

Title: emIAM v1.0: an emulator for Integrated Assessment Models using marginal abatement cost curves

Authors: Xiong et al.

Summary of review

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.

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.
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.
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).

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.
- Why is only REMIND shown in Fig. 8? Perhaps better to show the averages across all models?
- Table 6 is unreadable. I propose to make a selection of things to show rather than everything.