Authors’ response

We thank Prof. Sitch, Prof. Harrison, the two anonymous reviewers, and the Associate Editor Dr. Dekker for their helpful comments on the manuscript.

Some edits have been made since our initial replies. To reflect this, here we have included our replies with changes tracked (deletions in ~~red strikethrough~~, insertions in blue).

Reply to Reviewer 1

This article presents an ambitious modelling exercise that combines the LPJ-GUESS dynamic global vegetation model with the PLUM land-use model (now under the new name LandSyMM). While the model has been presented before in Alexander et al 2018, the coupled models have not been presented before from the LPJ-GUESS perspective which makes this an interesting article. Additionally, the large number of ecosystem service indicators make the presented analysis interesting on its own.

Before the article can be published in ESD however the structure of the article needs substantial improvement. While I acknowledge the challenge of describing such a complex model in a comprehensive as well as concise manner, I do think improvements can be made. I have three general points and a number of more detailed points that need to be addressed.

We thank the referee for their helpful comments.

The structure of the methods section is confusing. I would recommend to start with an overview of the models used and how they interact (a diagram might be helpful), then a detailed description of the different ecosystem service indicators analysed.
Also, I would recommend to reduce the size of the methods section by moving some of the information on detailed input for LPJ-GUESS and PLUM to the SI. On the other hand, much more information is required on the scenario setup. Details about the main assumptions should be presented. A table with an overview would be helpful. It is important that the reader does not depend on a different article to understand how the presented scenarios are defined.

The extra experimental scenarios that have been performed to improve understanding of the results are confusing. Maybe these can be added to a descriptive scenario table, either in the main text or the SI, and they could be given explicit names to make it easier to refer to them in the results section, e.g. SSP4-2.6-noCO2/noLUC. I think this is better than for example the long text 'in the constant climate+CO2 experiment' (line 295-296).

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.

A LandSyMM overview diagram (Figure 1) has been added to Sect. 2.3. Experiment nomenclature has been standardized and is explained in the new Table 1 (Sect. 2.3).

The results and discussion section is rather lengthy and is (as the names correctly states) a mix between results and discussion. I think this reduces the clarity of your story and is not in line with the standard outline of scientific articles. I think the article would greatly benefit from splitting up this section into a clear description of the results (topic by topic) and subsequently a discussion of the results in the context of the literature (again, topic by topic).

We agree that it would be more in line with common practice to move such comparisons to a separate Discussion section, but while this works well for most papers, we believe it would not in our case. First, some of our results with regard to land use area are so striking (and different from results from similar work) that it makes sense to address them immediately. This also provides the reader with valuable context for interpreting the rest of our results. Finally, because this paper touches on so many different ecosystem services, it is necessarily rather long. Postponing the literature comparison to a separate Discussion section would necessitate spending space there reminding the reader of our own results; in the interest of keeping this paper as concise as possible, we have chosen to avoid that.

Detailed comments:

Line 9-10: this is the first time biodiversity is mentioned in the abstract, while it is presented as one of the major outcomes.

We have added "biodiversity" to the list at Line 5.

Line 15: please rephrase 'larger than today's by anywhere from' to something more concise and more academic, e.g. 'an increase ranging from 1.5 billion to 6 billion'

This clause now reads "with a population increase by 2100 ranging from 1.5 billion to nearly 6 billion people".

Line 57-102: this section describes the ecosystem service indicators presented in the article. This should be a separate section. In addition, a lot of the text is introduction to why these indicators are important. I think this reduces the clarity of the methods section which should be a more technical description of the indicators presented in the article. Maybe you can move part of the text (in a more concise way) to the introduction.
Line 93-102: I assume the biodiversity indicator is not a standard output of LPJ-GUESS, correct? It is now part of the LPJ-GUESS section which is misleading. It should be more clear that this is calculated based on the downscaled PLUM results.

Description of the analyzed ecosystem service indicators has been moved to the new Section 2.5. More introductory or background text about the ecosystem services has been moved to the Introduction.

Line 109-110: The Popp et al 2017 article describes a large number of SSP scenarios from 5 IAM models. Please specify which scenario from which model has been used and preferably refer to a paper that presents the results specifically for this model.

We now specify that the bioenergy demand comes from the MESSAGE-GLOBIOM model.

Line 115: this is a very detailed start of this section. Consider restructuring the sentence. Maybe move this entire section to the SI as very detailed information for main text.
Line 137-138: a summary of the climate and land use data used should be given in the text, not in the SI. Consider moving this text to the SI and summarizing input data in the main text.

The relevant text (now Sect. 2.3) has been overhauled, and an overview figure (now Fig. 1) has been added.

Line 161: make sure your language is more academic. Don't use terms like 'briefly' or 'attempts'.

In this matter of style, we disagree with the reviewer. We have left in "attempts" because it is an accurate description of what the algorithm does. "Briefly" primes the reader to expect a non-comprehensive explanation of the algorithm, improving reading flow. That said, most of this text has been expanded and moved to the new Supplementary Methods Sect. SM3, and "briefly" has been removed.

Line 178: what is minimum 'non-agricultural area'. This sounds like a PLUM-specific technical term, please rephrase.

This has been clarified: "These included input and transport costs, tariffs, and minimum non-agricultural area (which places an upper limit on the total fraction of a gridcell that PLUM can allocate to cropland and pasture)."

Line 203: 'cropland area expands about 10% between 2050,. . .'. Between 2050 and what?

This has been corrected to read "between 2050 and 2100".

Line 214: 'what PLUM calls "other management"'. Please use more academic language and avoid usage of model-specific technical terms.

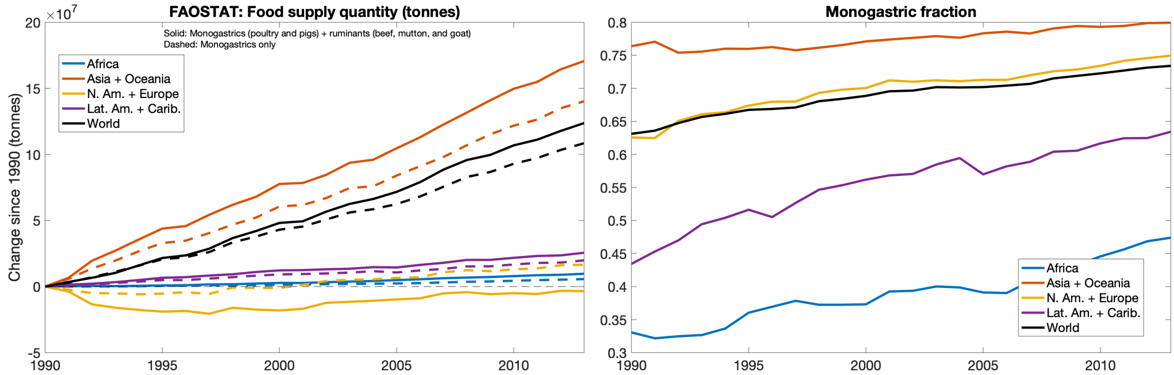
We have changed "what PLUM calls" to "PLUM's". However, we have left the reference to "'other management' intensity," since avoiding model-specific terms harms reproducibility. For readers not already familiar with the term, our explanatory parenthetical text—"(representing, e.g., pesticide application)"—should provide sufficient clarification.

Figure 1: why do you only show ruminants? Non-ruminants also have a very strong effect on the agricultural system due to high feed requirements.

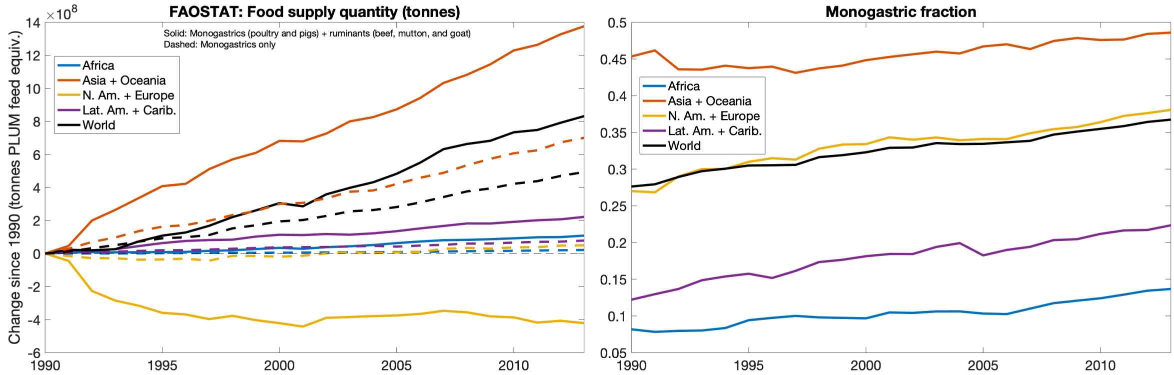
We originally excluded monogastrics from this figure in the interest of limiting its size, since the role of monogastrics demand is not discussed in the text, and interested readers could find its trajectory for each SSP in what is now Fig. SR2 (global demand trajectories for each commodity; formerly Fig. SR1). However, we have now added monogastrics demand to this bar graph (now Fig. 2).

