Dear referee,

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

The manuscript presents a detail description of the REMIND model. It is well written the content is mostly very clear. However, there are some comments that must be addressed before the manuscript could be considered for publication. The main concern is that it is not clear what new knowledge about REMIND this manuscript brings to the literature. There are several other research papers that describe several components of REMIND (cited in this manuscript as well) that should be better described here (not just refer the reader to other papers).

This manuscript introduces 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.

Additionally, it is not clear what is new in this version of REMIND, compared to previous versions that have been published. The manuscript is not clear about this and does not present a comparison or a section that explicitly presents the improvements.

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

Other minor/detail comments are presented next.

Comments Section 1:

Main comment: It is not clear from this section what is new in this version of REMIND to respect the older versions. This needs to be clearly stated. Other minor comments are described next:

1. *The use of SSPs helps to cover uncertainties regarding technological development for renewable or fossil fuel availability, but also social and behavioral development*

like population growth, dietary preferences, environmental awareness or international cooperation.

This sentence will be changed slightly: "...uncertainties regarding technological development for renewable energy or fossil fuel availability, but also social and behavioral developments like…"

2. *It is not clear what is meat by "damages" in the following text:* ***, the REMIND model represents some damages and can thus be used for cost-benefit analyses or least total cost analyses***

   In REMIND, "damages" refer to economic damages due to climate change. This sentence will be changed to "REMIND can endogenously represent macro-economic climate change damages based on recent damage function estimates (Kalkuhl and Wenz 2020)" in the revised version of the manuscript.

3. *In the next sentence, what is meant by self-consistent? Also, it is understood that the energy sector is modeled, while the economic (macroeconomic) is input. Hence, it seems a bit contradictory the next sentence: REMIND enables the exploration of a wide range of plausible developments and of possible futures of the* ***energy-economic*** *system exploring* ***self-consistent*** *transformation pathways.*

   The SSP scenarios provide a wide range of possible macroeconomic developments. While the population in REMIND is exogenous information and REMIND is calibrated for baseline scenarios to match GDP and final energy trajectories, investment decisions into the macro economic capital stock and energy carriers are endogenous. In particular, changes in macroeconomic variables from baseline scenarios to policy scenarios are endogenous. In the revised version of the manuscript we substituted "self-consistent" with "internally consistent", and the ensuing sentence explains that this refers to the simultaneous consideration of various sectors, including a simplified macro-economy, and the interplay of technology transformation and socio-economic changes.

4. *It is mentioned several times (up to page 3) that REMIND is an NLP, however, no description or ideas of the non-linear components are described. What make the model nonlinear? Some initial hints would be beneficial .*

   In REMIND several model components are represented via nonlinear equations (e.g. welfare function, CES production function, technological learning curve). This information will be added in the corresponding sections of the revised manuscript.

5. *"REMIND calculates economic and energy investments". What is meant by an "economic investment"?*

   By "economic investments" we mean investments into the macro-economic capital stock (i.e. physical capital that represents machinery, long-lived equipment, infrastructure, etc. beyond the core energy system represented explicitly via separate technologies). This sentence will be changed to "...calculates aggregate macro-economic as well as technology-specific energy-related investments…".

Introduce the concept of Pareto optimum. Pareto is normally use in the context of multi-objective optimization.

We apply the concept of Pareto optimality in an economic context (cf. Mas Collell et. al., 1995, p. 312 ff), specifically in a setting where different regions interact with each other. A solution is Pareto optimal if there is no other allocation of economic income and resources that would increase the welfare of one region without decreasing the welfare of another. By using the Negishi approach, REMIND directly finds a Pareto optimum that is, by construction, equivalent to the competitive equilibrium found by the Nash approach with full internalization of externalities. An appropriate explanation will be included in section 2.2.

7. *"The optimization is subject to equilibrium constraints, such as energy balances, economic production functions or the budget constraint of the representative household." It is not clear the mathematical structure of the REMIND model. Is it an NLP in the form of an optimization or equilibrium (MPEC) model? (MPEC: Mathematical program with equilibrium constraints).*

   The model is formulated as a standard nonlinear programming model with an objective function and a large number of constraints that represent different parts of the overall system. These constraints represent equilibrium or balance conditions, but are not explicitly differentiated as such. This way of model formulation is less strict than the MPEC models. This flexibility is needed as the model merges together different components of the energy-economy-climate system that are quite different in nature. An appropriate explanation will be included in section 2.2.

8. *"REMIND is usually run in a decentralized mode where each model region is optimized separately, and clearing of global trade markets ensured via iterative solutions (see section 2.2)." -> How the model ensures convergence? This needs to be clarified in the corresponding section.*

   In order to resolve trade interactions between model regions (i.e. clearing markets), the decentralized (Nash) approach applies a Walrasian type price adjustment algorithm. Regional actors start from an initial price vector and choose their trade patterns, acting as price takers. The regional solutions are subsequently collected, and the price for the next iteration is adjusted based on the surplus and deficits on the markets. Walrasian-type price adjustment algorithms are commonly used and convergence is conceptually proven under generous conditions (see also section 3.1.2). The implemented specification of the price adjustment algorithm (see details in Leimbach et al., 2017) makes use of parameters that play the role of price elasticities and help the model to converge. In order to guarantee convergence, two auxiliary mechanisms are applied: (i) anticipation of price changes, and (ii) penalty costs depending on the change of regional trade patterns over iterations. This information will be included in section 2.2.

