The paper presents and explores simulations with the coupled land use and vegetation model LandSyMM to quantify future land use change and resulting impacts on ecosystem service indicators. There is a lot of interesting and thought-provoking material here, and I am sure this paper will create a lot of interest. However, like many "future scenario" papers, there is a lack of consideration of plausibility or uncertainty. The authors do not help the reader to understand why these projections are better or more reliable than other estimates. The section on runoff and flood risk is not convincing, in part because the separation of responses takes no account of what is already known about impacts of CO2 changes on runoff and in part because no case is made for using mean P95month as a measure of flood risk in a model with no water redistribution instead of relying on a projections using explicit hydrological modelling. The writing style is overblown (particularly in the Introduction) and often obscure (for example in the methods section and in the results section). The messages here could be conveyed in a clearer fashion with some pruning and rewriting, and this would considerably improve the readability of this paper. Shortening the existing text would leave room for a proper discussion section that would allow key issues to be explored. I hope the specific suggestions below can help the authors improve this paper and clarify their arguments, because the reliable estimation of future changes in ecosystem services is important for many purposes and people.

We thank Prof. Harrison for her detailed and helpful comments.

Specific Comments

The Methods section is long, difficult to follow and at the same time does not give sufficient information to allow these experiments to be repeated. I think this needs rewriting, focusing on the information that is really needed to understand what is going on.

This section has been significantly reworked:

- Sect. 2.1 now focuses on LPJ-GUESS, with the text on ecosystem services having been moved to the Introduction and the new Sect. 2.5.
- Sect. 2.2 focuses (as before) on PLUM, now including some text about where it uses data from and gives data to LPJ-GUESS.
- Sect. 2.3 describes how the coupling works. This is necessarily very technical, but a flowchart figure is now provided for clarification.
- Sect 2.4, describing input data and scenarios, has been compressed significantly relative to the old Sects. 2.3.1–2. Technical information regarding data sources is now less prominent, with about half of the section serving instead to provide context about the SSPs and RCPs. For interested readers, the Supplementary Methods provide more technical detail.
- Sect. 2.5 focuses now solely on the ecosystem service indicators used in the study. Background information on the ecosystem services in question has been moved to the Introduction.

I think it might be helpful to provide a paragraph at the beginning of this section to explain the logic of the order of presentation – I found some information I expected in one section in somewhere else completely, for example.

Hopefully the reorganized Section 2 will avoid such issues.

Some of the information presented could be summarised in the form or a table and/or flowchart diagram, and this would certainly be helpful.

A flowchart is now included, as is a table describing the various experimental runs.

Section 2.1 LPJ-GUESS description. Given how important these simulations are for downstream results, it would be helpful to give a more detailed description of how the model simulates crops (i.e. what are the differences between the treatments of each crop type), how nitrogen limitation is handled, what information is used to specify nitrogen inputs to cropland etc. The information about how irrigation, water demand, water supply, and plant water stress are simulated may well be described in Alexander e al. (2018) but since these are crucial to the current simulations, the approach should be briefly described here. Even the description of how the model simulates natural vegetation types is given short shrift here, so that the claim that it handles CO2 fertilisation is unsupported.

In a revised version, we will add information about the performance of LPJ-GUESS relative to other dynamic global vegetation models with regard to primary production, CO_2 fertilization, and nutrient limitation. We will also briefly describe how N limitation and irrigation work in LPJ-GUESS. The fertilizer input datasets are described in the revised Sect. 2.4.

It is also unclear from the present description how some of the service "proxies" are calculated by the model. For example, how does LPJ-GUESS simulate runoff? Please provide a better description of how the model works, so that it is easier to understand its strengths and limitations.

The revised Sect. 2.5, excerpted here, better describes how LPJ-GUESS simulates some of the ecosystem service indicators we use:

> LPJ-GUESS simulates a number of output variables that here serve as the basis for quantifying ecosystem services. The carbon sequestration performed by terrestrial ecosystems is measured as the simulated change in total carbon stored in the land system, including both vegetation and soil. Ecosystem nitrogen in LPJ-GUESS is lost in liquid form via leaching (a function of percolation rate and soil sand fraction), and in gaseous form through denitrification (1% of the soil mineral nitrogen pool per day) and fire. Here we combine these into a value for total N loss. LPJ-GUESS also simulates the emission of isoprene and monoterpenes—the most prevalent BVOCs in the atmosphere (Kesselmeier and Staudt, 1999)—and accounts for three important factors regulating their emission: temperature, CO_2 concentration ([CO_2]), and changing distribution of woody plant species due to climate and land use change (Arneth et al., 2007b; Schurgers et al., 2009; Hantson et al., 2017).
>
> LPJ-GUESS simulates basic hydrological processes such as evaporation, transpiration, and runoff. The latter is calculated as the amount of water by which soil is oversaturated after precipitation, leaf interception, plant uptake, and evaporation.

