Reply to anonymous referee #1

This study analyzes an ensemble of future land use projections arising from a set of 5 LUMs/IAMs. These land use projections are then translated into changes in ecosystem services (ES) using the LPJ-GUESS model. One of the main conclusions is that there is a large spread in land use projections (and therefore ES) and that most of this spread originates from structural model differences (i.e. choice of LUM/IAM) rather than from socio-economic assumptions (i.e. choice of scenario storyline within a given model). Overall, this is an interesting study which makes an important point about the large uncertainties associated with future land use projections and their potential causes. However some aspects of the manuscript could still be improved and require additional clarifications.

We thank the reviewer for the expressed interest in our manuscript. In the revisions to the manuscript we will be addressing the raised questions as described below.

I found the whole discussion about the role of the baseline level very hard to follow. I understand from the analysis that there are very different starting states in 2000- 2004 across the models. But does it affect the conclusions of the study or not? One could for instance intuitively expect that in a model starting with lower present-day natural vegetation there would more room for future expansion. More specifically, in which way is the baseline level taken into account when looking at relative changes (L236-239)? And where is the assumption that "Effects of differences in the modeling protocol of CLUMondo/LUH1/LUH2 and IMAGE/MAgPIE simulations for projections of ecosystem dynamics affect especially the base level of ES indicators in 2000-2004 and the relative deviations over time, which this study is focused on, only to a small degree" coming from? This last sentence by the way is so convoluted that I might just have misinterpreted its meaning. On that same issue, it is somewhat disturbing that the IPSL-based climate forcing has been bias-corrected, implying a sort of harmonization between observed and projected climate, while no harmonization was performed for the land use forcing. Why not following the same logic for both climate and land use?

Starting with the last question first: the main reason for using a bias corrected climate is not so much the seamless historical-future transition (which is essential for e.g. carbon and water cycle simulations: without a harmonized climate time series, an artificial offset would be created between the end of the historical and the beginning of the projected future climate), but the fact that simulated climate output from GCMs is much more reliable in terms of anomalies, rather than in terms of absolute values – most GCMs are biased compared to "real" climate values in a grid cell. Therefore, projected raw daily GCM output has to be corrected, using the differences in the mean and variability between GCM and observations in a reference period.

The reviewer is correct that in principle one could also do that for land-use historical-to-future timeseries. However, doing so would to large degree affect the underlying objectives of our study. We aim to show here the large uncertainty/variability in terms of impacts on ecosystems that arises both from the very different (unknown) socio-economic futures as well as from the different LUC modelling approaches. This includes keeping the differences in the historical baseline, although the latter is less prominent in the manuscript by us concentrating on the relative changes. We will clarify this in the manuscript upon revision, and by doing so will also revise the sentence "Effects of…", which is indeed poorly phrased.

Suggested revisions: „The differences in the modeling protocol of CLUMondo/LUH1/LUH2 and IMAGE/MAgPIE simulations will affect the base level of ES indicators in 2000-2004 to some degree, although the impacts of slightly different historical model periods and spin-up and historical climate would have diminished by the beginning of the 21st century (baseline period). Larger effects arise from the differences in the individual land-use change models per se (see also Alexander et al., 2017). In principle, differences in the baseline land-cover maps could spill-over to the degree of change in the future scenarios. For instance, presence or absence of natural vegetation in the base-line maps might translate into different degrees of future (semi)natural vegetation re-growth. However, this would only be an important consideration for similar scenarios (and their underpinning storylines related to e.g. sustainability). By contrast, the alternative approach of harmonizing the different projections to the same starting point of land-cover would artificially mask some of the simulated differences in ecosystem services indicators which would be contrary to our objectives."

I have been wondering whether the positive trend in some ES indicators could be affected or even reversed if fires were properly accounted for. Referring to fire and other processes, section 4.2 has a rather elusive statement: "Because these processes are only to some degree implemented in LPJ-GUESS (see, e.g., Pugh et al., 2019), this could further increase the regional variability in ES indicators as indicated in this study." Something more explicit would be welcome, such as clarifying upfront in section 2.2 how fires are accounted for in the model.

LPJ-GUESS, like other DGVMs, models fire explicitly, but has a simplified representation of other forms of disturbances (which can arise from storms, or insect attaches, etc.). We will clarify this both in the methods, as well as in section 4.2.:

"Because only fire is explicitly simulated as a stand-replacing disturbance process in LPJ-GUESS, while others are subsumed in a stochastic background-disturbance, this could further increase the regional variability in ES indicators (see, e.g., Pugh et al., 2019)."

The conclusion that "some scenarios show questionable and possibly unrealistic features in their LULC allocations" could warrant some more in-depth evaluation of historical trends to be fully supported. Although this might be an ambitious task, the discussion could at least outline some evaluation strategies that should be deployed in future studies in order to pinpoint more specific deficiencies.

