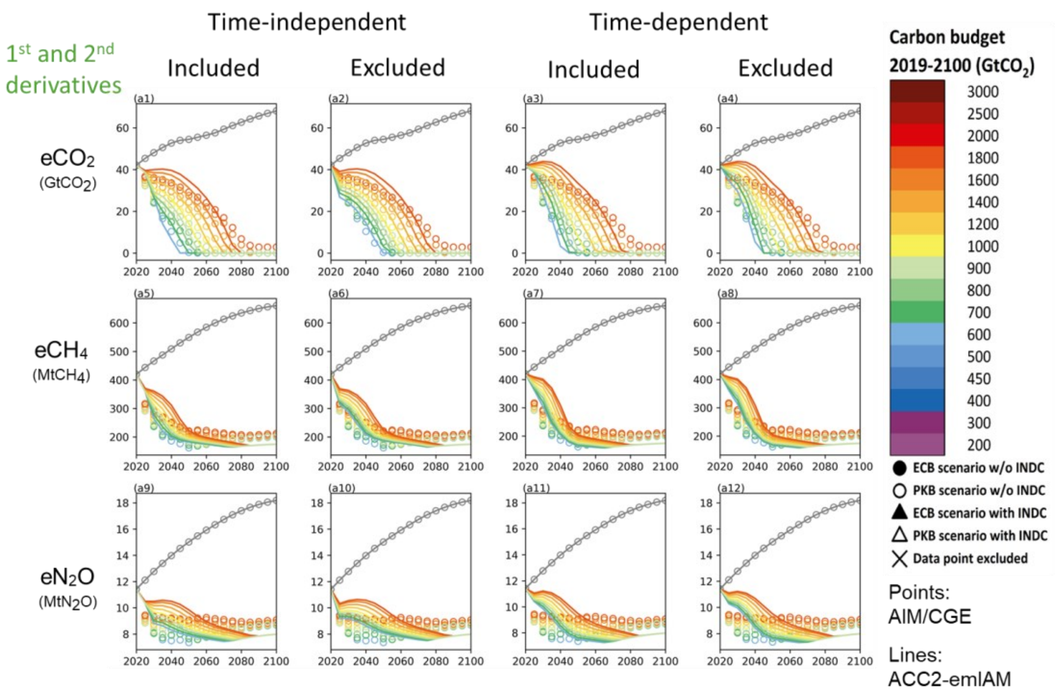


Figure R1. Global total anthropogenic time-dependent MAC curves from the AIM model as an example. The points are the data obtained from AIM in the ENGAGE Scenario Explorer and are shown with colors and markers as designated in the legend; the lines show the MAC curves derived for specific periods. The same color for the points and lines is for the same year. The time-independent MAC curve (black line) is derived from the approach described in the main text. This figure is included as Figure S241 in the Supplement.

Using these time-dependent MAC curves in ACC2-emIAM, we conducted Test 1 (i.e., the constraint on the cumulative emission budget of each gas) to examine the performance of the emulator. As discussed below, we also varied (retained or discarded) the assumptions on the upper bounds of the first and second derivatives of abatement changes (Table 2). Because these upper bounds strongly influence the near-term mitigation levels, we performed an analysis with and without such upper bounds. The results are shown in Table R1 and Figure R2.

Table R1. Validation results of different MAC curve approaches for total anthropogenic CO_2, CH_4, and N_2O emissions derived from AIM. All scenarios are shown here. For the definitions of the statistical indicators, see Section 4.4 of the main paper. This table is included as Table S8 in the Supplement.

