Reply to short comment by Judith Verstegen

This study analyzes 16 scenarios from five land use models (LUMs) and integrated assessment models (IAMs) and the effects of the resulting land use changes towards 2040 on ecosystem service (ES) indicator values. The text is easy to read, the work in well-embedded in existing literature, and the results are visualizaed comprehensively. I have two main questions/concerns with respect to this study, as detailed below.

Thanks for the overall positive statement on the manuscript. We answer the questions below & will incorporate these into the revised manuscript.

Model-coupling mismatches: To my understanding, you use a one-way coupling between the LUM/IAM and LPJ-GUESS. This can lead to mismatches. For example, the IAM has computed production of particular agricultural products for a region, and converted this into a cropland area per grid cell. In LPJ-GUESS, this grid cell contains a prescribed fraction of crop functional types, which can mismatch with the products in the IAM, and a certain yield, which can mismatch with the production (supply) of agricultural area of that grid cell in the IAM. I believe that such mismatches have a large effect on the results. I expect that large differences in ES indicator values between two models will occur when one of them has a good match with LPJ-GUESS (due to the use of the same data sources) and another a bad match. The comparison between ES indicator values is not fair in this case, because for the second model, its assumptions are violated by the LPJ-GUESS model (for example, demand in the IAM does not match supply computed by LPJ-GUESS). I think it would help to explain how the models are coupled exactly (what variable(s) is/are exchanged) and to bring this point up in your discussion if you agree with this potential issue.

We follow in the manuscript an absolutely standard approach in how land use and land cover changes from LUM/IAMs have been widely used in the ecosystem modelling community for many years (but also for e.g., species distribution models, hydrology models) to assess the impacts on ecosystem processes or biodiversity-related variables. Implicitly in such an approach is that some of the variables assessed in an ecosystem model would also be computed in the models that deliver the land use change scenarios – most notably crop yields or some carbon-cycle or water-cycle related variables. This is unavoidable when using projections of e.g. land use change from one model type, and using these in another model type. In fact, a similar criticism can be levelled on using climate projections for ecosystem models: the ESM that produces these climate projections would often have a different (and often a more simplistic representation of) vegetation in a gridcell (and this vegetation would affect the computed climate) compared to e.g. LPJ-GUESS.

However, these inconsistencies would only be an issue if variables computed with LPJ-GUESS would be directly compared to variables computed in the IAMs/LUMs (or: in an ESM). For instance, if we compared computed yields in the different models, or runoff. But this is not our objective here: we take a range of land use change projections at "face value" and investigate impacts on ecosystem variables in LPJ-GUESS. In principle we could perform a similar experiment with completely stylised or randomised land-use change scenarios that are not computed by a LUM/IAM at all. However, this would make the analysis a purely technical one and would not highlight the large impacts that different socio-economic scenarios have on ecosystems.

Therefore we don't fully share the expressed concern, but we will add a short note on this point to the revised manuscript, for clarity. For instance: "Some of the variables assessed in LPJ-GUESS would also be computed in the models that deliver the land use change scenarios – most notably crop yields or some carbon-cycle or water-cycle related variables. The spatial patterns of these would differ in the LUMs and LPJ-GUESS. However, this does not affect our analysis: here we take the LULC change projections in a uni-directional approach to assess impacts on ecosystem processes; we do not compare similar ecosystem output variables across different model types."

Scenario projections: It is recognized that LUMs/IAMs do not provide predictions (as weather predictions do), but instead projections, meaning that they are conditional (what if ... ?). What conditions are evaluated depends on the question at hand. The conditions can range from realistic (business as usual) to very irrealistic. Irrealistic scenarios can still be useful as thought experiments, to better understand the system, to serve as warnings for worst-case effects, or to evaluation potential policy interventions. As such, in lines 487 and further, you discuss that some scenario results don't seem plausible. But perhaps they aren't meant to be plausible. Therefore, given that the conditions, and thus the scenarios, depend very much on the question asked in the original study, it is not clear to me what you are exactly evaluating when assessing the variation is ES indicator outcomes over all these models and scenarios combined. I could see the added value of analyzing ES indicator outcomes of all business as usual scenarios, as that would show the effect of different assumptions about the working of the current system on ES impacts, but the value of comparing among the other scenarios (which could have easily been very different if a different questions were asked) is not quite clear to me. In the current version of the manuscript, you only comment about this seems to be "However, conclusions drawn here in regard to projected changes in LULC and ES indicators are inherently dependent on the selected set of LUMs and scenarios, evaluation time period and simulation set-up", which does not really help the reader to see what can and cannot be learned from the results given this dependence.

All scenarios analysed here have different socio-economic storylines underpinning them, which are linked to the SSP framework. These storylines indeed represent different, unknown futures. But none of them have been specifically designed to be unrealistic. By contrast, great effort has been put into making each storyline internally consistent (see e.g., O'Neill BC, Kriegler E, Ebi KL, Kemp-Benedict E, Riahi K, et al. 2017. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change* 42: 169-80). The original studies that led to the various LULC scenarios we investigate here were all designed to project land use change under a plausible future. Our purpose here is to not only compare same style of scenarios (e.g. BAU – which in itself would be a valuable analysis to do, we agree) but to highlight the large impact unknown future land use change has on ecosystems. Most people are well aware that different climate trajectories (RCPs) will greatly affect ecosystems. Even climate change projections for a single RCP when realised with different ESMs will result in large variability in computed ecosystem outcomes. The fact that similarly large variability can be introduced by land-use change (within or between e.g. an SSP) is less known – which is one of our objectives (we will highlight this in the revised version of the manuscript). At the same time, we also observe that the rate or direction of some of the land use change projections seem implausible, irrespective of the underlying storyline.

Minor comments:

481-486: In the context of this paragraph, which speculates about the potentially more valid small-scale changes of some of the models, you may be interested to know that, in a recent study of LUC in Brazil, we found that indeed the small-scale changes of a spatially-explicit LUM were more accurate than the larger-scale changes of an economic model, see Stepanov et al. 2020, doi:10.3390/land9020052

Thanks for pointing us to the Stepanov et al study, which is a good example to highlight in the context of this paragraph. We will use it to strengthen the argument we make in our revisions.

691-692: "We conclude that LUMs and IAMs have fundamental limitations in capturing all relevant processes related to LULC changes." I don't see how your results lead to this conclusion.

There are indeed a number of features that we would argue are important limitations. For instance, none of the models investigated in this study account for gross land changes – which in turn is crucial when assessing carbon cycle impacts ('slow in, fast out'). Additionally, these models would still

underestimate the dynamics observed for example with RS products (e.g. Fuchs et al 2014 - GCB / Pongratz et al. 2017 - GCB “Model meets data” / Bayer et al 2017 - ESD), e.g. the temporal and spatial cascades of transition types consisting of multiple land use changes (forestry dynamics, shifting cultivation, agro-forestry, crop-pasture rotation systems, land abandonment, etc.).