***Interactive comment on* "Impacts of future agricultural change on ecosystem service indicators" *by* Sam S. Rabin et al. (Stephen Sitch)**

Excellent attempt to evaluate scenarios for implications on ecosystem services including trade-off and co-benefits. LandSyMM represents state-of-the-art in modelling of the land-use & vegetation. This type of work is extremely relevant for ecosystem service provision in the context of the Paris Agreement (food-energy-water nexus and beyond). Follows on from earlier studies, e.g. Krause et al., 2017.

Overall I enjoyed reading the manuscript and warrants publication. I found it very informative and represents a substantial amount of work and scientific progress. A minor edit is warranted in the methods section which I found confusing in places. It could be improved as it was difficult to work out exactly how the models link together and finally the number of runs made. The authors also need to clearly distinguish general text explaining processes from what's actually included in the models. I found the results/discussion section comprehensive and complemented with extremely informative figures. Interesting new advances on attempting to relate flood/drought (see specific comment below), and the link to land loss in biodiversity hotspots.

We thank Prof. Sitch for his kind comments and helpful suggestions. Section 2 has been extensively 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.

Minor comments:

28% nitrogen pollution -> aquatic systems, air pollution? Unclear from line 75-90 how nitrogen pollution is quantified/considered in the LPJGUESS (this section is more an introduction to N pollution)

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

Line 24 "As global environmental and societal changes continue to accelerate over the coming decades,". Somewhat vague. Are we sure changes are and will accelerate.

We have deleted "to accelerate" from the sentence in question.

LandSyMM represents state-of-the-art in modelling of the land-use & vegetation. However this advance is not entirely clear. Lines 35-40 could be improved to demonstrate the advance beyond existing IAMs. e.g. "is unique among global land-use change models in the high level of spatial detail that it considers in the response of agricultural yields to management inputs, as well as in allowing short-term over and under-supply of commodities relative to demand (rather than assuming market equilibrium in every year)" . Vague – what does high-level of spatial detail mean? I wonder if a table contrasting CMIP6 IAMs with LandSyMM would be useful.

The text of the last paragraph in the Introduction has been modified to read (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…

Also perhaps a real world example can be given to demonstrate the importance of the non-equilibrium assumption (i.e. is this a detail or fundamental). Perhaps the authors can elaborate more on the differences between PLUM and other LU model approaches in section 2.2.

A new paragraph has been added to Sect. 2.2 (PLUM) explaining the significance and highlighting the novelty of the non-equilibrium assumption in PLUM (new text in **bold**):

> To solve for land use areas and inputs that satisfy demand, PLUM uses least-cost optimization, which allows for short-term resource surpluses and deficits. **Such imbalances can be significant in the real world: Global supply of major cereal crops frequently swings 5 to 10% out of equilibrium on an annual aggregate basis, and more extreme imbalances can be seen at the scale of individual countries (FAOSTAT, 2018a). These dynamics are not captured by equilibrium models, such as those used in other land use and integrated assessment models, which represent for each year the stable state that the economic system would move to eventually if the environment did not change. Because global agricultural markets are not in equilibrium, disequilibrium models are needed to capture the real-world process of moving towards---but not reaching---equilibrium in a constantly-changing economic and physical environment. Disequilibrium models have received varying amounts of attention in the literature over time (e.g., Kaldor, 1972; Mitra-Kahn, 2008; Arthur, 2010), and to our knowledge PLUM is the first land use model to incorporate one.**

Very good coverage of LPJGUESS, however somewhat generic in places. The authors should explicitly state which metrics are calculated for ecosystem services, e.g. how is nitrogen pollution calculated? (at first it gives the impression of some combination of health, water quality, air pollution impacts on vegetation indices; are these processes modelled in LPJGUESS?), i.e. this text gives a more general description of nitrogen pollution, rather than what is actually modelled in LPJGUESS.

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

Good attempt, but I'm not entirely convinced of the approach for changing flood risk based on monthly runoff – I think this should be raised again in the discussion as a potential limitation/uncertainty.

The following text has been added to Sect. 2.5:

> 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).

The following text has been added to Sect. 3.2.2:

> As discussed in Sect. 2.5, these values are not direct measurements of flooding or drought impacts, but they do serve as useful indicators.

As noted by the authors they apply the Asadieh and Krakauer 2017 approach but use monthly surface runoff data. The authors should state the temporal resolution the model is applied (I assume monthly?). As I understand LPJGUESS outputs per gridcell 1) monthly total N Loss 2) monthly runoff 3) annual land conversion in hotspots, and then interpreted in terms of nitrogen pollution (but no additional metrics used), flood/drought risk (using AK 2017 metric), and loss of land in biodiversity hotspots.

We have added the following to Sect. 2.1 (LPJ-GUESS): "Hydrological and most physiological processes are modeled at daily temporal resolution; vegetation growth, establishment, disturbance (including land-use change), and mortality happen annually."

Somewhat confusing section 2.3 on simulation details. For example, Line 115-120 "In each scenario, every grid cell is planted with each crop type, each of which is given six different management treatments in a factorial setup: fertilization of 0, 200, and 1000 kgN ha–1 and either no irrigation or maximum irrigation." Why? Somewhat comes out of the blue.