Section 2.1 Ecosystem services. Most of Section 2.1 is given over to a description of ecosystem services. What I was expecting here was information about what model outputs were used as indicators of specific ecosystem services. However, much of the text describes why a particular service is important – which should have been information provided in the introduction and indeed partly is provided there. The description of the simulated index is brief and uninformative. What I think would be more helpful would be to reshape this in the form of a table, listing the service and the model output (or outputs). This would save some space which could usefully then be used to provide more details in the model description so that it is clear how these outputs are obtained.

The new Section 2.5 is focused solely on how ecosystem service indicators are calculated. We have not provided a table, but hopefully the information should now be well-organized enough that one is not necessary. Background information on ecosystem services has been moved to the Introduction.

Section 2.2. Description of PLUM. Although a detailed explanation of the model is given in Alexander et al. (2018), it would be really nice to know a little more about it here. In particular, I am intrigued about the interface between the two models. What is the handshake, for example, between the four crop types in LPJ-GUESS and the seven crop types in PLUM? This is not explained here, nor is it explained in the description of the simulations.

The flowchart (Fig. 1) now points to this information, which is in the Supplementary Methods. Since this is rather technical model detail that has been covered previously (Alexander et al., 2018), we have decided not to put this information in the main text.

I do not understand how the crop demand optimisation works, and in particular whether this involves considering surpluses and surplus distribution (which should affect commodity prices) or whether it is assumed that there is always a surplus.

Text in the Introduction explains that PLUM "allow[s] short-term over- and under-supply of commodities relative to demand (rather than assuming market equilibrium in every year)." Text has been added to Sect. 2.2 saying that PLUM allows for short-term resource surpluses and deficits, and explaining the importance and novelty of this feature.

We consider other information regarding the optimization overly technical for most readers; those interested can find complete descriptions in the previous works cited in Sect. 2.2.

Section 2.3. Given the complexity of the experimental design, the complicated linking of different models, and the multiple sources of inputs, I think it would be extremely helpful for the reader if you included some kind of flow chart here to guide us through.

A flowchart is now included.

Section 2.3.3. The factor separation experiments are not well designed. Recycling 30 years of climate is not equivalent to a constant climate. As the results of the FireMIP experiments

show, it is difficult to compare these constant climate experiments with constant other experiments when the constant other is based on a single year.

Sect. 2.3 now explains, in the text as well as a footnote of the new Table 1, what actually goes in to the "constant-climate" run. New text in **bold**: "By holding either climate, atmospheric CO_2, or land use and management constant **(or for climate, looping through 30 years of temperature-detrended historical forcings)** over 2011–2100, …" While we acknowledge that looped climate such as we used can introduce artifacts that would be avoided by a random-sampling approach, we believe that clearly explaining this distinction would require too much space and would be overly technical.

Furthermore, the value of treating all climate variables as a single input seems a bit odd when thinking about productivity – it would be more interesting to diagnose what aspects of climate are crucial. In any case, a better factor-separation approach is needed. Alternatively, given that. these results are "mostly not presented" (line189) you might leave this out.

An experimental design to separate the influence of different climate variables would add some rigor, but it would also entail many more model runs, as well as the generation of new climate input datasets for LPJ-GUESS. We thus consider it beyond the scope of the present study.

Results and Discussion section. There is a lot of detail here, but the selection of things to highlight seems somewhat arbitrary. This is particularly the case in the delineation of geographic areas (what, for example, is meant by South Asia?). I was, for example, somewhat surprised by the lack of commentary on changes in China. Given that these kinds of assessments are of largely political interest, I wonder whether there should be some refocussing here – away from biggest changes to most important regions?

"South Asia" is now defined. If asked to submit a revised manuscript, we will add a bullet point discussing China, as befits its geopolitical importance.

Some thought should also be given to tabulating results.

The values provided next to the bars in the two bar graph figures are intended to serve this function while saving space relative to what would be required for a separate table.

I would strongly advise separating out the Results from the Discussion, creating a separate section. There are many issues affecting the results presented here, including the impact of methodological uncertainties, that really need to be discussed more fully in this paper. I am not suggesting that these issues invalidate the study, but I think it would be helpful to discuss the sources of uncertainty and I suggest that you add a Discussion section, where you can do this.

- *How sensitive are the results to specific inputs?*

  We considered a comprehensive evaluation of uncertainty related to climate model choice and PLUM parameter selection to be beyond the scope of this study.
- *what is the impact of mixing static and time-varying inputs?*

We acknowledge that looped climate such as we used for the "constant-climate" experiments can introduce artifacts that would be avoided by a random-sampling approach. However, we believe that exploring the possible impacts of this methodology would take too much space in an already lengthy paper, and in any case could not be properly quantified without additional model runs.