Upper bounds of 1st and 2nd derivatives		Time-independent MAC curves		Time-dependent MAC curves	
		Included	Excluded	Included	Excluded
r_P	CO_2	0.986	0.986	0.974	0.979
	CH_4	0.962	0.962	0.94	0.939
	N_2O	0.921	0.922	0.885	0.884
r_C	CO_2	0.981	0.981	0.964	0.964
	CH_4	0.957	0.957	0.927	0.927
	N_2O	0.916	0.918	0.874	0.873
MAE	CO_2	2.595	2.512	3.628	3.609
	CH_4	19.47	19.282	25.5	25.461
	N_2O	0.501	0.486	0.624	0.619
RMSE	CO_2	3.715	3.643	5.234	5.212
	CH_4	27.411	27.014	36.713	36.755
	N_2O	0.653	0.638	0.822	0.822



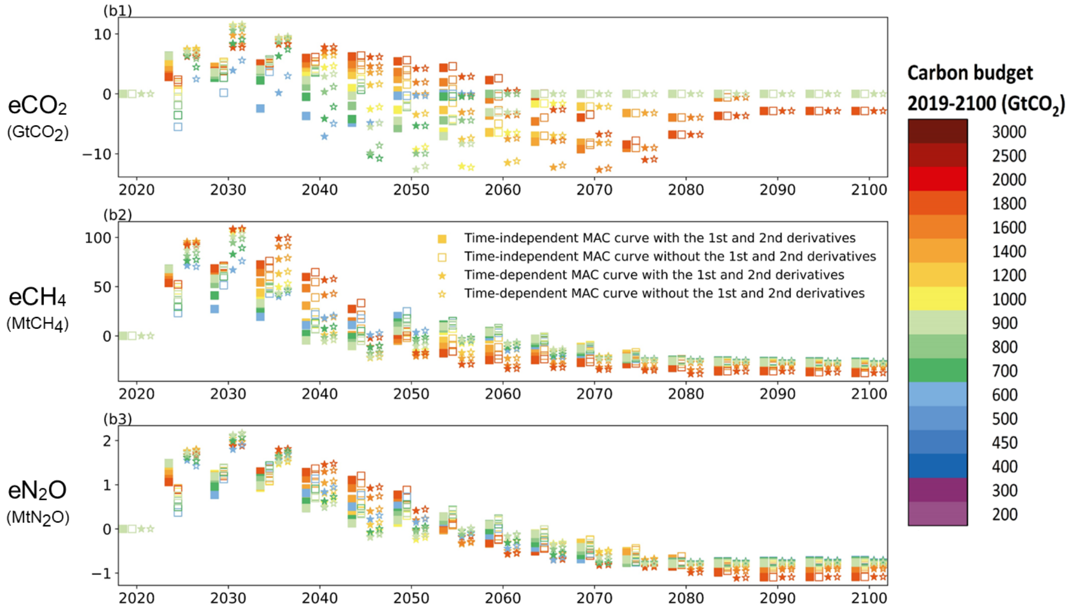


Figure R2. Validation results for ACC2-emIAM with AIM MAC curves. In the top panel, the points show the original REMIND emission pathways obtained from the ENGAGE Scenario Explorer; the lines show the emission pathways reproduced by ACC2-emIAM. The same color is used for each pair of original and reproduced pathways. The bottom panel shows the errors between the outputs of AIM and ACC2-emIAM, which use two types of MAC curves with/without the maximum first and second derivatives, respectively. For the sake of presentation, only the outcomes of the PKB scenarios w/o INDC are shown. This figure is included as Figure S242 in the Supplement.

The results indicate that the revised emulator using the time-dependent MAC curves is not superior to the original emulator with the time-independent MAC curves in terms of the reproducibility. With the time-dependent MAC curves, which penalize more the near-term mitigation, the near-term abatement became more limited for all three gases. This was expected, but in fact, the mitigation up to 2040 became too limited (relative to the output from AIM) because the time-dependent MAC curves are too high at low abatement levels for each period, which makes the near-term abatement more costly than the original model. The time-dependent MAC curves play the dominant role in shaping the near-term mitigation pathways. While the results from the original emulator with time-independent MAC curves showed a high sensitivity to the upper limits of the 1st and 2nd derivatives, those from the revised emulator with time-dependent MAC curves showed almost no sensitivity.

Overall, the time-dependent MAC curves did not improve the reproducibility of the IAM emulator in our example based on AIM. The results seemed puzzling to us at first because we expected an improved reproducibility with the time-dependent MAC curves. However, we came to the realization that the overall performance of the emulator is determined by a complex interplay of various factors, including the MAC curves and the upper bounds of the first and second derivative limits. We agree that the time-dependent MAC curve approach can potentially improve the reproducibility of the IAM emulator (despite our rather negative results) and should be further pursued if the reproducibility of the IAM emulator is the main goal. However, we speculate that the actual advantages of using the time-dependent MAC curves can be model- and scenario-dependent, requiring further analysis.