Text has been added ("under a range of irrigation-fertilization treatments") to Sect. 2.2 to introduce the idea of this setup.

Then in 2.3.3 experimental setups Lines 180-190 "In addition to the LPJGUESS runs forced with PLUM-output land use and management trajectories (harmonized as described in Sect. 2.3.1), six experimental runs were performed for each scenario, to disentangle the direct effects of climate change (including CO2 concentration increases) from those of land use and management change." This is another 6 simulations?

Yes, those 6 are different runs (as opposed to treatments within one run). This is hopefully clarified by the overhaul of the Methods section.

Then there's PLUM-forced runs. So I assume this is not the standard approach to running LandSyMM (i.e. PLUM coupled to LPJGUESS), i.e. how frequent is information from LPJGUESS and PLUM exchanged (annually, e.g. potential yields line 167?) Perhaps consider a table with a full list of the simulations would help the reader.

The new Sect. 2.3 is now clearer in describing that the PLUM-forced runs are part of the way that LandSyMM is designed to be run. The fact that the LPJ-GUESS to PLUM coupling happens at five-year timesteps is now mentioned (first sentence, third paragraph of new Sect. 2.3). A table of simulations has been added (Table 1), and Table SM1 provides additional clarification.

Line 140 "The calibration run was forced with climate data from CRU-NCEP version 7 (Le Quere et al., 2016), but with CRU TS3.24 precipitation (Harris et al., 2014) due to problems discovered in the CRU-NCEP precipitation data" What were the problems? (others are using CRUNCEP7 so would gain from this information)

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 143 IPSL-CM5A-MR – what's the climate sensitivity of the IPSL model? Why was this selected? (what are the characteristic features of this GCM future prediction), for example which areas are projected to have higher / lower precipitation, as this will govern the simulated flood/drought risk (and affect the other ecosystem services studied?). For example, around line 327 perhaps add some text / speculate on impact of using one climate model for regional runoff. E.g. Figure 5, perhaps it would be useful to include a map of change in precipitation (and temperature) to give the reader a feeling for the importance of choice of GCM.

Ahlström et al. (2012) looked at 18 CMIP5 climate model outputs and used them to force LPJ-GUESS. We used that analysis as a guide to selecting which model's forcings we would use. Initially, we wanted to use MPI-ESM-LR, as it represented a middle-of-the-road in terms of both mean land temperature rise and net ecosystem exchange response in LPJ-GUESS (Ahlström et al. 2012, Fig. S3). However, the MPI-ESM-LR outputs were not available for RCP6.0, which we needed for two of our RCPs. We instead chose IPSL-CM5A-MR, which had all RCPs available and is on the low side of the high end in terms of mean global temperature change. However, IPSL-CM5A-MR does simulate a large increase in precipitation around the Equator (ibid., Fig. S2).

Bars for global precipitation change have been added to what is now Fig. 2. This is now referenced at the beginning of Sect. 3.2.2. Additionally, a figure showing maps of mean change in temperature and precipitation for each RCP in this study has been added to the Supplementary Results. A reference to this figure has been added to the Methods (Sect. 2.4).

The following text has been added to the end of Sect. 3.1.1:

> Krause et al. (2017) used climate forcings from the IPSL-CM5A-LR model, which differs from what we used (IPSL-CM5A-MR) only in that the former was run at a lower resolution. Both have similar mean global land temperature changes: for RCP8.5, on the low side of the high end of 18 CMIP5 models examined in Ahlström et al. (2012). This temperature change is strongly correlated with net ecosystem carbon exchange (land-to-atmosphere C flux, excluding fire emissions), so a different choice of climate forcings could have resulted in a stronger C sink or even a C source (Ahlström et al. 2012, Fig. S3).

The following text has been edited in Sect. 3.2.2 (changed text in **bold**):

> Some differences between our results and those of Asadieh & Krakauer (2017) might be expected because we used monthly instead of daily values for P95**. Also, whereas that study used five climate models, we used only one---specifically, one that simulates a much larger precipitation increase around the Equator (in RCP8.5) than 18 other models examined in (Ahlström et al. 2012, Fig. S2). Finally,** LPJ-GUESS is not a full hydrological model: e.g., it does not include river routing.

Line 146 onwards. "Time-evolving historical land use fractions—i.e., the fractions of land in each gridcell that are natural vegetation, cropland, pasture, or barren—were taken from the Land Use Harmonization v2 dataset (LUH2; Hurtt et al., in prep.). The MIRCA2000 dataset (Portmann et al., 2010) provided crop type distributions for the year 2000, which were used for all historical years." I'm a bit confused what is used here. So LUH2 give the cropland coverage, but not which crops, that's given from MIRCA2000 and relative proportion of individual crop types stay fixed through time over the historical period?

Yes, that's the correct interpretation. This should now be clarified in text at the end of the new Sect. 2.4.

Lines 168 onwards describe the SSPs. I think this generic text would work better if it came earlier (SSPs have already been mentioned in several places already).

As part of the Methods section overhaul, the SSPs are now introduced at a more appropriate point.