Also, I am surprised about the very strong increase in ruminant demands. FAOSTAT actually shows that in recent decades demand for ruminants-based products has increased relatively little while monogastric-based products have increased much stronger.

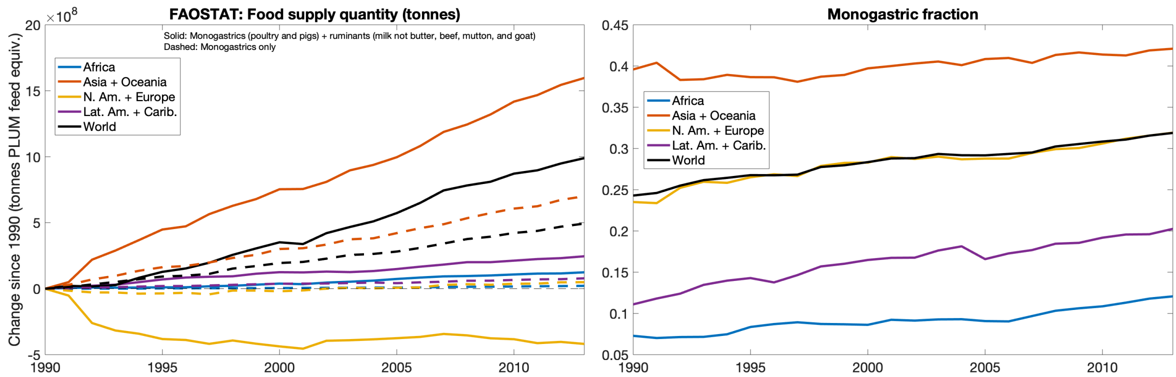
Monogastrics do make up a majority of meat supply by raw tonnage:



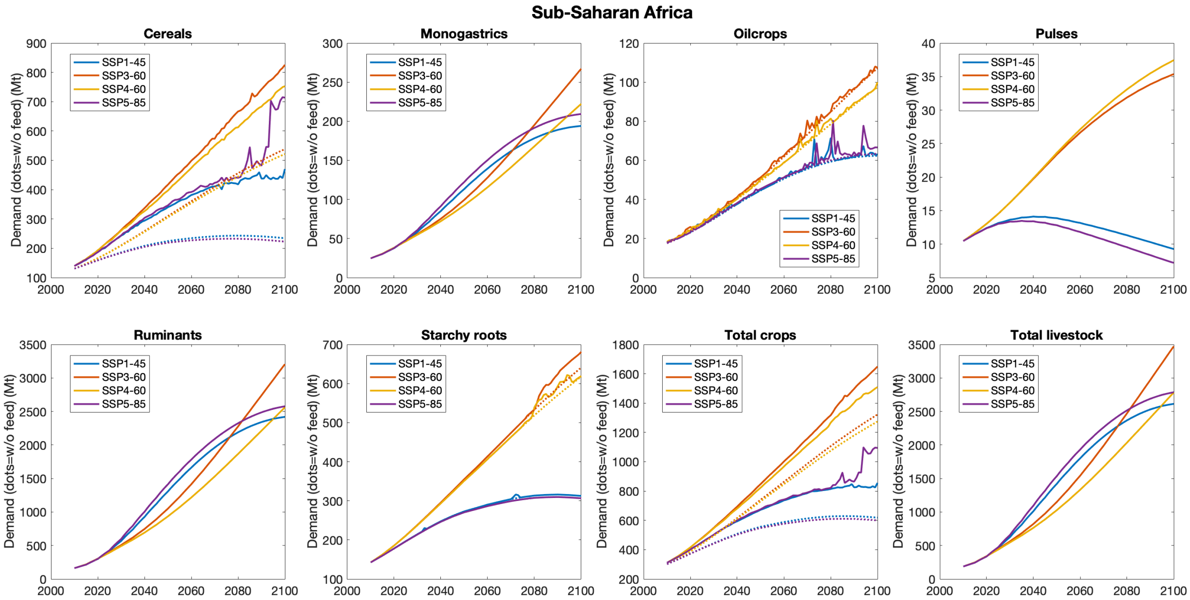
However, they make up a minority of meat supply once converted to the units of PLUM demand, which are tons feed equivalent: Ruminants require much more feed to produce a given weight of meat than do monogastrics. To convert, we multiply poultry meat, pig meat, mutton/goat meat, and beef respectively by 3.3, 6.4, 15.0, and 25.0 tons feed per ton product:

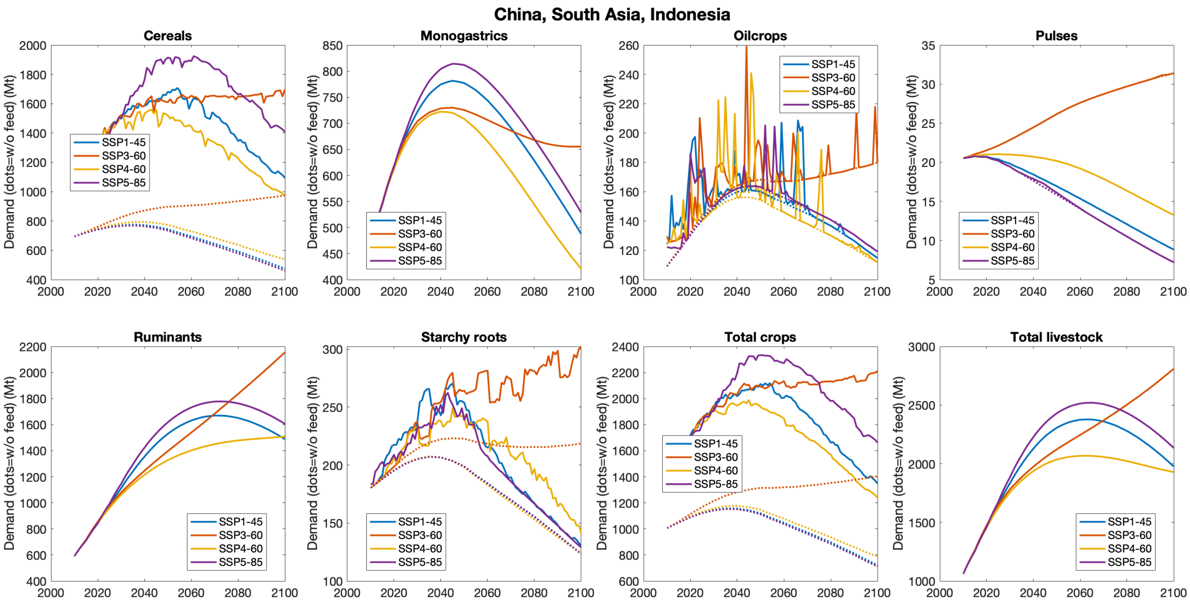


Additionally, ruminant feed requirements in PLUM include the production of milk, the consumption of which has been increasing more rapidly than ruminant meat. This further lowers the fraction of monogastric products in the feed-equivalent figures (0.7 tons feed per ton milk):

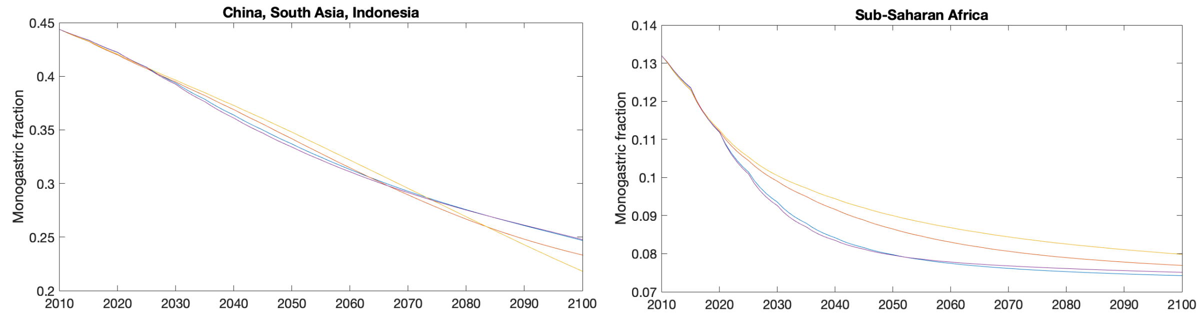


Furthermore, shifts in where demand is growing will change the makeup of the global livestock product landscape. The above figures show that recent trends in livestock meat demand have been driven by increasing consumption in Asia, where meat demand is mostly for monogastrics. In contrast, PLUM projects strong increases in livestock demand in sub-Saharan Africa as population and wealth increase there; that region has historically had a much lower fraction of livestock demand comprised of monogastrics. This is shown in the figures below, which present PLUM-projected demand for Sub-Saharan Africa as compared with combined China, South Asia, and Indonesia. (Region groupings are different between these plots and previous. Also note that colors here stand for scenarios, whereas colors previously stood for regions.)