Finally, we note that the use of the emulators for developing extended scenarios till 2300 is beyond the scope of the current paper (in fact, the paper does not say anything about that). As the editor is aware, we are indeed applying this method to extend emissions scenarios as part of Horizon Europe RESCUE and OptimESM projects. For these projects, we are further developing and fine tuning the emulator specifically to a newer and different version of REMIND-MAgPIE by taking into account the needs for the projects (e.g., explicit treatment of CDR technologies in the MAC curves to produce individual CDR pathways explicitly in our extended scenarios). On the other hand, our current paper aims to develop a more general approach. We intend to test a simple and common approach that could be applied to different models consistently and understand how well the simple MAC representation works for different IAMs under different scenarios.

Nevertheless, we once again thank the editor for suggesting the idea of time-dependent MAC curves. We have gained better understanding for some more complexity behind time-independent and time-dependent MAC curves. Since this issue of time-indepedency is important and also raised by the reviewers, we added a new subsection (4.5) to include the discussion above in the main text in a shorter form so that follow-up studies can be conducted potentially using different scenario data from different models.

Please also carefully consider the other remaining comments raised by the reviewers.

[Response] We further revised the manuscript based on the reviewers' comments. Please see below our point-by-point response. We hope that we have addressed all reviewers' comments and that our manuscript will be accepted for publication.

#Reviwer 1

The manuscript was substantially improved after the first-round revision and addressed most of the issues raised by the reviewers. I recommend that a version close to this one be accepted for publication. However, I still have one concern and three minor comments, as follows:

[Response] We appreciate the reviewer for the comments. They were very useful for improving the quality of our manuscript.

The authors simplified the MACCs by ignoring the temporal effects along the evaluation period, such that it can make the analysis more tractable and easier to communicate. In some cases, the assumptions necessary for a time-dependent MACC may introduce complexities that do not significantly improve the accuracy of the assessment. But it is not so clear to me in lines 686-687 (for the time-independent MAC curves): "A plausible explanation is that the use of percentage abatement levels relative to rising baseline can offset the effect of lowering mitigation costs through learning." It might relate to the learning costs; on the other hand, if there is a high degree of confidence that technology costs will not vary significantly over time, a time-independent MACC may be a reasonable assumption.

[Response] The issue discussed above is in some way explained by learning. As the reviewer pointed out, it is also true that if technology costs will not vary significantly over time, a time-independent MACC can be a reasonable assumption (under a stable baseline scenario). Given the reviewer's comment, we elaborated the text to the following:

"A plausible explanation is that the use of percentage abatement levels relative to rising baseline can offset the effect of lowering mitigation costs through learning over time. In other words, the higher the baseline scenario is, the larger the absolute amount of emission reduction is (for the same percentage emission reduction). If technology costs will not vary significantly over time, a time-independent MAC curve can be a reasonable assumption (under a stable baseline scenario)."

For minor comments:

1. In Figure 2, equations are not recommended to put in the caption text. Please move them to the methods.

[Response] We have moved these equations to a suitable place in Section 3 and changed to the following:

"We also calculate the confidence intervals of the fitted curves using $\hat{y} \mp t_{\frac{\alpha}{2}} * S_{\varepsilon} * \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}$ *(Thomson and Emery, 2014), where* $S_{\varepsilon} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y})^2}{n-2}}$*,* n *is the sample size,* $t_{\frac{\alpha}{2}}$ *is the critical value of t-distribution,* $\bar{x}$ *is the mean of samples,* $\hat{y} = f(x)$*, and*

x_i, y_i *are the original abatement level and carbon price result from the IAM, respectively.*"

2. Please check the journal requirements for the layout of table. The table caption text is on top of the table, rather than under the table.

[Response] We have moved the caption text of all tables to the top of the tables.