- *given that there are large differences between vegetation models in terms of their predictions, how reliable are the LPJ-GUESS productivity estimates? or perhaps, where are they situated with respect to other models? and how much does this matter to the final assessment? … A second issue that could usefully be included is "CO2 fertilisation" – given that this still appears to be controversial, that there is confusion about this is photosynthesis or WUE, that different models produce different strengths of fertilisation and so on.*

  In a revised version, we will add information about the performance of LPJ-GUESS relative to other dynamic global vegetation models with regard to primary production, CO_2 fertilization, and nutrient limitation.

- *How serious is the mismatch between PLUM outputs and the scenarios? How much of an impact does this have on the projections?*

  In a revised version, we will add a few sentences to the results explaining that the harmonization causes strong changes in the PLUM land-use area maps in only a few regions, and most of those discrepancies are reduced dramatically by the end of the century. We will also add a figure to the Supplementary Results illustrating this.

One additional issue that could usefully be included in the Discussion, but certainly needs to be treated somewhere, is the assumption that increased fertilisation will always produce an increase in production rather than a saturating relationship, shown by analyses of field data.

LPJ-GUESS actually does simulate, and PLUM does assume, yield as a saturating function of fertilizer application. This is now mentioned in the first paragraph of Sect. 2.2: "PLUM assumes that irrigation and fertilizer produce diminishing returns, such increasing them increases yield at low intensity levels, but less and less so at higher levels, approaching a yield asymptote."

In comparing LandSyMM results with other models, it would be useful to include a discussion of the. plausibility (or otherwise) of their/your assumptions. This would also deal with the questions: given that there are other simulation results, what does this paper add? and why should we believe the results are more plausible?

The text of the last paragraph in the Introduction has been modified to highlight advantages of LandSyMM relative to other model systems. It now reads (new/edited text in **bold**):

> "…. This coupled model system—the Land System Modular Model, or LandSyMM—is **among the state of the art in global land-use change models due to the high level of detail that it considers in the response of agricultural yields to management inputs. Whereas most integrated assessment models rely on generic responses of yield to changing climate, atmospheric carbon dioxide, and fertilizer, LPJ-GUESS simulates these processes mechanistically. Land use optimization also**

> **happens at a finer grain in LandSyMM (about 3400 gridcell clusters) than in other similar model systems (tens to hundreds of clusters). Finally, LandSyMM is unique in that PLUM allows** short-term over- and under-supply of commodities…"

I would seriously consider taking out the section 3.2.2, but in any case it needs rewriting. Runoff. The impact of CO2 on runoff is going to be strongly dependent on whether we are talking about semi-arid regions or not, and there is now considerable literature on this (which should be cited). I think a more logical way to organise this section would be around climate regions.

While it is true that CO_2 impacts on runoff are strongly regionally-dependent, we feel that describing its effects in our results for each climate region would require too much space relative to this issue's importance to this study. In a revised version, we will add some brief text and citations acknowledging the regional variation in the CO_2-runoff relationship.

The transition from global runoff increasing to "flood" and "drought" risk is abrupt and it would be helpful to actually explain regional patterns of runoff change first. The fact that LPJ-GUESS is not a proper hydrological model, i.e. it does not transfer water between grid cells, it does include groundwater recharge, it does not include surface storage etc. etc. is mentioned in passing here (line 345). But this is a key issue about what "runoff" means and what "flood risk" means. This has been alluded to earlier on by referring to meteorological flood/drought, but it potentially very mis-leading – not for the immediate readers of the paper but certainly for the "assessments" that will pick these results up and re-use them.

While LPJ-GUESS is not a full hydrological model, its predecessor model LPJ has been shown to perform comparably to such models at the basin scale, at least at the time of publication of Gerten et al. (2004). Since the simulation of runoff in LPJ-GUESS has not changed significantly since then, we feel confident enough in our results at the basin scale to leave this section in. However, we have removed all reference to non-basin-aggregated results. Text explaining this has been added to Sect. 2.5.

The following text has also been added to Sect. 2.5, clarifying that while the definitions of "flood risk" and "drought risk" used here are imperfect, they have been used many times previously in the literature:

> As Asadieh and Krakauer (2017) note, these metrics do not translate directly into impacts due to the mitigation capacity and nonlinear effectiveness of reservoirs, flood control mechanisms, and other infrastructure, as well as changes in demand and mean climate. However, changes in streamflow extremes have served as rough indicators in a number of previous global-scale studies (e.g., Tang and Lettenmaier, 2012; Hirabayashi et al., 2013; Dankers et al., 2014; Koirala et al., 2014).

To clarify the proper amount of meaning with which the reader should consider these results (referring the reader back to the new Sect. 2.5 text above), as well as to smooth the transition between results regarding average runoff and extremes, the following text has been moved/added to create a new second paragraph in Sect. 3.2.2 (new text in **bold**):

> Such regional patterns in runoff change are arguably more important than global means, since impacts of low water and flooding are actually felt at the level of individual river basins. **To evaluate regional impacts, we calculated how much land area was subjected to intensified**

wet and/or dry extremes (Sect. 2.5). As discussed in Sect. 2.5, these values should not be taken as direct measurements of flooding or drought impacts, but they do serve as useful indicators.

