Dear referee,

Thank you for your helpful comments, shown in italics below. Find our replies directly below each comment.

*Overall I found the paper to be well-written and to provide a reasonably comprehensive review of the different model components. I thought there was an appopriate level of technical descriptions, with links to references where individual topics are discussed in greater detail. At a high level, my biggest question pertains to the purpose of this paper in the peer reviewed literature; i.e., can the authors state in the text what the value added is of this paper (as opposed to the model)? As noted, there are already several published model documentation papers, as well as reasonably comprehensive online model documentation. Similarly, the results shown in this manuscript were a cursory review of SSP scenarios that were already published and documented in a number of papers four years ago. It doesn't seem appropriate to be re-publishing this scenario data as if it were new.*

The main purpose of this manuscript is to provide a comprehensive description of the new version 2.1 of REMIND. The last comprehensive documentation of REMIND described version 1.7. In the meantime, REMIND has improved substantially and become open source. Therefore, we would like to provide a complete description of the model which fills in all the missing information and interlinkages which are not included in previous publications focusing on specific aspects of the model. Now we can provide an important update of the SSP scenarios, serving as example results to introduce this new version of REMIND. In the revised version of this manuscript it will be made clear that this is an update of the previous SSP scenarios derived by REMIND 1.6, also pointing out that the scenarios reflect changes in systems representation and spatial resolution, but also are an update in that policy scenarios only start to diverge from 2020 onwards, instead of after 2010 for the original SSPs.

*One way that the authors could differentiate this from the prior literature would be to run scenarios that illustrate the value of new features that have been added to the model since the last documentation in 2017. For example if there's more sophisticated representations of variable renewable energy, the paper could show energy curtailment by region and scenario, or other variables that are interesting but that weren't reported in the SSP inter-comparison exercise and perhaps weren't available at that time anyway.*

A section on what is new in this version of REMIND, containing the following points, wil be added:

- flexible spatial aggregation for input data generation
- open source
- update of techno-economic parameters for most technologies to reflect latest market data
- updated bounds on developments until 2019 to reflect latest deployment and policy developments
- updated policy scenarios
- more detailed representation of demand sectors buildings, transport and industry
- possibility to include aggregated representation of impacts
- possibility of imperfect capital markets

*In terms of reviewing the model, I was generally impressed by the number of features and key interactions captured, but noted two weaknesses that should be explained in the text. First, why is the model calibration year 2005 when it is currently 2021, and the necessary data to calibrate the model to more recent years has been available for a long time? I'd think that the calibration year should be 2010 at a minimum.*

We calibrate variables to 2005 in order to have some years of overlap between model results and historic values, which are useful to confirm that REMIND can replicate observed trends. Significant departures from near-term developments are addressed by applying some bounds, e.g. on technology availability and trade volume. We now added the sentences: “Additional bounds for a few variables, mostly capacity (additions), up until 2019 make sure that the 2020 point of departure in current policy cases is close to actual developments. Being able to run the model also without those enables important comparisons of the model dynamics from 2005-2020 and real-world developments.”

*And, second, why are there only 12 global regions? For policy modeling purposes it's often advantageous to have single-country regions, and 4 of the 12 are single-country which is good, but that leaves some very heterogeneous regions. Canada-Australia/NZ seems an especially interesting market region given that they're at opposite sides of the world. Is there any sub-regionalization in the renewable energy markets, or any other way to prevent windy regions of Canada from supplying electricity to buildings in New Zealand? Similarly, "Other Asia" presumably includes a very wide range of development levels, as South Korea is mixed in with a large number of low-income countries. Can the authors comment on the level of effort/difficulty with adding regions to the model? Perhaps several components already include enhanced detail?*

The spatial aggregation of REMIND is flexible: input data can be provided in any spatial aggregation of countries and the model code is automatically adjusted based on a mapping which defines the spatial aggregation. In general, there is a limit to the number of regions due to the solution algorithm, but also due to the effort required for validating results especially for smaller countries/regions. Given the advent of parallelization in the Nash solution mode, the runtime does not increase substantially with higher numbers of regions. However, we only have limited experience as to the number of regions that the algorithms for market clearing can handle and still return a stringent market clearing solution. The 12-region spatial aggregation is considered REMIND’s default, for which we have validated the input data and model results. Each additional country/region that is modeled explicitly needs to be validated against detailed historical data. The first steps for increasing spatial detail -validating a version where Europe is split into 11 sub-regions - are currently underway. Further spatial detail is possible but would require validation of the model output. Therefore, for the time being, we have to live with the artefacts mentioned by the referee. They however are not too problematic on the global scale: CAN, AUS and NZ each have good renewable ressources, and rather limited populations. The demand estimation in many of the sub-modules calculates demands on country level, and aggregates to the region level, so that some of the heterogeneity within a region is partly reflected in the regional parametrization.