3. For the summary of the results (lines 664-687), please re-order the lists (the estimation of MAC curves should come first, then the reproduction results), which can be consistent with the context of the manuscript.

[Response] We have re-ordered the first two summary points for the estimation of MAC curves first and the reproduction results later as below.

- *Certain data points were difficult to capture by MAC curves. In particular, PKB scenarios with low carbon budgets can give very large carbon prices in the near-term. Such data points tend to deviate from the trend of other data points and were manually removed from the MAC curve fitting where appropriate (Figure 1 and Table 1). Except for these "outliers," no discernible difference in the data trend was found between ECB scenarios and PKB scenarios, supporting the use of common MAC curves for ECB and PKB scenarios. Note also that certain data points from GET at high abatement levels do not follow the trend of other data points and were also removed from the MAC curve fitting where appropriate. We speculate that these data points are affected by the limit on CCS capacity assumed in GET.*
- *Some IAMs were more easily emulated than other IAMs, reflecting specific model features such as solution methods, technology assumptions, and abatement inertia. The emulator can usually reproduce the emission pathways of an IAM better if the model response to carbon price are well fitted with a MAC function.*

#Reviewer 3

I think the paper significantly improved, primarily with the revision of the figures and streamlining the content. Also, the new figure 10 and discussion of differences between estimates from the emulator and individual models, as well as the inclusion of confidence intervals and the generalization/discussion at the end improved the manuscript. Again, I do think the idea of this emulator is interesting and useful, but I still have a few concerns. I will leave it to the editor to weigh these concerns for a final decision and I am open to go with the consensus of the other reviewers.

[Response] We appreciate the reviewer for the comments, which helped us to improve the quality of the manuscript.

General comments

1. Great that you included the confidence intervals in Figs. 2 and 6. However, how do these ranges propagate in the results of the emulator? In other words, could you also provide such ranges in the results of the emulator? Do they mean that the emulator output becomes very uncertain? I would propose, also to allow better comparison between the dots and the lines, that in Fig. 8 or 9 you omit some of the carbon budget levels (i.e., only focus on a few), and then also add confidence intervals of the emulator's output to get a feeling of how uncertain the output is.

[Response] We are thankful for the reviewer's comment. To estimate the uncertainty in the ACC2-emIAM due to the range of possible MAC curves derived from the IAMs, we utilized the confidence intervals at the 95% level for the fitted MAC curves, as represented as shaded bands in Figs. 2 and 6. In alignment with Fig. 8 in the main text, we take the REMIND MAC curves as an example. We, however, only report the reproducibility results based on the upper range of the MAC curve (95% confidence interval). We tried with the lower range of the MAC curve (95% confidence interval) as well, but we were not able to obtain reasonable results because of the negative segment of the lower MAC curve. The negative segment requires re-defining the problem as a new type of mathematical problem (a discontinuous nonlinear program (DNLP)), which either made it too complex to solve in our GAMS CONOPT3/4 computational environment or made the optimal solution unreliable (i.e., the solution becomes dependent on initial conditions).