We have also added brief explanations of meteorological and socioeconomic drought where those terms are introduced.

The logic of focusing on biodiversity hotspots is different from the logic employed with other ecosystem services, in the sense that with the other services you allow for increases/decreases and for changes in geographic regions where increases/decreases can happen. Wouldn't this be a useful approach here too? Is it possible that there would be increases in biodiversity in some regions that are not currently considered hotspots?

Yes, it's possible that increasing area of non-agricultural land could lead to a long-term increase in biodiversity in some regions. However, it's not possible to say where biodiversity is currently "limited" by available land—i.e., where, with enough available land, vegetation communities would see sufficient richness of vascular plant species to qualify under the CI definition. Text to this effect has been added to the explanation of our "biodiversity" indicator metric.

Conclusion. If you split the results and discussion section into two, then you could consider including the conclusions in your discussion section. The current conclusions are not very startling (storylines with high socioeconomic challenges to climate change mitigation consistently have the most severe consequences for ecosystem services) or are simply a repeat of how important this information could be (which was already in the introduction).

We have opted not to make a separate Discussion section, instead incorporating the additional discussion suggested in comment by Prof. Harrison and others into the Methods or Results. However, in a revised version, we will add some text to the Conclusions about the various elements of uncertainty that need to be explored in future work, including PLUM parameter uncertainty, vegetation and economic model choice, and selection of global climate model. This will allow the Conclusion section to be less repetitive than in the initial version of the manuscript.

Minor comments

Line 15-16. The statement about future population changes is expressed rather badly and is difficult to grasp, please rephrase.

The clause between the em dashes has been changed to: "with a population increase by 2100 ranging from 1.5 billion to nearly 6 billion people (KC and Lutz, 2017)".

Line 25. Is this really a feedback sensu stricto?

Yes: Land-use change and management affect climate via greenhouse gas emissions and biogeophysics, climate change affects agricultural productivity, changing agricultural

productivity affects land use and management, affecting greenhouse gas emissions and biogeophysics, etc. We do not (yet) model this in LandSyMM, but it is indeed a feedback.

Lines 47-48. The processes operate on the plant functional types rather than among them. Can you rephrase this to describe the model more clearly.

Here "among" has been changed to "for".

Line 51. When you say C3 cereals sown in winter and spring, presumably these are considered as two PFTs, so it would be clearer to say "C3 cereals sown in winter, C3 cereals sown in spring ...."

This change has been made.

Lines 68-75. It is impossible to judge whether these measures provide reasonable proxies for water availability, freshwater ecosystem condition, or flood risk because there is no information on how runoff is generated in LPJ-GUESS. is runoff simply the difference between P and ET in a gridcell? or is there transfer of surface water between gridcells? is there a contribution from groundwater?

The runoff paragraph in the "ecosystem services" section has been edited to clarify: "LPJ-GUESS calculates runoff as the amount of water by which soil is oversaturated after precipitation, leaf interception, plant uptake, and evaporation; note that runoff flow is not modeled (e.g., from one gridcell to another)."

Line 75. If you mean that hydrologic drought is not the same as meteorologic or socioeconomic drought, why not simply say so? This sentence is unnecessary, and begs the question: what is e.g. socioeconomic drought.

Meteorological and socioeconomic drought are now briefly defined.

Line 76-81. How does LPJ-GUESS calculate total nitrogen loss? Do you separate out nitrogen loss from natural ecosystems and agricultural systems?

The following sentence has been added to Sect. 2.5: "This is the combined rate of dissolved nitrogen losses (a function of percolation rate and soil sand fraction) and gaseous losses from denitrification (1% of the soil mineral nitrogen pool per day) and fire."

Line 82. The linking of climate change and human health here led me to believe that you were going to look at ecosystem services that mitigated the impact of climate change on human health. Apart from the mention that BVOCs affect ozone which in turn can have impacts on health, you don't really go into this in any depth. For example, you don't mention e.g. mineral dust and the role that vegetation plays on mitigating dust emission pace China. Perhaps changing the emphasis here to plant emissions (which have multiple effects, including on climate and on health) would be a better way to introduce this section.

"Human health" and "ecosystem services" have been swapped at the beginning of the first sentence of this section.

Line 94-102. I can see why the focus on the hotspots is attractive, but in this modelling framework is would also be possible to make a more general assessment of biodiversity loss and this would also be valuable.

A more comprehensive evaluation of the biodiversity impacts of land use change is indeed possible in this framework, but since this paper is broadly-focused, we have decided to not do that here. We believe that effort to be more appropriately directed at a paper focused specifically on biodiversity.

