Point-by-point response: https://doi.org/10.5194/essd-2024-68

Reviewer comment	Author response
RC1 We thank the authors for addressing the majority of our comments by providing an overview of the used variables in Table 1, including bootstrapping for a more robust analysis, adding use cases to the discussion, describing the categories and the different regression models used. However, there are some minor points that are not yet fully clear to me.	**AC1** We thank the reviewer for taking the time to re-evaluate our revised manuscript and for the constructive comments. Below, we respond to the remaining comments point-by-point.
I wonder why in the recent manuscript, the k-nearest neighbor method performs better compared to gradient boosting as in the previous version? Is that due to including bootstrapping in your analysis?	Generally, the two models perform comparatively well for the global scenarios. Correct, the k-nearest neighbors model performs slightly better than gradient boosting when including bootstrapping for the global scenarios – the difference gets clearer when looking at the regional scenario variants (see the new Figure 4a and the first paragraph in the Results section).
Which overall metrics do you use to decide which regression model performs best for Fig. 2A and Fig. 3A? Is the decision purely based on visual comparison or do you include some weighting of the four metrics? There is no clear explanation in the results why the k-nearest neighbor method performs best.	We did not weigh or aggregate the four evaluation metrics. As mentioned above, k-nearest neighbors and gradient boosting perform comparatively well for the global scenarios. For the regional variants, k-nearest neighbors outperforms gradient boosting. Another argument against gradient boosting is that this model partly predicted slightly negative values in the target variable, which is conceptually inconsistent with a clean definition of Land CDR, which should have a uniform removal sign. In lines 177-186 in the revised manuscript, we described why we selected the k-nearest neighbors model instead of gradient boosting.
It would ease understanding if you clearly describe the difference between Fig. 2a and 3a at the beginning of the results section.	Thank you for this feedback. We have slightly rephrased the first sentence of the Results section to make the distinction

	between the old Figure 2a and 3a clearer (now Figure 3a and 4a).
RC2 The authors have substantially improved the manuscript based on the reviewer comments. The description of methods is much clearer, the discussion has improved, and the full paper is more comprehensible. In a few instances, I still had difficulties to fully understand the content, as I felt that some explanations are still missing. I thus have a few remaining points that should be considered before publishing the manuscripts. Remaining points:	**AC2** We thank the reviewer for the constructive and detailed feedback on our revised manuscript. This was very helpful to improve the manuscript further! Below, we respond to the remaining comments point-by-point.
• Please explain the meaning of all categories C1-C8. These are central to the analysis, as they are used in two out of the three figures of the manuscript. The categories C1-C8 first appear in Figure 1 but are not explained in the caption. Their meaning should also be explained in the caption of Figure 1.	We have now introduced Table 2 which specifies the meaning of all eight AR6 scenario categories. We have also pointed to Table 2 in the captions of the figures to facilitate the read.
• It remains a bit unclear whether the R10 regions are 10 or 11 regions (including rest of the world), and whether the imputation dataset has data for 10 or 11 regions. Please specify this more clearly.	The R10 regional categorization comprises 10 regions plus an additional category for rest of the world (see new Figure 4b). Most IAMs have no more than 10 regions but there is a handful of models with 11. The models were trained based on all available data without excluding any of the R10 regions – the same applies for applying the model to impute missing data. We have adjusted the wording in lines 84-88 to make this explicit.
• Line 79-87: I cannot follow how the numbers in this section are derived. It also remains unclear how the final number of scenarios (404 and 2358) is connected to the numbers listed in this section. A clearer derivation of the numbers would thus be helpful.	We impute 404 global and 2358 sub-global scenario variants, which lack data on Land CDR in the AR6 Scenario Database. To do so, we focus on the vetted and complete global scenarios (n=783) and vetted and complete regional scenario variants (n=6162). We have now adjusted the wording to make this more