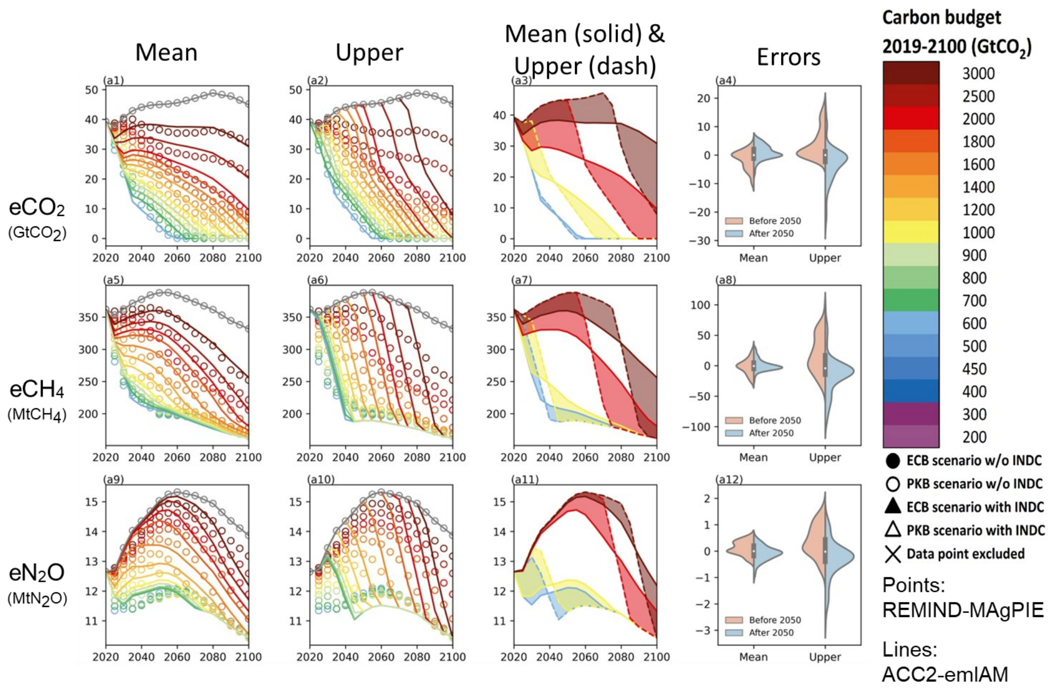


Figure R3. Validation results for ACC2-emIAM with mean and upper MAC curves from REMIND. The points show the original emission pathways from REMIND obtained from the ENGAGE Scenario Explorer; the lines show the emission pathways reproduced from ACC2-emIAM by using the mean (the first column) and upper (the second column) MAC curves. The third column presents the uncertainty (shaded band within two emissions pathways) of ACC2-emIAM by using different MAC curves (only cases with the carbon budgets of 600, 1000, 2000, and 3000 $GtCO_2$ are shown here). The fourth column shows the errors from the reproduced scenarios (ACC2-emIAM) relative to the original scenarios (REMIND). Positive values indicate ACC2-emIAM gives higher estimates than REMIND and vice versa. The same color is used for each pair of original and reproduced pathways. For the sake of presentation, only the outcomes of the PKB scenarios without INDC are presented. This figure is added to Figure S243 in the Supplement.

The reproduced results using the default MAC curve and the upper MAC curve (95% confidence interval) are compared in Figure R3. We found that the use of the upper MAC curve weakens the emulator performance, as most clearly indicated by the abrupt emission declines for all gases, which did not occur in the original scenarios. This points to the need for assessing how to make use of the uncertainty in the MAC curve. It is also an issue of interpretation how the uncertainties in the MAC curves can propagate to the uncertainties in the reproduced scenarios gererated by new optimizations. The uncertainty propagation is different from more intuitive, forward uncertainty propagations, such as those along the cause-effect change of climate change: emissions → concentration → forcing → temperature change → impacts (e.g.,

Figure 8.27 of IPCC AR5 WGI Chapter 8). While analyses could be extended for other models and scenarios, such exploration falls beyond the primary focus of our paper.

We added the statement below in the main text and put Figure R3 as Figure S243 in Supplement.

"Uncertainty is reported in all MAC curves derived in this study. While such uncertainty is useful to indicate the confidence level of the MAC curve, it is not necessarily very obvious how to make use of the uncertainty range in reproducing scenarios from the IAM emulator (Figure S243)."

2. In your response to my question on time variance of MACs and percentage abatement, you quote the text "The behaviors of IAMs that contain various time-dependent processes were generally well captured by the time-independent MAC curves. A plausible explanation is that the use of percentage abatement levels relative to rising baseline can offset the effect of lowering mitigation costs through learning." I am not an expert on this particular matter, but could you elaborate on this? For example, has it been studied before to what extent, when merely looking at (percentage) abatement levels, time-invariant MACs are fine? I would expect that in the finer details (e.g., lifestyle changes, energy mix), this time invariance does not hold anymore. Also, see comment (3) below on the performance indicators you are using.