PLUM also projects that the fraction of livestock products (meat and milk) provided by monogastrics will decrease as these regions get wealthier:



Line 243: I was taught to avoid the term 'forecast' (sounds like something a fortune teller would say) but rather use the term 'project', but maybe this is a personal preference.

This has been changed to "projects".

Line 251-252: this sentence is extremely vague and unhelpful. Please make more explicit.

The last two sentences of this paragraph have been changed to the following: "While we do not expect LandSyMM's results to necessarily match those of other models, such a large, qualitative difference requires explanation. Several factors related to experimental setup and overall model structure likely contribute."

Line 276: it is impossible to have 500% or 700% higher land use, using current agriculture (~50 Mkm2) this would mean 250 or 350 Mkm2 which is more than the terrestrial area of the world.

These values refer to *relative* difference from the Alexander et al. (2018) differences. That is, the range of cropland area among scenarios in the present work is six times what the range was in Alexander et al. (2018). The text has been changed to read (new text in **bold**): "The **spread** in land-use area **projections** between the most extreme scenarios is much higher in this work than…"

Line 282: what do you mean with infrastructure efficiency? Also, if I understand correctly you make similar SSP-specific assumptions in PLUM (SM6). These should also lead to a higher spread in land-use projections, right?

PLUM does make some SSP-specific assumptions, but not the ones described in this sentence. The sentence has been changed to read (new/amended text in **bold**):

> As described above, IMAGE makes a number of assumptions (based on the SSP storylines) **that PLUM does not** regarding future deviations from historical 'business-as-usual' trends and relationships, including dietary shifts, **reductions in food losses during transport**, and forest conservation.

Line 276-278: I don't see why the RCPs lead to a much larger spread in scenario results?

After the cited sentence we have added the following: "The wide variation among the SSPs in population and economic growth trajectories, along with SSP-specific PLUM parameters (Sect. 2.4~~3.2~~), contribute to this increased spread."

Also, is PLUM informed about the yield effects of climate change as these would impact for example trade and food security?

Informing PLUM about the yield effects of climate change (and changing CO_2 concentration) is indeed the reason we feed it with LPJ-GUESS-simulated potential yields. Text has been added throughout the manuscript to emphasize this point.

Line 325: is this realistic? Could not reduced feedback effects from lower evapotranspiration from forests in fact reduce runoff?

Line 336-337: how can increased agriculture reduce the risks on droughts? Please explain.

The end of the paragraph originally ending with Line 326 has been changed to read (new text in **bold**):

> Deforestation in central Africa, for example, is the primary driver of increasing mean annual runoff there **because of reduced evapotranspiration relative to existing vegetation. Note, however, that LandSyMM can only represent the effect of land cover change on evapotranspiration and runoff directly—to include the impact of these flux differences on rainfall would require a coupling with a climate model.**

We have also added the following text in the Methods: “The CMIP5 runs did include land-use change, but not the trajectories output by PLUM. As such, and as with all models that are not climate-coupled but rather use offline forcings, we do not consider the effects of our simulated land-use change on climate.”

Line 348-349: I do not understand what the 'fraction of included land area' means and why it shows that not including routing is not a big issue (it sounds disconcerting to me). Please explain better.

~~In that sentence, “the fraction of included land area in any class” has been changed to “the results for any class”.~~Because LPJ-GUESS results here may be more reliable when aggregated at the basin scale, for simplicity, we have removed the non-aggregated results from the text and Table 2 (formerly Table 1). We have also added a sentence noting this in Sect. 2.5.

Line 359-360: why is the estimate more consistent if there is less over application of N? In reality many countries exceed the N uptake rate of plants (most notably China). It sounds less realistic to me that N application only increases by 2%. Does this not imply a major break with historical trends?

The integrated assessment models whose output was used in Krause et al. (2017) actually simulated these very high levels of nitrogen input as resulting in very high yields—i.e., they did not simulate the real-life phenomenon of overapplication. We do not attempt to account for overapplication, either, but in Alexander et al. (2018) we showed that LandSyMM nevertheless does reproduce historical N application levels globally.

To clarify our point here, we have changed the text to read (edited text in **bold**):

> … used fertilizer information from IMAGE and MAgPIE**. Strong increases in fertilizer in those models resulted in strongly increased yields, but nitrogen limitation is alleviated at much lower levels in LPJ-GUESS. IMAGE and MAgPIE fertilization rates thus** often exceeded what plants in LPJ-GUESS could actually take up, resulting in high amounts of N loss. Coupling LPJ-GUESS with PLUM provides for a more internally consistent estimate of future N losses**, while still reproducing historical fertilizer application well (Alexander et al., 2018)**.

Line 369-370: you cannot state 'and other models' here if you refer to three articles that are if I am not mistaken all based on LPJ-GUESS. It is quite logical then that the estimates are similar. Please add an independent reference.

We have replaced this sentence with the following (edited/new text in **bold**):

> Global combined BVOC emissions over 2001–2010 totaled ~546 TgC yr^{-1} (~503 and ~43 TgC yr^{-1} for isoprene and monoterpenes, respectively), which compares well with estimates from LPJ-GUESS **using different land use scenarios (Arneth et al., 2008; Hantson et al., 2017; Szogs et al., 2017) and the MEGAN model (Sindelarova et al., 2014).**

Line 430: I don't think 'storylines' is the right word here as you don't calculate storylines but scenarios that are described by a certain storyline.

We have replaced "storylines" with "scenarios" there. To avoid repetition, in the previous sentence we replaced "scenarios" with "possible futures".

Line 435-439: would it be possible to draw stronger conclusions based on the scenario assumptions on how certain future developments should rather be avoided etc?

To maintain objectivity, we have decided to avoid such prescriptivist language.

SI:

Figure SR2: why are the starting points in 2010 so different for irrigation and fertilizer? This should be historic data I assume? Also, there is hardly any trends in the irrigation results, why is this the case?

This figure (now Fig. SR3) presents PLUM outputs pre-harmonization; we have added text to this effect in the caption. These raw PLUM outputs are not necessarily expected to align exactly with historical data, as evidenced by the need for the harmonization routine. Nor are they necessarily expected to align with each other at the beginning of the period, because of scenario-specific parameters in PLUM derived from the SSPs—particularly regarding the cost of irrigation and fertilizer. Text to this effect has been added to Sect. 2.2 and the caption of what is now Fig. SR3.

If PLUM were to begin with some "historical" parameter values and gradually phase in scenario-specific values, this would improve agreement among scenarios at the beginning of the future run. However, as there is no obviously "correct" way to design this phase-in, we have made the parsimonious decision to apply all scenario-specific parameters at the beginning.

Regarding irrigation, the following text has been added to the end of the second paragraph of Sect. 3.1:

> PLUM prescribes lower irrigation rates by the end of the century for most scenarios (Figs. 2, SR3~~2~~). This is enabled by higher global mean rainfall in all RCP scenarios, as evidenced by the bars for runoff in Figure 4, as well as by improved water-use efficiency for crops other than C_4 cereals due to increased CO_2 concentrations. Crop demand increase in SSP3-60 outweighs these effects, however, resulting in higher irrigation in that scenario.

Figure SR3: I don't understand the top figure on livestock demand. The order of magnitude makes it likely the results are on crop production, but the title suggests this is total *production (?) of livestock products. But the bottom figure shows feed for livestock so if the top figure is also about feed demand for livestock it would not be useful. Please explain and improve.*

The former Fig. SR3 (now Fig. SR4) has been simplified to show only the fraction of ruminant food that is provided by feed crops, which is the only information required to support the corresponding assertion in the main text (i.e., that feed becomes much more important in raising ruminant livestock beginning around 2090).

Figure SR8: similar figures are shown for south asia and sub-saharan Africa. Why not consistently show results for all regions? Maybe in slight smaller panels and without the figures with maps in between? Also, why is this figure shown as a delta instead of absolute amounts and how can it be that the demand is so extremely jumpy for oil crops? This seems very unrealistic.