Line 109. The plant name Miscanthus should be in italics.

This has been corrected throughout the paper.

Line 115. Is 500 years really sufficient to bring the carbon pools into equilibrium? or is the phrase a realistic starting point mean to imply that they are not necessarily in equilibrium?

Spinup information is now located in the Supplementary Methods, Sect. SM1:

> All runs are preceded by a 500-year spinup period using a temperature-detrended version of the relevant climate forcings (CRU-NCEP v7 CRUp for the calibration run; IPSL-CM5A-MR for the yield-generating and PLUM-forced historical runs.) This includes a routine that analytically solves for equilibrium soil carbon content, bringing carbon pools into equilibrium before the beginning of the actual run.

Line 125-126. This first sentence should be moved into the description of the model, as confirmation that PLUM works reasonably well. It is not relevant to a description of the modelling protocol.

This has been taken care of as part of the Sect. 2.3 overhaul.

Line 127. Given that PLUM has 7 crop types and LPJ-GUESS only four, how do you input PLUM land use into LPJ-GUESS?

This information can be found in the Supplementary Methods. The overhauled Sect. 2.3 and new flowchart (Fig. 1 in revised text) point interested readers there.

Line 129. Please can you bring this flowchart and the table into the main text?

A flowchart is now in the main text.

Lines 133-135. Please indicate that the details for these sensitivity tests are given in a following section and reference the section here.

The following has been added to Sect. 2.3: "Details regarding the inputs of these experimental runs can be found in Sect. 2.4 and the Supplementary Methods."

Line 139. Surely this should be: Viovy, N. 2018. CRUNCEP Version 7 – Atmospheric Forcing Data for the Community Land Model. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. http://rda.ucar.edu/datasets/ds314.3/.

That appears to be a current version of the dataset, but we accessed the data in a different format, from a different server at a different time. We have added a corresponding citation to: Viovy, N.: CRUNCEP Version 7: Atmospheric Forcing Data for the Global Carbon Budget 2016, 2016. A footnote in Supplementary Methods Section SM1 gives the URL and date of access.

Line 140. Either spell out what these problems are or refer to a paper that does. Maybe Tang et al. (2017)?

This is now explained in a footnote in Supplementary Methods Section SM1: "The CRU-NCEP algorithm was designed to match CRU TS3.24 monthly precipitation totals, but it produced unrealistically high numbers of wet days—days with precipitation of at least 0.1 mm—in the tropics and boreal regions in the early part of the 20th century."

Line 147-149. I am having difficulty with this description. You use time-varying allocations of cropland area per gridcell but a static data set of what these crops were. How did you apply this? Simply assuming that the area might change but the crop remains the same? How much uncertainty does this introduce in your results? Unless you address this in a Discussion section (as suggested above) you need to say something here.

This is clarified in the last sentence of what is now Sect. 2.4: "Historical crop distributions (i.e., given LUH2 cropland area in a gridcell, what fraction was rice, starchy roots, etc.) came from the MIRCA2000 dataset (Portmann et al., 2010) and were held constant throughout the historical period."

Line 148. Is this still in prep.?

Yes.

Line 154-155. This sentence is a bit unclear. The manure N is held constant in the calibration run but varying in the other runs?

Yes, that's the correct interpretation—probably hard to understand because of a typo. The sentence (now in Sect. SM1) now reads as follows: "Manure N was added in the historical period according to the annually-varying maps given in Zhang et al. (2017b), but in the calibration run was held constant at year 2000 levels to match the use of the AgMIP fertilizer data."

Line 179-180. Did you estimate these values or are they provided?

We estimated them.

Line 184-185. Recycling 30-years of climate is not "constant climate"

This is true, but we consider "constant climate" to be an acceptable shorthand that should not mislead a reasonably careful reader.

Line 210. "first two or so ...." please state what period it actually decreases over.

The relevant part of this sentence has been changed to: "… in SSP5-85 it decreases through about 2050, after which it increases slowly, ending at a slightly lower global extent …"

Figure 2. This is unreadable at the size reproduced here and, with the grey background, the paler colours are not sufficiently visible. You need to find a way of making the changes more visible. Maybe splitting this into two figures would help. (Note the same comments apply to Figure 4, 5).

This figure has been updated to use discrete colors (rather than a gradient) with the gray now darker to improve visibility. It has also been enlarged to fill the page (pending editorial approval to exceed the "two-column" width of 12 cm), and rearranging labels has allowed minor additional enlargement.

Line 222. I confess that I find this agricultural expansion in Alaska a bit implausible, even for a high-end scenario, given the topographic constraints and the issue of permafrost. I would be intrigued to know when the permafrost disappears in this scenario. And how infrastructure (or lack of it) would impact this expansion?