[Response] We appreciate the reviewer's thoughts. We agree with the reviewer's point that the time invariance may not hold for individual details. Our argument goes only at the aggregated level (global/regional total emissions from all sectors). While we hope to gain more insight at individual process levels, it is practically impossible for us to investigate how these are represented in each of the ten IAMs. (Note that we are just users of the data from the ENGAGE project. We are not part of the project, so our access to model details is limited.) This is also why we stick to the aggregated level and keep our scope to test how well our simple time-independent MAC representation works for different IAMs. Please also see the related comment from Reviewer #1 (first comment).

Time-independent MAC curves are not new as such and have been applied before. At the beginning of Section 3, we state that "*While MAC curves are more commonly time-dependent or for a specific point in time, time-independent MAC curves have also been used for long-term pathway calculations (Johansson et al., 2006; Tanaka and O'Neill, 2018; Tanaka et al., 2021) and short-term assessments (De Cara and Jayet, 2011)*". However, we argue that our study for the first time extensively applied the time-independent approach for capturing the behaviors of various IAMs. No other studies pushed this approach as far as our study did.

Given the reviewers' comment, we have modified and expanded the text. The revised text reads:

"A plausible explanation is that the use of percentage abatement levels relative to rising baseline can offset the effect of lowering mitigation costs through learning over time. In other words, the higher the baseline scenario is, the larger the absolute amount of emission reduction is (for the same percentage emission reduction). If technology costs will not vary significantly over time, a time-independent MAC curve can be a reasonable assumption (under a stable baseline scenario)."

3. The correlation metrics in Tab. 5 and 6 should be changed. These are correlations over time, I believe? If so, it makes perfect sense that you get high correlations purely because in both cases the emission pathways drop. I would prefer to use root-mean-square-error, to actually get an idea of the error (in average or cumulative Gt CO2, for example).

[Response] It is true that if two variables drop proportionally each time step, we only find perfect correlations. However, the data we are dealing with are not so idealistic and do deviate from this perfect setting, resulting in low correlations in certain cases. We maintain two indicators: i) ordinary Pearson's correlation coefficient r_P and ii) Lin's concordance coefficient r_C. r_P is a commonly used indicator that can be used to test the *strength of linear relationships*, which is a reference for our comparison. However, it is inappropriate for testing *agreement*, making r_C a more appropriate choice for measuring agreement between two sequences.

Following the suggestion from the reviewer, we added two more indicators (i.e., the root-mean-square-error (*RMSE*) and mean-average-error (*MAE*) (see Figures S110-S128, S148-S166, S185-S202, and S222-S240 in the Supplement). We find that these two indicators provide added values as they capture the magnitude of the deviation. However, to avoid making the manuscript even longer, we keep the results of these two indicators in the Supplement.

Minor comments

• I still think that the paper is lengthy. You may want to consider moving some of the text and

figures to a supplementary file to streamline it even more.

[Response] We have thoroughly edited and streamlined the entire manuscript in response to the reviewer's comment for the previous review round. We have once again checked the manuscript to trim down the text for the current review round. However, due to the nature of this paper, we find it impossible to compress the text any more significantly. This is a methodological paper that requires a full description of our approach and underlying data, even though the text becomes lengthy. We thought about moving some text to the Supplement, but we still prefer to keep them in the main paper and leave only additional figures in the Supplement (the text becomes not easily accessible for readers if it is put in the Supplement). It is also our observation that many other papers in Geoscientific Model Development are as long as or even longer than our manuscript. This is ultimately an editorial decision, but we hope that our current paper format is acceptable for the journal.

• Why is only REMIND shown in Fig. 8? Perhaps better to show the averages across all models?

[Response] We appreciate the reviewer for this suggestion. IAMs have very different baseline emissions and mitigation behaviors, which would make the averaged results challenging for interpretation. We take the results for REMIND only as an illustrative example for our verification processes. The results of other IAMs can be found in Figures S91-S109, S129-S147, S167-S184, and S203-S221 in the Supplement.

• Table 6 is unreadable. I propose to make a selection of things to show rather than everything.

[Response] We chose different colors to represent the intervals of specific values. That is, the darker the color, the higher the level of the reproducibility for ACC2-emIAM, thereby allowing the precise numerical values to become less central to the table’s interpretation. Based on this visualization strategy, we consider the present format of the table to be suitable.