There are an enormous number of possible regions and region groupings that could have such figures made; we present only what is necessary to support specific assertions made in the main text. Similarly, we have presented percentage change rather than absolute amount because the former is more directly explanatory of assertions in the main text. Showing absolute amount would, in some cases, make relevant trends difficult to discern.

The "jumpiness" of individual crops is due to shifts in which crops are used as animal feed. These shifts are due primarily to changes in relative prices of the different crop commodities. Note that the dotted lines, which exclude demand for animal feed, are much more stable. It is indeed unrealistic to expect, e.g., oilcrop production to triple from one year to the next, as would be required to satisfy the demand increase seen in the US and Canada in the early 2040s (Fig. SR7~~9~~, formerly Fig. SR8). For the purposes of our ecosystem services analysis, however, gross decadal trends in total agricultural area and management inputs are much more important than exactly what is being grown on cropland, and those gross trends are much smoother. Text explaining this has been added to Sect. 2.2~~.~~:

> The composition of livestock feed (in terms of which crops are used) is assumed to be flexible, which can result in large interannual fluctuations in demand and production of individual crops as their prices change relative to one another. This is seen, for example, in Supplementary Results Fig. SR7, where oilcrop demand in the US and Canada triples from one year to the next. This assumption is not expected to materially affect the results in terms of gross decadal trends in total agricultural area and management inputs.

Figure SR10: please write complete description instead of referring to another figure.

Complete descriptions have been added to the former Figs. SR 5~~–7~~, 10, and 13 (now SR 6~~–8~~, 9~~11~~, and 11~~14~~). The former Figs. 6 and 7 have been removed as they are not (or are no longer) referenced anywhere.

Reply to Reviewer 2

The main contribution of this paper is in coupling the PLUM and LPJ-GUESS models to project land-use change impacts in future scenarios, in terms of biodiversity impacts or greenhouse gas emissions.

As such, my main criticism of the paper is that the method section is not very detailed about (1) the assumptions of the two models, (2) the working of the two models, and most crucially (3) how these were combined. While I appreciate the difficulty of communicating complex models in a brief section, seeing that this is the central contribution of the paper, the reader should not be forced to go through the supplemental materials (which is also very densely presented) to understand the models and their interplay. This could potentially be presented as multiple tables and a joint figure exploring the interactions and basic properties of the models.

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.

Conversely, I would suggest to shift large parts of the input data sections to the SM, as (especially in the case of PLUM), these are largely technical details on the modeling side. Instead, the manuscript should spend more time in detailing the scenario setup, as well as how the "holding constant of certain variables" for the purpose of robustness checking was implemented, as based on the abstract and introduction this is a central part of the paper.

Technical details are significantly less prominent in the new Sect. 2.3. The new Table 1, which describes the experimental runs, should clarify what it means for certain variables to be "held constant."

Minor comments

The error bars in Fig. 1 and Fig. 3 are largely cosmetic, as the processes depicted here are highly persistent (e.g. population, cropland), and the error bars merely measure the standard deviations within a decade. The authors themselves do not interpret them within the text, so *for the clarity of information they could be also left off. Indeed, if the authors would like to highlight the temporal dynamics, a representation of the whole time-series would be better suited.*

The error bars in these figures (now Figs. 2 and 4) have been removed.

In Fig. 2 it would be good to either have a different color scheme for the two columns or the same scale.

This figure (now Fig. 3) has been updated. Among other visual changes, the two columns are now on the same scale.

In the SM, figures for commodities and exports are presented (SR8, 10,11, 13). Here the trade patterns exhibit highly cyclical behavior, which right not be fully realistic. This should be contrasted with past export dynamics in the same crops and regions.

The "jumpiness" of individual crops is due to shifts in which crops are used as animal feed. These shifts are due primarily to changes in relative prices of the different crop commodities. Note that the dotted lines, which exclude demand for animal feed, are much more stable. It is indeed unrealistic to expect, e.g., oilcrop production to triple from one year to the next, as would be required to satisfy the demand increase seen in the US and Canada in the early 2040s (former Fig. SR8, now SR7~~9~~). For the purposes of our ecosystem services analysis, however, gross decadal trends in total agricultural area and management inputs are much more important than exactly what is being grown on cropland, and those gross trends are much smoother. Text explaining this has been added to Sect. 2.2 (PLUM).

Reply to comment by 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~~ text 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."~~

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.

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~~ text 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."~~

> 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.

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 (Fig. SR1). 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, compared to 18 other models examined in Ahlström et al. (2012, their Fig. S2), IPSL-CM5A-MR in RCP8.5 simulates a much larger precipitation increase around the Equator, where we see the largest increase in runoff (Fig. SR14a).~~—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: "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."~~.~~

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.

Reply to comment by Sandy Harrison

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 (Fig. 1), as is a table describing the various experimental runs (Table 1).

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~~ We have added 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, as well as how N limitation and irrigation work, to Sect. 2.1 (new text in **bold**):~~.~~

> Nitrogen limitation on plant growth is modeled, with cropland able to receive fertilizer applications (Smith et al., 2014; Olin et al., 2015b). The mechanistic representation of wild plant and crop growth accounts for the CO_2 fertilization effect, by which productivity can be enhanced due to improved water use efficiency and (in C_3 plants) reduced photorespiration (Smith et al., 2014). **In an intercomparison of eight vegetation models over 1981–2000 (Ito et al., 2017), LPJ-GUESS simulated a mean global gross primary productivity very close to the ensemble average, although with the second-steepest increasing trend. LPJ-GUESS also has been shown to realistically simulate the effects of fertilizer application and elevated CO_2 on temperate cereal yield (Olin et al., 2015b), although the latter effect is stronger than in other crop models (Pugh et al., in prep.).** Changes to irrigation, water demand, water supply, and plant water stress as described in the Supplementary Information of Alexander et al. (2018) were included. **Most importantly, these changes include (a) increasing maximum irrigation to allow it to bring soil to moisture levels well above the wilting point, and (b) a factor reflecting how soil moisture extraction gets more difficult as the soil gets drier.**

~~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 new 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~~.~~ (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.**

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 (Fig. 1).

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.~~ We have also added a bullet point discussing China:

> China's crop demand peaks by about 2040; by the end of the century, it has either returned to (SSP3-60) or dropped past 2010 levels (by 30%, 40%, and 25% for SSP1-45, SSP4-60, and SSP5-85, respectively; Fig. SR12). Crop imports decrease from 14% of demand to less than 6%. This fits well with apparent net losses of cropland area in all scenarios, but note that harmonization switched SSP1-45's projection from an 8.5% gain to a 15% loss. Moreover, whereas PLUM projected cropland abandonment to occur in the montane shrublands and steppe of the Tibetan Plateau, after harmonization it occurs throughout the eastern temperate and subtropical forests. Slight cropland expansion projected by PLUM in China's subtropical moist forests is increased 300–600% by harmonization in all scenarios except SSP1-45 (+21%).

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~~We have added 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 to Sect. 2.1.
- *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.~~ Detailed comparison of the original PLUM outputs with the harmonized time series reveals that harmonization increases the total amount of land undergoing land-use change. While this complicates interpretation to some extent, sometimes necessitating clarification on whether a change was present in the PLUM outputs or not, overall our results are not greatly affected. We have also added text to the Results as necessary noting where apparent strong regional effects of land-use change result from changes that were not present pre-harmonization. This phenomenon is explored in new discussion (plus a new table and figure) in Supplementary Methods Sect. SM3.

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~~ second paragraph of Sect. 2.2: "~~PLUM assumes that~~ [irrigation and fertilizer] are assumed to 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.~~ We have revised part of the first paragraph of Sect. 3.2.2 to read, "While the impacts of increasing CO_2 levels on runoff can be strongly regionally dependent (Zhu et al., 2012), we see that overall more CO_2 means less runoff at a global level."

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 comments 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.~~ We have added the following text to the Conclusions: "However, various elements of uncertainty—related to PLUM parameter values, global climate model selection, and model design—affect these results and remain to be explored."

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 that~~:~~ "~~LPJ-GUESS calculates runoff~~ [runoff] is calculated 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).~~." Additionally, text has been added to that paragraph explaining that flow between gridcells is not modeled.

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.”~~“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.”

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 (Sect. SM2). 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 (now Fig. 3) 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. Similar color changes have been made to the former Fig. 4 (now Fig. 5).