According to the US Department of Agriculture[1], there are already several hundred farms in Alaska. As the climate warms, permafrost extent is expected to decline across the Northern-Hemisphere boreal zone, especially in RCP 8.5, suggesting that more area might become arable. While the version of LPJ-GUESS used in this study does not have a complete representation of permafrost dynamics, it does include limitations on various plant and soil processes based on air and soil temperature. Thus, LandSyMM might be overly optimistic with regard to the arable area in Alaska by the end of the century, but we do not feel it to be qualitatively implausible. Text to this effect (including citations of two papers projecting permafrost extent) has been added to the first bullet point in Sect. 3.1.

PLUM does not account for limitations on expansion due to lack of infrastructure, implicitly assuming that if conditions are appropriate—in terms of production capacity given demand—for production of agricultural commodities, the necessary infrastructure will follow.

Line 228. Since South Asia is not a widely-recognised geographical term, it would be helpful to define where exactly you mean here. Are you including southern China here?

We use "South Asia" to refer to a set of PLUM country groups: India, Sri Lanka, Pakistan, Afghanistan, Bangladesh, Nepal, and Bhutan. Text has been added explaining this.

Line 234. What climate change produces more favourable growing conditions in South Asia?

The relevant sentence has been edited to read (changed text in **bold**): "… it also depends markedly on yield boosts due to **increased rainfall (Fig. SR1)** and rising CO_2 …"

---

[1]https://www.nass.usda.gov/Publications/AgCensus/2017/Full_Report/Volume_1,_Chapter_1_State_Level/Alaska/st02_1_0009_0010.pdf

Line 236. Even larger ... even larger than what?

This clause has been changed to “Sub-Saharan Africa sees crop production increases even larger than South Asia”.

Line 250 et seq. I find this discussion of other model results here confusing. I think you want to separate this from the presentation of all the results from your experiments and move this type of comparison into a separate discussion session. This would allow you to discuss the plausibility of the other assumptions compared to the assumptions encapsulated in your simulations.

The advantage of our current structure is that it allows us to immediately address the most striking pattern in our maps of projected land use change, which is the pasture expansion in central Africa. As we explain, there are reasons behind the patterns that we see in our land use trajectories. We want to provide those to the reader here so that readers have the appropriate context for interpreting the rest of our results.

Line 267. I do not understand what you mean by "friction" here.

We have replaced “‘friction’” with “cost”.

Line 269-274. Please take out this speculation about the impacts of including forest products in LandSyMM based on work that has not yet been done.

We appreciate that the current phrasing is overly speculative regarding work in progress. However, we feel the idea expressed here is important to fully explain the issue at hand. We have thus changed “Work currently underway to include… may” to “Including… could”.

Line 275-283. And so what? You appear to be saying that you have different results from one study because they used unrealistic inputs, and that you have different results from another study because they made a set of different assumptions. In the first case, perhaps you could assume that a reader might guess that your results are "better", although you never actually use the "unrealistic" word. In the second case, however, you might give use a hint about the assumptions made by IMAGE would produce more/less realistic results and why.

We acknowledge that most readers will probably recognize why the results of this study are different from and more internally consistent than those of Alexander et al. (2018). However, as the explanation takes only one sentence, we have decided to leave it in.

We do not consider LandSyMM more or less “realistic” or “plausible” than other state-of-the-art models. It may be that assumptions similar to those made in IMAGE (such as deviation from historical GDP-diet composition relationships) would be necessary in order to restrict PLUM to a solution space that satisfies the radiative forcing values of each RCP scenario; however, LandSyMM does not yet include a climate model, and so we cannot yet assess that possibility. While we do not include forestry or payments for carbon storage, LandSyMM does have other advantages, as explained in the text. We thus consider this work

to be another contribution to the body of research exploring possibilities for the future of land use and terrestrial ecosystems, and leave it to the reader to make their own judgment about relative plausibility if they care to do so.

Line 290. "intermediate carbon fertilisation ..." Not phrased felicitously, since it implies that C-fertilisation itself has multiple levels of working (off, half-on, on). Please rephrase.

To us, "intermediate" does not necessarily imply a measurement of discrete values. Rather, it simply implies "somewhere near the middle of two extremes," which allows for our usage in reference to continuous values.

Line 299. Sorry, I might have missed this – what do you do about the conversion of secondary vegetation to pasture in terms of carbon. Are we looking at gross or net here?

The following has been added to Sect. 2.2:

> Land use areas are calculated as net change, which neglects certain dynamics—such as shifting cultivation—that can have significant impacts on modeled carbon cycling especially in some regions (Bayer et al., 2017). Other ecosystem services could be affected as well. LandSyMM does not capture these dynamics, but this was considered an acceptable trade-off for computational efficiency.

Line 302-303. I am not sure why you are picking out one model from this study. I think you should give the range of estimates. I don't know whether the quoted value for IPSL-CM5A-LR is low-end or high-end.