	comprehensible. We have also introduced the new Figure 2, which provides a better overview of how the numbers were derived.
• Sentence in lines 93-96: This sentence is unclear to me. What are the training bins? Why are scenarios not split and have they been split for other analyses? (and if so how and why have they been split?) I think that more explanation is needed here for readers who are not familiar with the methodology.	This means that we did train the model separately based on global scenario data and on regional scenario variants. However, we did not further distinguish and separately train scenarios by AR6 scenario category or R10 region. This was done to keep the number of training data points as large as possible. We have now rephrased this section to better clarify what we did.
• The discussion mentions that in some cases net negative AFOLU CO_2 yields larger CDR than the predicted Land CDR. This already becomes evident in Figures 2 and 3 (e.g. for category C5 in Fig. 2 or region EUROPE in Fig. 3). I think it would be good to already point to this behaviour in the results section (as it is a result of the study).	Thank you for this suggestion. We have now added two sentences to the results section (lines 226-229) and refer to the discussion.
Minor points: • Line 14: Maybe add "not" before "consistently reported" (for clarity).	Thank you for this suggestion. We have adjusted the text accordingly.
• Line 16: It is not yet clear what R10 means. I'd suggest adding a short explanation here.	We have now slightly adjusted the wording in the abstract to address this feedback.
• Line 76: Not only the R10 regions but also global, right?	The 404 refers to global and the 2358 refers to the R10 regions. We changed the wording to "for 404 global scenarios and for 2358 sub-global scenario variants across R10 regions" to make the separation between global and R10 clearer.
• Line 80: Here, the term "net emissions" confuses me. To me "emissions" denote CO_2 fluxes from the land to the atmosphere, whereas CDR are "removals". A more neutral term including	Thank you for this reflection. We have now adjusted the wording and referenced Table 1 again to make it clearer.

both emissions and removals could be "fluxes". If the paper uses "emissions" in a broader sense (i.e., including emissions and removals), this should be clearly stated.	
• Line 81: Shouldn't the reference point to Figure 3b? Also, I suggest adding a map that shows the regions, as there are so many different definitions for regions out there that their naming remains unclear unless a clear depiction of their extent is shown.	Very well spotted, thanks for pointing out this typo! We have changed it accordingly. We agree that a visual overview of the different regions would be desirable. However, as we state in the Methods section, the underlying models, which produced the scenarios don't always use the identical classification of regions. To our understanding, the R10 regions were assigned by colleagues of IAMC and IIASA during the post-processing and curation of the AR6 Scenario Database, which brings the different models and scenarios together. We hope that our revised text still allows for a better understanding.
• Line 170 and following: This only refers to the k-NN approach, right? Please specify.	Correct, we refer to the k-nearest neighbors. We have now made this explicit.
• Line 172 and following: Check the superscript for yr-1.	Thanks for spotting this typo! We have now corrected the superscript.
• Figure 3: The caption for panel (b) is the same as for Figure 2b, although the results are shown here for different regions and not for scenario categories. Please check. Also, how are the different scenarios aggregated to derive the regional timeseries shown in Figure 3b?	Thank you for pointing this out. We have adjusted the caption accordingly. The regional scenario variants across the R10 regions are not aggregated in the old Figure 3b (now Figure 4b). Figure 4b shows the median and 5-95 percentile range across the regional scenario variants for the different R10 regions. In the subplot titles, the number of underlying scenario variants in the regression validation dataset is specified. The total number of regional scenario variants in this dataset is 617 but the number of variants varies across regions.
• Lines 230-232: I suggest adding references to the figures, scenarios, or regions, where this is the case.	In our imputation dataset, we have specified for which scenarios the issue of net negative CO_2 emissions being larger

	than gross removal occurs. The issue of inconsistent removal baselines is more like a cross-scenario issue. We have adjusted the wording to highlight that we made the former issue explicit in our imputation dataset.
• Line 233-236: How is the data replaced in the adjusted dataset? Are only single years replaced or is the full time series replaced? Also, what do the numbers mean in lines 235-236?	The full timeseries is replaces as shown in the imputation dataset. We made this explicit in line 259. The numbers indicate the number of scenarios for which we performed the replacement in the adjusted dataset.
RC3 Thank you for your revisions. I thank the authors for addressing all my comments and expanding on the methods in particular (This helped me greatly in derstanding the dataset and I'm sure it will be helpful to the community). I also appreciate the regional analysis. I also went through the other reviewer's comments and appreciate the value this dataset adds to the community. I would recommend one small technical correction before publication. While the methods described here have been applied to AR6 data, authors should acknowledge whether or not the methods posted here can be extended to other datasets (such as the GCP for example). I don't think its an issue if the methods described here cannot be extended at the moment, but a comment on the same would be helpful for the community in my opinion.	**AC3** We thank the reviewer for the positive feedback on our revised manuscript. Below, we respond to the remaining suggestion. The regression model-based imputation approach could also be adapted and applied to other datasets to infill incomplete variables given that sufficient training data is available to tune and evaluate the regression model. Generally, the merit of our data description paper is not the method itself but the generated dataset, which is of use for researchers working with AR6 scenario data and facing the issue of incomplete reporting of Land CDR. While our approach allows to work with a more complete set of scenarios, it is still a workaround to deal with lacking data availability. For the next generation of mitigation scenarios, the issue should ideally be addressed by improving the reporting of Land CDR across integrated assessment models. See lines 264-266 and 290-293 in the revised manuscript, where we flagged this.