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

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

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 …"

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

This clause has been changed to "PLUM expects Sub-Saharan Africa to see~~s~~ 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…”

Impacts of future agricultural change on ecosystem service indicators

Sam S. Rabin[1], Peter Alexander[2,3], Roslyn Henry[2], Peter Anthoni[1], Thomas A. M. Pugh[4,5], Mark Rounsevell[1], and Almut Arneth[1]

[1]Institute of Meteorology and Climate Research / Atmospheric Environmental Research, Karlsruhe Institute of Technology, Germany
[2]School of Geosciences, University of Edinburgh, UK
[3]Global Academy of Agriculture and Food Security, The Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK
[4]School of Geography, Earth and Environmental Sciences, University of Birmingham, UK
[5]Birmingham Institute of Forest Research, University of Birmingham, UK

Correspondence: Sam S. Rabin (sam.rabin@kit.edu)

Abstract. A future of increasing atmospheric carbon dioxide concentrations, changing climate, growing human populations, and shifting socioeconomic conditions means that the global agricultural system will need to adapt in order to feed the world. These changes will affect not only agricultural land, but terrestrial ecosystems in general. Here, we use the coupled land use and vegetation model LandSyMM to quantify future land use change and resulting impacts on ecosystem service indicators ~~including~~ relating to carbon sequestration, runoff, biodiversity, and nitrogen pollution. We additionally hold certain variables, such as climate or land use, constant to assess the relative contribution of different drivers to the projected impacts. While indicators of some ecosystem services (e.g., ~~flood and drought risk~~extreme surface water flow levels) see trends that are mostly dominated by the direct effects of climate change, others (e.g., carbon sequestration) depend critically on land use and management. Scenarios in which climate change mitigation is more difficult (Shared Socioeconomic Pathways 3 and 5) have the strongest impacts on ecosystem service indicators, such as a loss of 13–19% of land in biodiversity hotspots and a 28% increase in nitrogen pollution. Evaluating a suite of ecosystem service indicators across scenarios enables the identification of tradeoffs and co-benefits associated with different climate change mitigation and adaptation strategies and socioeconomic developments.

1 Introduction

Exploring how the agricultural system might shift under different plausible future climate and socioeconomic changes is critically important for understanding how the future world—with a population ~~in~~ increase by 2100 ~~larger than today's by anywhere~~ ranging from 1.5 billion to nearly 6 billion people (KC and Lutz, 2017)—will be fed. ~~However, the implications of changing agricultural land area~~ In addition, land-based mitigation—reducing deforestation, increasing sequestration in natural and agricultural lands, and expanding biofuel use—might be an important piece in the strategy to achieve warming targets laid out in the Paris Agreement (Rogelj et al., 2018; van Vuuren et al., 2018). The implications of resultant shifts in land use areas and management inputs go far beyond food security. ~~The conversion of forests and other ecosystems to croplands or~~

~~pasture—combined with the inputs required to produce food on those lands—emits large amounts of greenhouse gases, reduces the ability of the natural world to buffer against anthropogenic carbon dioxide emissions, results in pollution of important freshwater and marine ecosystems, harms biodiversity, and affects water availability and flood risk. Such impacts are not merely academic, because they affect the livelihoods of human societies which depend~~ Human society depends on a wide range of ecosystem services which broadly fall into three categories (IPBES, 2018a): regulating (e.g., greenhouse gas sequestration, flood control), material (e.g., food and feed production), and non-material (e.g., learning and inspiration). These have all historically been strongly affected by land use and management.

Declining biodiversity due to the loss and degradation of habitat (Jantz et al., 2015; Newbold et al., 2015) raises moral and ethical questions regarding extinction, represents a loss of a non-material ecosystem service *per se*, and indirectly harms other ecosystem services by impairing system function (Simpson et al., 1996; Tilman et al., 2014; IPBES, 2018b). The conversion of forests and other ecosystems to croplands or pasture has also, by releasing carbon from vegetation and soil pools, caused about a third of humanity's CO_2 emissions since 1750 (Ciais et al., 2013). Land-use change also alters how vegetation intercepts rainfall and takes up water from the soil, affecting the amount and timing of runoff and thus water supply and flood risk (Wheater and Evans, 2009; Haddeland et al., 2014). This affects both human and natural systems, as do changes in runoff quality: Nitrogen (N) compounds from fertilizer dissolve in soil water and are transported from agricultural land to freshwater and marine ecosystems. There, this nitrogen pollution can cause eutrophication and affect various ecosystem services, including fishery production (Vitousek et al., 1997). Fertilizer also produces air pollution in the form of nitric oxides (which contribute to respiratory illnesses; Yang and Omaye, 2009) and the greenhouse gas nitrous oxide (the third-largest component of anthropogenic climate change; Fowler et al., 2009; Myhre et al., 2013; Shcherbak et al., 2014). Where nitrogen oxides are elevated, they can react with biogenic volatile organic compounds (BVOCs), which are emitted by plants—especially woody species (Rosenkranz et al., 2014)—for a variety of physiological functions. These reactions produce tropospheric ozone, which is harmful to human health (Ebi and McGregor, 2008), can negatively affect photosynthesis (Ashmore, 2005), and is a greenhouse gas (Myhre et al., 2013). BVOCs also have other, more complicated implications for regulating and material ecosystem services. They can warm the planet by increasing methane lifetime (Young et al., 2009), but on the other hand they help form tropospheric aerosols, which increase reflectance and boost photosynthesis via diffuse radiation (Rap et al., 2018; Sporre et al., 2019). The latter can improve crop yields, but BVOC-enhanced ozone formation can work against that effect (Feng and Kobayashi, 2009).

As global environmental and societal changes continue ~~to accelerate~~ over the coming decades, it is critical that we understand not just the impacts on the natural world, but how those impacts feed back onto humanity.

To explore the possible future evolution of the Earth system and society, models have been developed that simulate the global economy, natural world, and their interactions. A substantial body of research has been built up using such models to examine how future land-use change will affect individual ecosystem services such as carbon sequestration (Brovkin et al., 2013; Lawrence et al., 2018), biodiversity (Jantz et al., 2015; Hof et al., 2018; Di Marco et al., 2019), and water availability and flood risk (Davie et al., 2013; Elliott et al., 2014; Asadieh and Krakauer, 2017). Much less work has been undertaken to evaluate the future of a suite of ecosystem services in an integrated way (Krause et al., 2017; Molotoks et al., 2018). However, such analyses provide critically important evidence for balancing the many competing demands on the land system while achieving climate and societal targets such as those laid out in the Paris Agreement and Sustainable Development goals (Eitelberg et al., 2016; Benton et al., 2018; Verhagen et al., 2018).

Previously, we used the PLUM land-use model to estimate future land use and management change, based on changing socioeconomic conditions as well as climate effects on agricultural yield provided by the vegetation model LPJ-GUESS (Alexander et al., 2018). This coupled model system—the Land System Modular Model, or LandSyMM—is ~~unique among~~ among the state of the art in global land-use change models ~~in~~ due to the high level of ~~spatial~~ detail that it considers in the response of agricultural yields to management inputs~~, as well as in allowing~~. 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 relative to demand (rather than assuming market equilibrium in every year). Here, we take advantage of the mechanistic modeling of terrestrial ecosystems provided by LPJ-GUESS to explore how PLUM-generated future land use and management trajectories—under different scenarios of future socioeconomic development and climate change—differ in their consequences for a range of regulating and material ecosystem services.

2 Methods

2.1 LPJ-GUESS~~and ecosystem services~~

The Lund-Potsdam-Jena General Ecosystem Simulator (LPJ-GUESS) is a dynamic global vegetation model that simulates—here, at a spatial resolution of 0.5 degrees—physiological, demographic, and disturbance processes ~~among~~ for a variety of plant functional types (PFTs) on natural land (Smith et al., 2001, 2014). Hydrological and most physiological processes are modeled at daily temporal resolution; vegetation growth, establishment, disturbance (including land-use change), and mortality happen annually. Agricultural land is also included, with cropland and pasture being restricted in the types of plants allowed and experiencing annual harvest. Transitions among land use types are given as an input, with LPJ-GUESS calculating the associated change in carbon pools and fluxes (Lindeskog et al., 2013). Four crop functional types (CFTs) are represented: C~~$_3$~~$_3$ cereals sown in winter~~and~~, C_3 cereals sown in spring, C~~$_4$~~$_4$ cereals, and rice (Olin et al., 2015a). Nitrogen limitation on plant growth is modeled, with cropland able to receive fertilizer applications (Smith et al., 2014; Olin et al., 2015b). The mechanistic representation of wild plant and crop growth accounts for the CO_2 fertilization effect, by which productivity can be enhanced due to improved water use efficiency and (in C~~$_3$~~$_3$ plants) reduced photorespiration (Smith et al., 2014). In an intercomparison of eight vegetation models over 1981–2000 (Ito et al., 2017), LPJ-GUESS simulated a mean global gross primary productivity very close to the ensemble average, although with the second-steepest increasing trend. LPJ-GUESS has also been shown to realistically simulate the effects of elevated CO_2 on temperate cereal yield (Olin et al., 2015b), although the latter effect is stronger than in other crop models (Pugh et al., in prep.). Changes to irrigation, water demand, water supply, and plant water stress as described in the Supplementary Information of Alexander et al. (2018) were included. Most importantly, these changes include (a) increasing maximum irrigation to allow it to bring soil to moisture levels well above the wilting point, and (b) a factor reflecting how soil moisture extraction gets more difficult as the soil gets drier.

