Reviewer 3 (Robert Huber):

I think this is a valuable contribution to the discussion on how computational models can inform political and social efforts toward more sustainable land systems on a large spatial scale. I find it very important to explicitly discuss underlying paradigms in models of land-use change and I think this contribution is an important step to improve our understanding of how the paradigm affects the interpretability and validity of such models. In my view, however, the current version of the manuscript could be improved by describing the model paradigms more explicit and by a more careful presentation of the input-output relations in the result Section. In addition, I think the contribution would gain from discussing the implications of the different paradigms to inform what the authors call "efforts to limit climate change and reverse biodiversity loss". I would like to specify my general comment below.

Many thanks for the positive feedbacks and useful suggestions, all of which we propose to adopt.

1) I think it would be important to introduce the two model paradigms earlier in the manuscript. In my understanding, the first part of the discussion (lines 239-260) is something that defines the research design and should not appear as something the authors would like to discuss. In addition, I think it would help the reader if the authors would also discuss and classify/categorize these two paradigms a bit broader e.g. in the context of their own work on ABMs and their theoretical and philosophical background (Arneth et al., 2014; Brown et al., 2016).

We will follow this suggestion, moving the text referred to back in the manuscript, detailing the paradigms more carefully in the new model descriptions and relating them to earlier publications as suggested.

2) In the same vein, I think that the discussion of the paradigm should also include implications from the different mathematical model implementations. If I understood the models correctly, IAP maintains equilibrium between the supply and demand for food while agents in CRAFTY compete for land-uses having a satisficing behavior including non-economic benefits. The point I'd like to make is that models based on rational economic behavior usually are characterized by switching from one corner of the mathematical solution space to another. I do not know whether this describes IAP adequately. However, the outputs seem to suggest that CRAFTY results are always more balanced than the economically driven IAP results. Thus, I would suspect that IAP jumps to corner solutions. If this is true, then this would also be known before the comparison. There might be other direct implications of the mathematical implementation of the models for the interpretation of the output. This could be introduced and discussed in the context of model paradigms.

A good point, and we will highlight the issues of this sort that are clear prior to the comparison, with those relating to these specific models in the model comparison table and those relating to these paradigms more generally in new text. This attribution is slightly complex – for instance the IAP simulates and aggregates up to 40 clusters in each grid cell that produce different solutions, with changes also possible within a land use class (e.g. different crop selections), but there are certainly elements of model design that mathematically pre-define model outcomes, yes.

3) I also found it difficult to follow the input-output description in the text but also the figures. The authors use the pre-defined abbreviations for the different climate and socio-economic scenarios. I understand that there are reasons not to give explicit names to these scenarios. Nevertheless, it makes the presentation of the comparison in this contribution very demanding. As a reader unfamiliar with the exact definition of each of the socio-economic scenarios, I always had to cross check what SSP3 or SSP1 now exactly implies with respect to the input assumptions. Since there is no description of the socio-economic scenarios in Section 2.2, I had to use O'Neill et al. to be able to

follow the result Section. In addition, I did not really understand how the convergence scenario was developed. I think the manuscript would profit from a concise description of the socio-economic scenarios and how these scenarios affect the underlying assumption in the model exercise e.g. production functions, demand levels etc. This would help the reader to connect input- and outputs in the different models. Personally, I would find it also helpful if there would not be abbreviations for the socio-economic scenarios. This could make it easier and more accessible to the reader e.g. in Figure 1.

Thanks for highlighting this. We will add a scenario description and implementation table, and also avoid the acronyms where possible (while perhaps keeping them where the full names would otherwise be frequently repeated).

4) In this context, I also had the impression that the authors did not adequately address and discuss the uncertainty with respect to model inputs. For example, the author writes that there is a "greatly increased productivity" in the scenario RCP8.5-SSP5 and consequently, the IAP model suggests that the supply of crops and meat can increase more than 30% with less than a third of the area of intensive agriculture (comparing Figure 1 and Figure 2a). The increase in productivity, in contrast, did obviously not affect land-use in CRAFTY. However, I would expect that a change in the productivity increase would considerably lower the extreme solution in the SSP5 scenario. Maybe that is something the authors wanted to address with the "convergence" comparison: Look at the sensitivity to input parameters of specific importance. I think this would deserve more attention. Maybe the authors can include more than just two input variations (increase in imports and food values) and discuss the results in the context of input uncertainty that seems to have very different impacts in the two model paradigms.