The beginning of this paragraph has been modified to clarify (new text in **bold**):

> Brovkin et al. (2013) examined the change in land carbon storage over 2006–2100 for a number of climate and land surface models. **This included IPSL-CM5A-LR: the same IPSL-CM5A Earth system model that produced our forcings, except run at a lower resolution (hence, -LR instead of our -MR). They found that IPSL-CM5A-LR,** when forced…

Figure 3. Please don't abbreviate emission on this Figure.

In this figure (now Fig. 4), "emissions" is now spelled out.

Line 305. "probably"? You could establish this by looking at what the difference in loss of non-agricultural land between these simulations and yours is if you take out the pasture expansion and the expansion of cropland in Alaska.

"The difference" here refers not to the difference in area, but rather to the difference in C sequestration. We qualify with "probably" because while there are *definite* differences in terms of where and how much non-agricultural land is lost, quantifying the difference in *C sequestration* due to this would require maps of C stocks and fluxes for the Brovkin et al. model outputs.

Line 308 et seq. So the difference is caused by the differences in the scenarios, right? But later on you imply its because the models don't include nitrogen-limitation. I think you need to make it clear what you think is giving rise to these differences, scenarios or model set-up.

It would, of course, make it interesting to run your experiments with the older scenarios – and this would be helpful in terms of uncertainty analysis.

The text comparing our C sequestration results to those of Brovkin et al. (2013) is intended to convey that the differences could be due to *both* (a) differences in where and how much non-agricultural land is lost, as well as (b) the fact that photosynthesis is limited by N in our model but not those in Brovkin et al. (2013).

We performed the "back-of-the-envelope" calculation suggested by Prof. Harrison below (methodology explained below), which showed that only a small part of the difference can be explained by land-use change scenario. The end of the Brovkin et al. discussion now reads as follows:

> A rough estimate (not shown) shows that running LPJ-GUESS under RCP8.5 with the same land use change as Brovkin et al. (2013) would have increased total carbon gain by 10–15% at most. Instead, most of the difference is likely because none of the models in Brovkin et al. (2013) limit photosynthesis based on nitrogen availability.

Line 309–310. Your comparisons with other simulations are unbalanced – having described the results from the Brovkin et al study in some detail you here say that the results are low compared to the Nishina et al. (2015) study even when comparing just to the models in that study with nitrogen limitation. But no details. How low? what was the range simulated by the N-enabled models in Nishina et al.? Do you have any idea why you get a different result?

Upon re-reading Nishina et al. (2015), it was discovered that instead of assuming constant land use (as we first understood it), those simulations did not include land use at all. This explains the large difference between their results and ours, but makes the comparison rather trivial. We have removed the reference.

Lines 311-318. So, different models produce estimates less than LandSyMM as well as above. What do we learn from this? You imply this is because of differences in scenarios (while hedging your bets in terms of climate forcing), but what is needed here is a back-of-the-envelope calculation of whether the differences in cropland and pasture area would produce a comparable estimate in LandSyMM. If you wanted to be ultra-realistic, you could use the areas where they show biggest changes in area.

The following text has been added: "A rough estimate (not shown) shows that running LPJ-GUESS under RCP8.5 with the same land use change as Brovkin et al. (2013) would have increased total carbon gain by 10–15% at most." Because we did not save by-LU carbon pools, we estimated this by taking gridcell mean carbon density and dividing it by non-agricultural fraction to get non-agricultural carbon density, effectively making the extreme assumption that agricultural land had zero carbon. (This estimate thus produces an upper limit to the difference that would have occurred using LUH1 land use areas.) We then multiplied that carbon density by the area of non-agricultural land in the PLUM outputs and LUH1 for 2006–2010 and 2096–2100, and calculated the difference. We excluded grid cells where PLUM had <0.1% non-agricultural land.

Line 319. Photosynthesis scaling parameters ..... what scaling parameters?

The end of this sentence has been changed to read (new text in **bold**): "… different climate forcings and **a different photosynthetic scaling parameter (which accounts for real-world reductions in light use efficiency; Haxeltine and Prentice, 1996)**."

Line 338. Why did you not run it in coupled mode then?

Our group does not have a version of LPJ-GUESS coupled to a climate model.

Line 341. Please can you explain what was done in the Asadieh and Krakauer (2017) analyses. Were these full hydrological models?

We have added text explaining that Asadieh and Krakauer (2017) included full hydrological models.

Since this was an ensemble, presumably there is a range of estimates for at-risk of flood and at-risk of drought? Please give these ranges in the text

Asadieh and Krakauer (2017) only presented their multi-model average results, not the range of results across all models.

How much of a difference does using monthly versus daily values make to the estimates of area affected?

We have added a sentence to Sect. 3.2.2: "We expect that our results for land area with increasing and decreasing flood risk would have been lower and higher, respectively, had we used daily values for P95 as Asadieh & Krakauer (2017) did, instead of the LPJ-GUESS-output monthly values." Quantifying this difference does not seem possible without adding code to LPJ-GUESS allowing daily runoff outputs, then performing the runs again.