9. *"CH4 emissions from fossil fuel extraction and residential energy use" What about CH4 from the agro-industrial (food, beef-lamb) sectors? If not considered, it must be clarified.*

CH4 emissions from the land-use sector as well as from waste handling are considered. However, they are only calculated via emission baselines and marginal abatement cost curves (MACCs) - not directly by source. We reformulated the text to mention CH4 emissions from fossil fuel extraction and residential energy use under section 3.4.1 together with the GHG emissions from other sources.

10. *"Historical data for the year 2005 is used to calibrate most of the free variables (e.g. primary 130 energy mixes in 2005, secondary energy mixes in 2005, standing capacities in 2005, trade in all traded goods for 2005)." Why there are not updates to the base calibration year? 2005 seems a bit old to account for new trends.*

    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."

Comments Section 2:

1. *"(for more information about the modular structure see Dietrich et al., 2019 - Appendix A." It would be good to introduce some of this information in this article, since it is such a critical piece of the model structure of REMIND.*

    A short description of the modular structure will be added in the revised version of the manuscript.

2. *"This paper focuses on realizations which are active in default scenarios. More detail about all modules and their interlinkages can be found in the model documentation". I still believe that this is information is relevant and should be somehow described and discussed in this manuscript.*

    The module structure in general enables a clear structure of the code and splits it into meaningful components with clear interfaces. A bit more explanation about this module structure will be added in the revised version of the manuscript. Although the focus of this manuscript is on the default realizations of a module, in most cases a short description of alternative realizations is given (e.g. description of the damage module in section 3.1.5). Previous publications which focus on the more detailed realizations are often mentioned as well. In some cases there exists only one realization of a module. This will be indicated in the revised version of the manuscript. Nevertheless, a complete list of all available module realizations is provided by our model documentation.

3. *"By default REMIND calculates results for the 12 following world regions:" A table with regions and other important information would be better than just listing countries/regions.*

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

4. *"By parallelizing the calculation of the individual regions in decentralized optimization mode (see section 2.2) the computation time increases only moderately with increasing spatial detail." It would be interesting to have a general idea of the computational complexity of the model (minutes, hours, days?) depending on the type of scenarios.*

   REMIND converges within 1 to 3 hours for scenarios with the default settings. If some modules are used in a more complex version, this can increase to 5-8 hours. The runtime only slightly varies with policy type, with intermediate ambition policy scenarios without many policy-specific constraints having longest runtimes. In line 199 of the first manuscript, we provided the rough estimate: "(thanks to the possibility of parallel computing, scenarios converge within one to a few hours)". We now specify further: "both baseline and policy scenarios converge within one to a few hours, mostly depending on the specified module detail."

5. *"Time represents a separate dimension" -> What is meant by this? Not clear at all.*

   In economics, static models are widely used. They just analyse economic activities in a given point of time and the model dimension is mainly determined by the number of actors (firms, consumers) and goods (sectors, markets). In a dynamic model like REMIND, the dimensionality, and therefore the numerical demand, is largely influenced by the time dimension, i.e. time horizon and time steps. We specified further: "...represents a separate dimension within all equations (next to the other general dimension "region", and further equation-specific dimensions relating to technologies, emission species, etc.), increasing the overall dimensionality of the model."

6. *"In essence, the time dimension only increases the number of markets for which the algorithm has to find an equilibrium" I would be extremely careful about the use of the concept "equilibrium". To this point, the model has been introduced as a NLP optimization problem, with some equilibrium constraints. But there is not clear mathematical structure to really understand what the model does. If it is a pure optimization model, what talk about equilibrium? Why Pareto optimal is mentioned earlier? Please be clear and consistent with the type of solution that is obtained.*

   There is a clear context in which we use the concept of equilibrium: trade interactions of model regions. As part of the model solution, an equilibrium is found that is characterized by a set of prices for tradable goods that clears all markets. While we do not prove the existence of this equilibrium for the model, we apply algorithms that are known to generate an equilibrium: Walrasian-type price adjustment.

   A corresponding context is given for the concept of Pareto optimality. Here, we repeat our response to question 5 in section 1: We apply the concept of Pareto optimality in an economic context (cf. Mas Collell et. al., 1995, p. 312 ff), specifically in a setting where different regions interact with each other. A solution is Pareto optimal if there is no other allocation of economic income and resources that would

increase the welfare of one region without decreasing the welfare of another. By using the Negishi approach, REMIND directly finds a Pareto optimum that is, by construction, equivalent to the competitive equilibrium found by the Nash approach with full internalization of externalities. An appropriate explanation will be included in section 2.2.