LPJ-GUESS ~~simulates indicators of a number of provisioning and regulating ecosystem services (see also Table 1 in). Along with providing some background information, the rest of this section will detail the relevant LPJ-GUESS outputs for the ecosystem service indicators analyzed in this paper having to do with carbon sequestration, water supply, nitrogen pollution, biogenic volatile organic compounds (BVOCs), and biodiversity.~~

~~The conversion of forests and other ecosystems to croplands or pasture has been responsible for about athird of humanity's CO_2 emissions since 1750. Land clearance for agriculture affects carbon storage by emitting vegetation and soil carbon from existing pools, as well as by reducing sequestration potential . Land-based mitigation strategies—including reducing deforestation, increasing sequestration in natural and agricultural lands, and expanding biofuels' contribution to energy supply—could play a critical role in whether warming targets laid out in the Paris Agreement can actually be achieved . The change in total carbon stored in the land system is used here as a measure of the carbon sequestration performed by terrestrial ecosystems.~~

~~Land use and management change can also impact how ecosystems regulate water quantity, quality, and flood risk . Land use change affects runoff by changing how vegetation intercepts rainfall and takes up water from the soil. We consider average annual runoff as it contributes to water levels in lakes and rivers, which is important for not only freshwater ecosystems but also water availability for irrigation and other human uses . After , we also use the difference between 1971–2000 and 2071–2100 in the 95th percentile of monthly surface runoff ($P95_{month}$) as a proxy for changing flood risk (although note that those authors used daily values), and the difference in the 5th annual percentile ($P5_{year}$) for changing drought risk. Note that we are referring to hydrologic drought, which can be contrasted with, e.g., meteorological or socioeconomic drought .~~

~~Crop fertilization is the main anthropogenic source of nitrous oxide (N_2O), a potent greenhouse gas responsible for the third-largest contribution of anthropogenic climate change . Some is also emitted as nitric oxides (NO_x), which contribute to respiratory illnesses via surface-level air pollution . Dissolved nitrogen compounds leach from agricultural land into freshwater and marine ecosystems, where they can contribute to their degradation via eutrophication and affect various ecosystem services , including fishery production . Here, we examine total nitrogen loss from terrestrial ecosystems as output by LPJ-GUESS.~~

~~Ecosystem services related to climate change and human health can be strongly affected by biogenic volatile organic compounds (BVOCs), which are emitted by plants—especially woody species —for a variety of physiological functions. In regions where nitrogen oxides are elevated, their reactions with BVOCs produce tropospheric ozone, which is harmful to human health , can negatively affect photosynthesis , and is a greenhouse gas . BVOCs also warm the planet by increasing methane lifetime , but on the other hand they help form tropospheric aerosols, which increase reflectance and boost photosynthesis via diffuse radiation . The latter can improve crop yields, but BVOC-enhanced ozone formation can work against that effect . LPJ-GUESS simulates the emission of isoprene and monoterpenes—the most prevalent BVOCs in the atmosphere —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 .~~

~~Biodiversity underpins a wide range of ecosystem services due to its importance for ecosystem functioning and the possible use of wild species in improving crops and developing new medicines . Land clearance has also been—and will continue to~~

~~be—responsible for declining biodiversity due to the loss and degradation of habitat . In addition to raising moral and ethical questions regarding extinction, this negatively impacts ecosystem services and thus people's livelihoods . Here we assess how much land is converted to agriculture within the Conservation International (CI) hotspots, a set of 35 regions covering less than 3of the Earth's land area but containing half the world's endemic plant species and over 40of the world's endemic vertebrate animal species . These regions each contain at least 1500 endemic vascular plant species and have already lost at least 70of their original natural vegetation, thus representing highly diverse areas presently at high risk of habitat loss~~simulates variables that can be used as indicators of a number of provisioning and regulating ecosystem services (see also Table 1 in Krause et al. (2017)); these are described in Section 2.5.

2.2 PLUM

The Parsimonious Land Use Model (PLUM) is designed to produce trajectories of land use and management based on socioeconomic trends and gridcell-level crop and pasture productivity at a resolution of 0.5 degrees (Engström et al., 2016b; Alexander et al., 2018). Food demand is projected into the future based on ~~the Shared Socioeconomic Pathway (SSP)scenarios~~ external scenario projections of country-level population and gross domestic product (GDP), using the historical relationship of per capita GDP to consumption of each of six crop types—C_3 cereals, C_4 cereals, rice, oilcrops, pulses, and starchy roots—plus ruminant and monogastric livestock (FAOSTAT, 2018a, b). Demand of a seventh crop type—dedicated bioenergy crops such as ~~Miscanthus—is specified according to the SSP2 scenariodescribed by ; demand for bioenergy from food crops is specified to double from 2010 by 2030 and thereafter remain constant.~~*Miscanthus*—is specified based on an exogenous scenario. PLUM calculates the demand for food crops both for human consumption and feed for monogastric livestock, plus any ruminants not raised on pasture.)

Demand is satisfied at the country level by either domestic production or imports, the balance between which is determined ~~using a least-cost optimization~~ considering commodity prices~~as well as the cost of~~ , management costs (fertilizer, irrigation~~water, and land conversion. A more detailed explanation of PLUM can be found in .~~ , land conversion, and "other management" such as pesticide use), and changing LPJ-GUESS-simulated productivity due to climate change and CO_2 under a range of irrigation-fertilization treatments. The latter are assumed to produce diminishing returns, such that increasing them increases yield at low intensity levels, but less and less so at higher levels, approaching a yield asymptote.

2.3 ~~Simulation details~~

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.

~~After a spinup of 500 years to bring vegetation and soil pools to a realistic starting point, a historical run spanning 1850 to 2005 was performed in LPJ-GUESS. The model state at the end of this run was usedas the starting point for runs generating potential yields under different Representative Concentration Pathway (RCP)scenarios of climate and atmospheric~~ ~~CO_2~~ ~~. In each scenario, every grid cell is~~ The composition of livestock feed (in terms of which crops are used) is assumed to be flexible, which can result in large interannual fluctuations in demand and production of individual crops as their prices change relative to one another. This is seen, for example, in Supplementary Results Fig. SR7, where oilcrop demand in the US and Canada triples from one year to the next. This assumption is not expected to materially affect the results in terms of gross decadal trends in total agricultural area and management inputs.

As outputs (feeding into LPJ-GUESS for use in LandSyMM), PLUM produces half-degree gridded maps of land use area (cropland, pasture, and non-agricultural land), crop distribution (fraction of cropland planted with each crop type~~, each of which is given~~), irrigation intensity, and nitrogen fertilizer application rate. 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.

2.3 LandSyMM: Combining LPJ-GUESS and PLUM

This section and Figure 1 provide an overview of how LPJ-GUESS and PLUM are combined in the LandSyMM runs presented in this work. More details on this coupling can be found in the Supplementary Methods.

The first step in running LandSyMM is to perform "yield-generating" runs in LPJ-GUESS. A simulation of the historical period generates a model state, which is needed so that vegetation and soil condition can be fed into subsequent runs (Fig. 1). From that state, we perform a series of runs that generate "potential yields" in every gridcell for each crop under six different management treatments in a factorial setup: fertilization of 0, 200, and 1000 kgN ha^{-1} and either no irrigation or maximum irrigation. ~~PLUM then uses these potential yield estimates to generate time series of land use and management that satisfy crop and livestock demand calculated based on socioeconomic projections from the SSPs. Each socioeconomic scenario (SSP) is paired with a climate scenario (RCP) based on what sort of climate change could be expected under each SSP's storyline: SSP1 uses RCP4.5, SSPs 3 and 4 use RCP6.0, and SSP5 uses RCP8.5. More details on the setup of the calibration and yield-generating runs can be found in the Supplementary Methods (Sect~~Changing pasture productivity is accounted for using annual average net primary productivity; for simplicity, we include pasture when using the phrase "potential yields." These potential yields account for changing productivity given changing climate and atmospheric CO_2 concentration.