*The final thing I was wondering about the model pertains to the renewable energy supply curves, which appear to include uninhabited lands of Russia, Canada, Australia, etc. Am I*

*correct in understanding that all land area is included in these supply curves, starting in the base year, and that there's no consideration of transmission line distance? While this would be a difficult thing to do well (chicken and egg issue with the transmission lines), that does seem a pretty major omission that would tend to make much more wind energy available for much cheaper that it should, for countries like those named above that have large tracts of uninhabited land, thousands of kilometers from any population centers.*

The renewable potentials we use include only areas with a maximum of 100km distance to existing settlements. Accordingly, we potentially slightly underestimate the renewable potential for large regions with low population density. However, as these regions do not have a scarcity of wind and solar supply, increasing the renewable potential would not have any impact on our results.

Furthermore, the parametrization of grid demands (which are a function of wind share in power generation) takes this into account, at least in an approximate way. In a recent paper soon to be published, we see that REMIND rather even overestimates the grid investment requirements for integrating solar and wind, at least in comparison with estimates from other models and the IEA.

*What follows are some minor questions and requests for clarification:*

** How is proprietary data masked or filtered and re-processed for distribution, given that the model is open source but (presumably) not all data used in its calibration is free?*

Our strategy for proprietary input data distribution is to release the necessary excerpts at REMIND's default level of regional, sectoral and temporal aggregation, so that the comprehensive proprietary source data doesn't need to be published. We have contacted all of our data providing institutions for approval, but are still awaiting an official agreement from the last one (IEA). In the meantime, our fully open-source data processing routines (R packages) help users to generate input data locally, but the user must have access (a license) to the raw data at the moment.

** Line 34 - should be "example", not "exemplary".*

This is changed in the revised version of the manuscript.

** Fig 1 - I believe "labor efficiency" should be re-named "labor productivity" for consistency with the literature.*

We adjusted the figure in the revised version of the manuscript.

** Line 166 (and others): The China region is called "CHA" on line 166 and "CHN" on line 170. My preference would be to always use CHN, the official 3-digit ISO code, similar to the handling of the other single-country model regions (USA, IND, and JPN).*

The region "CHA" of the default regional aggregation of REMIND contains China (CHN), Hong Kong (HKG),Taiwan (TWN) and Maccao (MAC). Because of this we do not use the official 3-digit ISO code for China for our region (which we used in the previous REMIND versions that only mapped CHN to the region). This is clarified in the revised version of the manuscript in Appendix B, which shows the region mapping. Regarding macro-economic

development and climate policy mitigation, we assume that our region “CHA” is dominated by China. We adjusted the regional description in the text to “CHA - mainly China” to make it clear that “CHA” contains more countries than just China.

** Can the authors provide a country-to-model-region mapping list in an Appendix? A number of the boundaries are unclear from the descriptions (e.g., Latvia/Estonia/Lithuania, Turkey).*

We added Appendix B containing a table of regions and countries belonging to those regions in the revised version of the manuscript.

** Line 286 - should be "modes", not "models" (I think; please check)*

This is changed in the revised version of the manuscript.

** Line 615 (about hydropower potential): "The regional disaggregation is based on information from a background paper produced for this report (Horlacher, 2003)" I'm wondering if this is a typo, or perhaps copied from an older document? Otherwise I can't see how a paper published 18 years ago was produced for this report.*

Please excuse our formulation, which could be misunderstood. What we wanted to say is that there is a 2003 report that states global technological hydro potentials, and a 2003 background paper to that 2003 report that provides regional detail. We reformulated to “These estimates are based on the technological potentials provided in the report (WGBU, 2003) and the background paper produced for this report (Horlacher, 2003)”

** Lines 700-715: for the more detailed version of the buildings module, can the authors comment on how this was calibrated? The disaggregation of energy consumption to the services is not something readily available in external data sources, and the paper cited is under review so a brief description here would help.*

As for the simple realisation, the calibration of the REMIND buildings module is based on the EDGE-Buildings projections. EDGE-Buildings disaggregates IEA Energy balances by end-uses in accordance with additional datasets, and can therefore project energy demand for end-uses as well as energy carriers. The methodology is described in detail in the paper (Levesque et al., 2018, https://doi.org/10.1016/j.energy.2018.01.139). An explanation will be added in section 2.4.2 relating to the buildings calibration (“EDGE-Buildings projections are disaggregated both by energy carrier as well as by energy service and can therefore be used to calibrate the different buildings module realizations (see section 3.3.2”). The paper cited in the section describing the buildings module is now published (https://doi.org/10.1088/1748-9326/abdf07), and the citation will be modified accordingly.