The reviewer is certainly right that a full evaluation of pros and cons of the implementation of scenarios' storylines is beyond our study. We aspired to put scenarios' projected LULC changes in a historic context because this is our most realistic reference which led us us to the presented conclusion. We suggest to add the following text to the manuscript in order to take up the suggestions (section 4.1.);

"In this context it seems worthwhile for the land use modelling community to evaluate future land use changes against historic trends, in spatial, temporal and thematic aspects. This may help to avoid some of the deemed questionable/unrealistic land use change effects seen in this study. However, current data products of historic land use change are often themselves associated with high uncertainty in historic trends, due to data limitations. In order to build improved historical products, merging multiple data sources could support the evaluation of future projected LULC changes. In addition, a clear declaration of potentially possible regional or global LULC pathways in comparison to LULC changes going beyond historical exemplars would be desirable."

I am still confused about the procedure by which crops are prescribed in LPJ. The total crop fraction evolves according to the given land use scenario, I assume, while the particular mixture of crop types is prescribed to be constant in time after 2006 (table S1). If this mixture is constant in time, then how was it possible to represent crop adaptation by "simulating the adequate selection of suitable crop varieties under changing climate" (section 2.2)?

The mixture of crop types in terms of C3 crops vs. C4 crops vs. rice (respectively 4 crop types in IMAGE/MAgPIE simulations) follows the MIRCA2000 dataset in the version of Fader et al. (2010) with increasing amount of irrigated crops until 2006. After 2006 crop fractions do not change. This is stated in Table S1. The adaptation of crop types under changing climate, i.e. the dynamic calculation of PHU, is indeed meant to represent adaptive improvement (or selection) of crop *varieties* under climate change. It does not imply different crop *species*. As already stated in section2.2 “Adaptation to climate change is partially accounted for by a dynamic calculation of potential heat units (PHU) needed for the full development of a crop before harvest, simulating the adequate selection of suitable crop varieties under changing climate (see Lindeskog et al., 2013)”. We suggest to add in section 2.4 additional text such as “Yields respond to changes in climate and CO_2, including also adaptation as arising from the calculation of dynamic PHU (see 2.2). Adaptation related to e.g. choosing different crops species in a gridcell was not considered here in simulations of the future period.”

Could you please clarify if NPP is aggregated over all ecosystem types including crops? I suppose this is the case, which would imply that there is some information overlap between the NPP and crop productivity indicators. Would it be possible to show some disaggregated results for NPP (i.e. separately for the 3 types of vegetation represented in LPJ)? It might help to reveal some ecosystem-specific responses.

Yes, NPP is aggregated over all plant and crop functional types and therefore there is an information overlap between NPP and the crop production indicator. NPP could in principle be split into NPP from crops and other plants. However, from the ecosystem service point of view, NPP as the entirety of the biomass produced in the ecosystem is relevant and not so much the crop/plant types that it comes from. A detailed evaluation of this would be besides the main topic of our study. Therefore, we would not want to distract the reader further by splitting up NPP and with this adding complexity to tables and figures. Also by showing crop production explicitly, an indication is given, where NPP is significantly influenced by crop NPP. We will clarify the definition of NPP and the information overlap with crop production in sect 2.4.

L31-33: please provide an uncertainty range for all variables along with the median value.

Will be done by including the means from section 3.2: “The variability in ecosystem service indicators across scenarios was especially high for vegetation carbon stocks (9.2% ± 4.1%) and crop production (31.2% ± 12.2%).”

L229: Could you clarify how C storage from CCS was quantified, given that BECCS is not represented in LPJ (table S1)?

Following Krause et al 2017, we assumed 80% of the harvested C from bioenergy crops to be captured and stored. This may be optimistic, but was similar to the assumptions in MagPIE and IMAGE. We will add this to the revised text in section 2.3.

“… LUMs, with cropland also including bioenergy areas and pasture including degraded forests rangeland and grazing land. Following Krause et al. (2017) we assumed in the BECCS scenarios 80% of the harvested C from bioenergy crops to be captured and stored.”

L244: it would be nice to include table S2 in the main text.

This could certainly be done, and we would change table references in the main manuscript then accordingly (Table S2 would become Table 2).

L256-260: this part would fit better in the method section. Moreover, it would be good to include a cross-walking table in the method section to explain how the land classes of the respective LUMs/IAMs were translated into the 3 main types in LPJ. i.e., this needs to be clarified not only for CLUMondo.

Good point, we will integrate these into methods and suggest to add the following table on how LUM scenario classes were translated to the LPJ-GUESS land cover classes to the SI.

Table S2. Translation of LUM land use information to three LPJ-GUESS land use types.

LPJ-GUESS landcover / LUM landcover	cropland	pasture	natural
CLU-Mondo	regionally varying composition of each CLUMondo land use system in natural, pasture and cropland area following Eitelberg et al., 2016		
IMAGE	cropland	pasture	forest, urban, other natural
MAgPIE	cropland (irrigated, non-irrigated)	pasture	forest, urban, other natural
LUH1	cropland (inclusing bioenergy cropland for 26BE scenario)	pasture	primary vegetation, secondary vegetation, urban
LUH2	C3/C4 annuals, C3/C4 perennial, C3 nitrogen fixing	managed pasture, rangeland	primary land, secondary land, urban

Fig 3: please add the units for all variables.

Fig. 3 shows the change and uncertainty in ES indicators in % as is stated in the figure caption. Adding "[% change]" behind each ES indicator name would crowd the Figure with a lot of txt. However, we suggest to place the unit to a more prominent position in the revised caption to better highlight it.