PLUM then combines the future potential yields from LPJ-GUESS (averaged over five-year timesteps) with its own estimates of future commodity demand to project land use areas, fertilizer application, and irrigation intensity (Fig. ~~SM1).~~

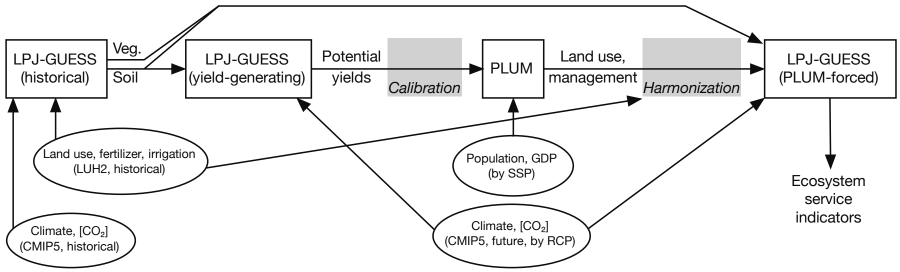


Figure 1. LandSyMM structural overview. Ovals represent external input data and white rectangles represent model runs, with arrows indicating data flow from one model run to the next. Gray rectangles represent model coupling processes, whose external inputs have been excluded for simplicity; more information on these can be found in the Supplementary Methods.

1). PLUM has been found to perform well in this coupled system; its recreation of historical patterns and projections into the future are discussed in Alexander et al. (2018). Here, ~~after extending the historical run to 2010, we feed~~ PLUM's ~~projections of 2011–2100 land use , nitrogen fertilizer application, and irrigation intensity back into~~ demand estimates are driven by scenario-specific population and GDP data (Sect. 2.4).

The outputs of land use and management from PLUM for a given 2011–2100 scenario are fed into a final LPJ-GUESS ~~to simulate their impacts over that period. We refer to these experiments as the "PLUM-forced" runs. The Supplementary Methods include a figure illustrating the overall workflow~~ run in order to produce projections of the ecosystem service indicators analyzed here (Fig. ~~SM1)~~1). However, the PLUM outputs must be processed first, because at the beginning of the future period they do not exactly match the land use and management forcings used at the end of the historical period. Feeding the raw PLUM outputs directly into LPJ-GUESS—causing large areas of sudden agricultural abandonment and expansion between 2010 and ~~table summarizing a summary of the run types~~ 2011—would thus complicate interpretation of the results, especially of carbon cycling. We developed a harmonization routine, based on that published for LUH1 (Hurtt et al., 2011, http://luh.umd.edu/code.shtml), that adjusts the PLUM outputs to ensure a smooth transition from the historical period to the future. While global totals are conserved in almost all cases, harmonization can produce notable differences at the regional scale. Details on this routine can be found in ~~Table SM1~~the Supplementary Methods (Sect. SM3).

In addition to ~~PLUM-forced runs with time-varying future climate and atmospheric CO_2~~the LPJ-GUESS runs forced with harmonized PLUM-output land use and management trajectories, we perform several experiments to examine the impact of different factors on the land-use and -management projections generated by PLUM and thus the ecosystem service indicators simulated by LPJ-GUESS in the PLUM-forced runs. 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, we can estimate the contribution of each to changing ecosystem service indicators in the future. ~~The indirect contributions of changing climate and atmospheric CO_2—i.e., how they affect the land use and management pathways chosen by PLUM—are not considered.~~

2.3.1 ~~Input data: LPJ-GUESS~~

~~Here, we describe the climate and land use data used in LPJ-GUESS. A summary~~ Details regarding the inputs of these experimental runs can be found in Sect. 2.4 and the Supplementary Methods~~, Tables SM2–SM5.~~ ~~The calibration run was forced with climate data from CRU-NCEP version 7 , but with CRU TS3.24 precipitation due to problems discovered in the CRU-NCEP precipitation data. All other runs used the atmospheric CO~~. The results of these experimental runs were used to inform interpretation of the results but are mostly not presented here. Instead, the Supplementary Results file contains figures (numbers prefixed with SR) supporting claims derived from the experimental runs, in addition to other figures that were not included here to conserve space. Runs are referred to using the naming convention described in Table 1. Note that all PLUM outputs consider LPJ-GUESS yields under changing climate and CO_2 ~~concentrations and climate forcings from the Fifth Coupled Model Intercomparison Project : specifically, the IPSL-CM5A-MR forcings , which were bias-corrected to the 1961–1990 observation-based climate used by the calibration runs. Because not all SSP-RCP combinations are equally plausible, the PLUM-forced runs used future climate forcings corresponding to the most likely RCP for each SSP, based on the SSP-RCP probability matrix from : i.e., RCP4.5 for SSP1, RCP6.0 for SSP3 and SSP4, and RCP8.5 for SSP5. Three instances of the yield-generating runs were performed: one with each of the RCP climate scenarios.~~ ~~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 provided crop type distributions for the year 2000, which were used for all historical years. Some mapping between MIRCA,~~ concentration, even when those outputs are fed into LPJ-GUESS ~~, and PLUM crop types was required, details of which can be found in the Supplementary Methods (Sect.SM2) . Fertilizer application for the calibration run was taken from the dataset prepared for the Global Gridded Crop Model Intercomparison exercise of the Agricultural Model Intercomparison Project . In the historical period of other runs, nitrogen fertilizer application rates were taken from LUH2 (Hurtt et al., in prep). Manure N was added in the historical period according to the annually-varying maps given in held constant at year 2000 levels in the calibration run to match the use of the AgMIP fertilizer data. Simulation years outside the dataset's 1860–2014 range used 1860~~ runs with constant climate and/or CO_2. Our analyses thus account only for the direct effects of changing climate and CO_2 on ecosystem service indicators, rather than their indirect effects via land use and ~~2014 values, respectively.~~ management.

~~The PLUM outputs at the beginning of the future period do not exactly match the land use and management forcings used at the end of the historical period. Feeding the PLUM outputs directly into LPJ-GUESS—causing large areas of agricultural abandonment and expansion between 2010 and2011—would thus complicate interpretation of the results, especially of carbon cycling. We developed a harmonization routine based on that published for LUH1 . Briefly, the code begins with land use from LUH2 in 2010, then attempts to apply changes in land use area between PLUM's 2010 and 2011 outputs. In gridcells with no space to apply those changes, the algorithm attempts to find space in neighboring gridcells, expanding the search radius until the changes are satisfied. This process is then repeated, with the PLUM changes for 2011–2012 being applied to the~~

Table 1. Naming convention for LandSyMM runs analyzed in this work, based on land use and management (LU, mgmt.), climate, and CO_2 inputs. **Bold** indicates factors held constant in experimental runs. `X` refers to one of the SSPs (1, 3, 4, 5); `YY` refers to one of the RCPs (4.5, 6.0, 8.5). Unless otherwise specified, land use forcings are harmonized outputs from PLUM run fed with RCPY.Y-forced LPJ-GUESS potential yields and SSPX socioeconomic data and assumptions, and climate and CO_2 forcings are from RCPY.Y.

Experiment name	LU, mgmt.	Climate	CO_2
`sXlum_rYYclico2` (all varying)	2011–2100	2011–2100	2011–2100
`rYYclico2` (constant LU/mgmt.)	**2010** [a]	2011–2100	2011–2100
`sXlum_rYYco2` (constant climate)	2011–2100	**1981–2010** [b]	2011–2100
`sXlum_rYYcli` (constant CO_2)	2011–2100	2011–2100	**2010** [c]
`sXlum` (constant climate and CO_2)	2011–2100	**1981–2010** [b]	**2010** [c]
`rYYco2` (constant LU/mgmt. and climate)	**2010** [a]	**1981–2010** [b]	2011–2100
`rYYcli` (constant LU/mgmt. and CO_2)	**2010** [a]	2011–2100	**2010** [c]

a: From LUH2 (Hurtt et al., in prep.) and Zhang et al. (2017).

b: Historical (not RCP) climate with temperature detrended. These 30 years are repeated throughout the future period: 2011 uses 1981 climate, 2012 uses 1982 climate, etc.

c: Approximately 389 ppm.

~~harmonized 2011 land use map, and so on.More details on the harmonization procedure can be found in the Readme of the harmonization code (see Code Availability) .~~

2.3.1 ~~Input data: PLUM~~

2.4 Input data and scenarios

~~In addition to the potential yields generated by LPJ-GUESS, PLUM considers a number of socioeconomic factors in estimating both future demand for commodities as well as how best to satisfy that demand. The~~ The experiments treated here are based around combined future climate-socioeconomic scenarios. Future population growth and economic development are derived from the Shared Socioeconomic Pathways ~~(SSPs) provide a framework for generating scenarios of those socioeconomic factors' future evolution~~ (SSPs; O'Neill et al., 2014; IIASA, 2014). We use four of the five SSPs, which together cover a wide spectrum of possible storylines for the future evolution of the climate and society (O'Neill et al., 2014, 2015)~~. Four of the five SSPs are used here~~. SSP1 characterizes a world shifting to a more sustainable pathway, with low population growth and strong technological and economic developments. SSP3 describes a pathway with strong population growth and intensive resource usage, low technological development, and lessening globalization. SSP4 is a pathway of inequality with the potential for competition over resources and resource intensification. SSP5 is a pathway dependent on fossil fuels with low population growth, strong globalization, and high economic and technological growth~~with strong globalization~~. (SSP2, a "middle of the road" pathway intermediate between the other four SSPs, is not considered here.) ~~Scenario data on population and GDP per country through 2100 were~~

Scenarios of future climate change and atmospheric CO_2 concentrations are based on the Representative Concentration Pathways (RCPs; van Vuuren et al., 2011). SSPs are paired with RCPs based on what sort of climate change could be expected under each SSP's storyline: SSP1 with RCP4.5, SSPs 3 and 4 with RCP6.0, and SSP5 with RCP8.5. RCP numbering refers to each scenario's average global radiative forcing (W m^{-2}) in 2100.