Line 349. How many classes are there? You need to spell out that you are talking about increases/decreases in flood risk and drought risk areas).

There are four classes, as given in what is now Table 2. A reference to Table 2 has been added to the sentence in question.

Line 379. Please explain why high CO2 suppresses BVOC formation and provide references here.

The following has been added here: "The exact cellular regulatory processes of this '[CO_2] inhibition' remain enigmatic; recent evidence suggests that reduced supply of photosynthetic energy and reductants plays a major role (Rasulov et al., 2016)."

Line 382-383. This sentence is confusing because it seems to say that boreal forests are causing declining monoterpene emissions – whereas I think the idea is that decreases in boreal forest area coupled with less effective BVOC production in the surviving boreal forest area are responsible. Please rewrite.

We have replaced "drivers" with "areas".

Line 393. Please italicise Miscanthus

This has been corrected throughout the manuscript.

Line 396 et seq. This paragraph states that predicting the effects of changing BVOCs is difficult because the model framework doesn't include atmospheric chemistry. I am not sure what "a surface-level discussion of possible effects" means here. You can predict potential changes in BVOC emissions, and so perhaps this is the point to stop at. There is no need to go further and speculate about what the impact of these changes might be on atmospheric chemistry, climate and/or health.

We think it is helpful for readers to be given some sort of context for the results, but acknowledge that this discussion is indeed speculative nature of this discussion. To make this clear, we have:

- Swapped the last two paragraphs of this section.
- Added the following to the end of what is now the second-to-last paragraph: "However, we wish to provide context for the benefits and detriments associated with changing BVOC emissions, as well as some limitations related to our model setup."
- Replaced the last sentence of what is now the last paragraph with the following (new text in **bold**): "Moreover, the loss of natural land is itself associated with myriad negative health impacts (Myers et al., 2013) **which are not simulated in LandSyMM**, so it would be shortsighted to view deforestation-induced BVOC reductions as a public health boon. **Testing whether and to what extent any of the mechanisms described in this paragraph would make a difference to regional climate and human health would require significant extension of LandSyMM, including the incorporation of new sub-models."**

Line 403. The fact that this one region is not defined as a hotspot, and that it has a big impact on the results, makes a good case for extending this analysis to consider changes in biodiversity everywhere.

Theoretically it would be possible to extend this analysis to areas not currently classified as hotspots by surveying the literature to determine the floristic diversity of all ecoregions and then—where the "at least 1500 vascular plant species" requirement is met—including all ecoregions that will have lost at least 70% of their natural vegetation by the end of the century. However, the effort that would require would better be spent on a more comprehensive analysis of not only area loss but corresponding extinctions (see below). Such an analysis would be valuable, but is outside the scope of this paper.

Line 410-415. Please rewrite this section to first make the comparison and then explain possible sources of differences. ... Line 415. Are you saying that species-area curves are an inappropriate tool for estimating extinction rates rather than species numbers?

The text has been edited to clarify that species-area curves are correct in accounting for how the number of species lost per hectare of land conversion decreases as total area converted increases. This comparison has also been rearranged to highlight the reason for citing Jantz et al. (2015) at all: to illustrate (a) the importance of this nonlinear relationship

and (b) that our analysis did not take this into account, meaning that our results do not correspond directly to extinction estimates.

Lines 419-420. I hadn't realised that climate changes and CO2 could have an effect on models too! Please rephrase this.

Hof et al. (2018) used a tool, species distribution modelling, that has not yet been mentioned in this manuscript. Such models use climate and land-use change as inputs.

Line 423. "We may see a similar effect ...." Please clarify: do you or don't you?

The following has replaced the part of this paragraph beginning "We may see…":

> We see a similar effect: If ignoring *Miscanthus* area, loss of natural land in CI+CSLF hotspots is reduced (respectively for SSP1-45, SSP3-60, SSP4-60, and SSP5-85) by about 100%, 45%, 39%, and 17%. However, because land cleared for biofuel is not available for other crops, a full accounting of the contribution of biofuel expansion to land conversion and thus biodiversity would require PLUM runs with no biofuel demand.

We have decided not to perform those extra PLUM runs, believing that effort would serve better in work more focused on the future impacts of land-use change on biodiversity rather than the more general review here.

Line 429-430. Considering that the results from LandSyMM have been compared to a range of other model simulations in this paper, the first sentence really doesn't make sense. Maybe this is more comprehensive in terms of the range of scenarios and the range of outputs, but what else is different here?

This first sentence is indeed intended to highlight this work's novelty due to its comprehensiveness; "comprehensively" has been added to stress this point. The beginning of the second sentence has been modified to highlight other advantages of LandSyMM as mentioned in the revised Introduction (new text in **bold**): "**Using a uniquely spatially-detailed, process-based coupled model system,** we show…"