Also, it was mentioned that CONOPT is used to solve the NLP problem, hence, it is also questionable when authors refer to the "algorithm" use to solve REMIND, since it is in fact a solver who does this process and authors have not developed an algorithm. In the case that an algorithm is indeed implemented, then this has not been clearly stated and differentiated from the NLP-CONOPT process.

In section 2.2 will be started with referring to CONOPT before introducing Nash and Negishi. These are the two algorithms which can be employed to find the singular market clearing solution within all of the Pareto optimal solutions identified by the CONOPT solver.

7. *Based on the previous comment, I found then that there is indeed a NASH mode in REMIND. This helps to understand the concept of equilibrium. However, there is still not clarity in terms of what the base structure of REMIND is, how different structures are solve, what type of solution is obtained, solution algorithm, etc. This needs to be further clarified.*

   REMIND, as a composition of different modules and components, is mathematically coded as a nonlinear programming model. As such, it is computed by the solver CONOPT. Basic features of the underlying solution algorithm are hidden. There is however a second layer of solution structure. This is related to algorithms that we use in order to generate solutions that are meaningful from an economic point of view (i.e. Pareto optimum and competitive equilibrium, respectively). In order to find such equilibrium solutions with the REMIND model, trade interactions of regions have to be reconciled. Two corresponding mechanisms are implemented: Nash and Negishi algorithm. Both are iterative algorithms. This information will be added at the beginning of section 2.2. Within each iteration, the entire NLP model (which itself is solved within thousands of iterations) is computed. The Nash and Negishi approach is briefly described in section 2.2. The presentation of this section will be revised, but still refers to Leimbach et al. (2017) for more details in the discussed context.

Comments Section 3:

1. *"It is possible to compute the Pareto-optimal global equilibrium including inter-regional trade as the global social optimum using the Negishi method (Negishi, 1972), or the decentralized market solution among regions using the Nash concept (Leimbach et al., 2017)". This is interesting but needs further clarification. In practice, a Nash solution is an equilibrium, that can be categorized in some conditions as Pareto optimal. In fact, it has been studied that Pareto optimal strategies are a subset of Nash Equilibrium strategies (see paper D**OI:** 1109/ICCCNT45670.2019.8944817)*

   See reply to comment 6, section 2.

2. *"REMIND considers the trade of coal, gas, oil, biomass, uranium, the composite good (aggregated output of the macroeconomic system), and emissions permits (in the case of emissions-trading-system (ETS) based climate policy).". Are ETS global in REMIND? Or can be defined for particular regions? If global, how are allowances distributed?*

   The default climate policy implementation is via carbon prices, which can either be prescribed exogenously, or can be iteratively adjusted to achieve a prescribed near-term emission target or long-term cumulative carbon budget. There is, however, also the option of defining a global emission budget with various alternative prescribed distributions of permits which then can be traded across regions. This option has been used most recently in Leimbach and Giannousakis 2019 (https://link.springer.com/article/10.1007/s10584-019-02469-8). In response to your comment, we changed the text in parentheses to "(in the case of emission-trading-system (ETS) based climate policy, which is not the default but has been used in some studies, most recently in Leimbach and Giannousakis 2019)".

3. *"To match 2005 values in the IEA statistics, REMIND adjusts the regional by-production coefficients of combined heat and power (CHP) technologies." This refers to the calibration of REMIND in the Energy Sector? If so, it is still not clear how the full calibration process works since it will also depend on the macroeconomic results and other sectors that may not be linked to CHP plants (transport for instance?).*

   Yes, this sentence refers to the energy transformation part of REMIND. What we mean by this sentence is that in different world regions, CHP plants have different electricity to heat output ratios according to the IEA statistics.
   We reformulated this paragraph to make clearer that module "04_PE_FE_parameters" calculates regional conversion efficiencies and CHP coefficients from IEA energy statistics, and module "05_initialCap" calculates the vintage capital stock needed to produce and convert these amounts of energy.

4. *"represent challenges and options related to the temporal and spatial variability of wind and solar power" Please elaborate on this. If the mode runs with 5 years' time steps, how the temporal variability is considered? It is not inter-annual? Curtailment rates, which are mentioned later, will also depend on the increased levels of demand in future periods as well as the inclusion of other flexibility technologies, such as electrolyzer, which can transform excess electricity into a different energy carrier. Hence, it does not seem correct to consider curtailment rates.*

   Bridging the time scales of investments (5 year time steps) and power sector variability (inter-annual) has been one research focus of our group over the past 10 years. In order to make clearer how we have a parametric representation of variability (which functions on time scales of hours and days) in our model with 5-year time steps, we changed the sentence "These drivers are parameterized for a range of wind and solar PV generation shares, as well as for the regional-specific temporal matching of electricity demand and renewable supply." to "These variables are linked via specific equations to the shares of wind and solar PV (variable

renewable energy, VRE) generation, with higher VRE shares resulting in higher requirements for storage and grid. The parametrization of these equations also takes into account the regional-specific temporal and regional matching of electricity demand and renewable supply, so that regions with better match (e.g. having large noon demand peaks for air conditioning) require less storage, and regions with higher geographical vicinity of VRE and demand require less grid."