An interesting point, and it's quite correct that we haven't dealt with uncertainty/sensitivity in any depth here – we agree that we should include more on this. While we're wary of adding more experiments here in addition to substantial extra explanation as suggested by the reviewers, both models have been quite extensively assessed in sensitivity and uncertainty analyses in the past, including with respect to scenario conditions, and we therefore propose to include a summary of these findings and particularly their bearing on the differences between the models that we identify. We believe this would indeed improve interpretation of the findings.

5) With respect to the methods, I acknowledge that these are well documented and state-of-art models that are suitable for comparing the effect of different model paradigms on future land-use change. However, one sentence in the manuscript confused me. The authors write (lines 113ff): "CRAFTY-EU is parameterised on the basis of the IAP, taking IAP outputs as exogenous conditions and replacing only the land allocation component to provide alternative land use projections under identical driving conditions." What is implied here by taking the output of IAP as exogenous conditions for CRAFTY. It would not make sense to use outputs as inputs in another model and then conclude that the models have different outputs. I'm sure this is a misunderstanding (culpa mea). However, the authors should be clearer in what they do here. What are these conditions (except land-use) and how do they affect the comparison? Maybe the solution here goes hand in hand with the reply to my comment 3. However, I would suggest that the authors explain the data exchange between the models in more detail.

Yes agreed, we will explain this properly in the revision, including via a model diagram that shows the relationship of the two models and their input/output sharing.

6)The last comment I'd like to make is probably also the most difficult to address. When looking at the results, I had the impression that the two model paradigms lead to really large differences despite using the same scenarios (e.g. in Figure 3). The authors also state that the "greatest value of

these two approaches may therefore lie in their ability to provide alternatives". But if these models should inform "efforts to limit climate change and reverse biodiversity loss" what do these alternatives imply? Obviously one would come to very different conclusion what to do concerning e.g. biodiversity loss depending on the model paradigm (irrespectively of the scenario). Given the potentially contradicting (policy) conclusions from these "alternatives", critics of mathematical modelling could argue that this "invalidates" such simulations. One can get any result by choosing the "right" paradigm. I'm aware that the contribution does not attempt to address all of the caveats in model design, analysis and interpretation mentioned in the Introduction. However, I had the impression that the authors take refuge in discussing "technical integration" of models. But how could such a hybrid modelling approach solve potential contradictions? In climate change modeling, model ensembles are a way of addressing different underlying functionalities of models. However, it seems to be impossible when looking at the results of this exercise. I think this point should at least be discussed: what if model paradigms prevent instead of foster discussions on how to use modelling of more sustainable land systems on a large spatial scale? I have the impression that the authors should also discuss the value of theoretical underpinnings and conceptual frameworks (which may be more important in this context) than just "more data from another discipline on another spatial level" (which is my simplified interpretation of the last paragraph).

We find these excellent suggestions and fair criticisms. It is probably true that we take refuge in technical issues to some extent! This is partly because we wish to establish basic differences here, but we should have better addressed this overarching issue. We will therefore add text in the discussion to link our findings to the motivating question of model uses, and actually believe we can suggest some useful ways forward in terms of converging on more balanced representations that account for the different effects highlighted in the comparison.

Minor comments

What is the unit of the Y-axes in Figure 1? I would prefer if the difference between IAP/CRAFTY in the figures would not be represented by the level of shading only. Maybe the authors can use a different pattern or something that makes it easier to distinguish the models. I found the caption in the Figures not self-explaining (and I have to say a bit cryptic in the beginning). I do not really understand why some specific information is given in one Figure but not in the other. I think that the authors should try to make the caption self-explaining (in a way). On line 302ff, the authors state that "Conversely, (constrained) optimising models like the IAP produce idealized results that (. . .) can use flexible spatial dependencies as proxies for processes such as imitation, diffusion of knowledge or the formation of social norms (). Are you sure that knowledge diffusion and social norms fit into the economic framework of IAP? Not sure I understood this sentence.

All to be changed as suggested.