We use climate input data from the Fifth Coupled Model Intercomparison Project (CMIP5; Taylor et al., 2012) outputs of the IPSL-CM5A-MR climate model (Dufresne et al., 2013). Maps of temperature and precipitation change over the simulation period for each RCP are presented in the Supplementary Results (Fig. SR1). The CMIP5 runs did include land-use change, but not the trajectories output by PLUM. As such, and as with all models that are not climate-coupled but rather use offline forcings, we do not consider the effects of our simulated land-use change on climate. We also use just one climate model, and as such the only uncertainty explored in this work is uncertainty related to scenario choice.

Future socioeconomic data – country-level population and GDP projections – are taken from version 0.93 of the SSP database (IIASA, 2014).

~~In addition to the input data provided by the SSPs, and following the approach described in , PLUMparameters—such as~~ Demand of dedicated bioenergy crops such as *Miscanthus* is specified according to the SSP2 scenario from the MESSAGE-GLOBIOM model described by Popp et al. (2017); demand for bioenergy from food crops is specified to double from 2010 by 2030 and thereafter remain constant. The SSP narratives also affected parameters within PLUM. These included input and transport costs, tariffs, and minimum non-agricultural ~~area—were also affected by the SSP narratives~~area (which places an upper limit on the total fraction of a gridcell that PLUM can allocate to cropland and pasture). Values were estimated for each SSP based on an interpretation of the storylines (O'Neill et al., 2015; Engström et al., 2016a) and can be found in Table SM6. Because of these scenario-specific parameters, the raw PLUM outputs are not necessarily expected to match at the beginning of the period.

2.4.1 ~~Experimental setups~~

Historical land use areas (cropland and pasture fractions), irrigation, and synthetic nitrogen fertilizer application levels were taken from the Land Use Harmonization v2 dataset (LUH2; Hurtt et al., in prep.). Historical manure application rates (simplified upon import to LPJ-GUESS as pure nitrogen addition) come from Zhang et al. (2017). 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.

2.5 Ecosystem service indicators

~~In addition to the~~ LPJ-GUESS ~~runs forced with PLUM-output land use and management trajectories (harmonized as described in Sect.??), six experimental runs were performed for each scenario, to disentangle the direct effects of climate change (including CO_2 concentration increases)from those of land use and management change. A "constant-climate+CO_2" 2011–2100 run used repeating 1981–2010 climate and constant 2010~~ 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 ~~; separate “constant-climate” and “constant-CO_2” runs were also performed. A “constant-LU” run used constant 2010 land use maps and management inputs. Finally, “onlyCO~~([CO_2~~” and “onlyClimate” runs were performed using constant land use~~]), 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 ~~either CO_2 concentrations or climate forcings from the RCP scenarios. The results of these experimental runs were used to inform interpretation of the results but are mostly not presented here.The Supplementary Results file contains figures (numbers prefixed with SR) supporting claims derived from the experimental runs, in addition to other figures that were not included here to conserve space~~evaporation. We present change in average annual runoff as a general indicator of trend in water availability. After Asadieh and Krakauer (2017), we also use the difference between 1971–2000 and 2071–2100 in the 95th percentile of monthly surface runoff ($P95_{month}$) as a proxy for changing flood risk (although note that those authors used daily values), and the difference in the 5th annual percentile ($P5_{year}$) for changing drought risk. Note that we are referring to hydrologic drought, which can be contrasted with, e.g., meteorological drought (a long time with no or little precipitation) or socioeconomic drought (water supply levels too low to satisfy human usage demand; Wilhite and Glantz, 1985). 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 of changing risk 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). While LPJ-GUESS does not model the physical flow of water within and between gridcells, the predecessor LPJ model has been shown to compare well to dedicated hydrological models when aggregated to basin scale (Gerten et al., 2004). As such, where discussing geographic patterns, we will refer to basin-level results only.

Finally, we assess how much land is converted to agriculture within the Conservation International (CI) hotspots, a set of 35 regions covering less than 3% of the Earth's land area but containing half the world's endemic plant species and over 40% of the world's endemic vertebrate animal species (Myers et al., 2000; Mittermeier et al., 2004). These regions each contain at least 1500 endemic vascular plant species and have already lost at least 70% of their original natural vegetation, thus representing highly diverse areas presently at high risk of habitat loss. Note that our chosen metric does not consider areas where agricultural abandonment could lead to a long-term increase in biodiversity, because it is impossible to determine where and how soon, given enough newly available land, there would be sufficient vascular plant richness to qualify as a biodiversity hotspot.

3 Results and Discussion

3.1 Land use areas and management inputs

LandSyMM simulates net global loss of natural land area over the 21st century in all scenarios (Fig. 2), with SSP3 seeing the greatest loss of area (10%), SSP1 seeing the least (3%), and SSPs 4 and 5 seeing an intermediate loss (6%). These patterns are mostly reversed for pasture area change, in which all scenarios see an increase, although the trajectory for SSP5 is more similar to that of SSP3. PLUM also simulates net increased cropland area globally in all scenarios, with SSPs 1 and 5 seeing the least increase, SSP4 seeing more, and SSP3 seeing the most.

Cropland expansion happens at a more or less constant rate in SSP3 and SSP4, but these scenarios experience very different trajectories of crop commodity demand: SSP4 approximately levels off around midcentury, whereas SSP3 experiences only a brief slowdown in growth followed by constantly increasing demand through 2100 (Fig. ~~SR1~~SR2). The majority of the increased demand in the first half of the century is satisfied by fertilizer application, which increases by more than 75% from the 2010s to the 2050s while crop area increases by less than 15%. However, management inputs per hectare in SSP3-60 approximately plateau after midcentury (Fig. ~~SR2~~SR3), while crop demand rises 16%. Cropland area expands about 10% between 2050 and 2100, with boosted productivity—thanks to climate change and/or CO_2 fertilization—helping to satisfy the rest of the increased demand. Since SSP4-60 experiences the same climate and CO_2 fertilization but with level crop demand during the second half of the century, management inputs decrease after about 2050. PLUM prescribes lower irrigation rates by the end of the century for most scenarios (Figs. 2, SR3). This is enabled by higher global mean rainfall in all RCP scenarios, as evidenced by the bars for runoff in Figure 4, as well as by improved water-use efficiency for crops other than C_4 cereals due to increased CO_2 concentrations. Crop demand increase in SSP3-60 outweighs these effects, however, resulting in higher irrigation in that scenario.

Although population growth in SSP5-85 is more than twice that of SSP1-45, PLUM simulates very similar trajectories of global crop demand in both: an increase until about 2040 followed by a decrease for the rest of the century, with SSP5-85 crop demand ending slightly higher. SSP5-85 livestock demand increases about 20% more than in SSP1-45, which explains the rest of the difference in global caloric needs between the two scenarios (Fig. ~~SR1~~SR2). However, because SSP5-85 experiences much stronger climate change and CO_2 increase, the two scenarios differ importantly in how they satisfy their crop demand over the century. Whereas cropland area increases more or less constantly in SSP1-45 (slightly slowing throughout), in SSP5-85 it decreases ~~over the first two or so decades, increases slowly after about~~ through about 2050, ~~and ends~~ after which it increases slowly, ending at a slightly lower global extent than in SSP1-45 despite a jump in the early 2090s as feed becomes more important in raising ruminant livestock (Fig. ~~SR3~~SR4). Crop production remains similar between the two scenarios, especially in the first half of the century, because SSP5-85 applies much more fertilizer and irrigation water per hectare (Fig. ~~SR2~~SR3). This gap in these inputs narrows in the second half of the century as climate change and the CO_2 fertilization effect become even stronger in SSP5-85 relative to SSP1-45, although the latter also begins to increase ~~what PLUMcalls~~ PLUM's “other management” intensity (representing, e.g., pesticide application).

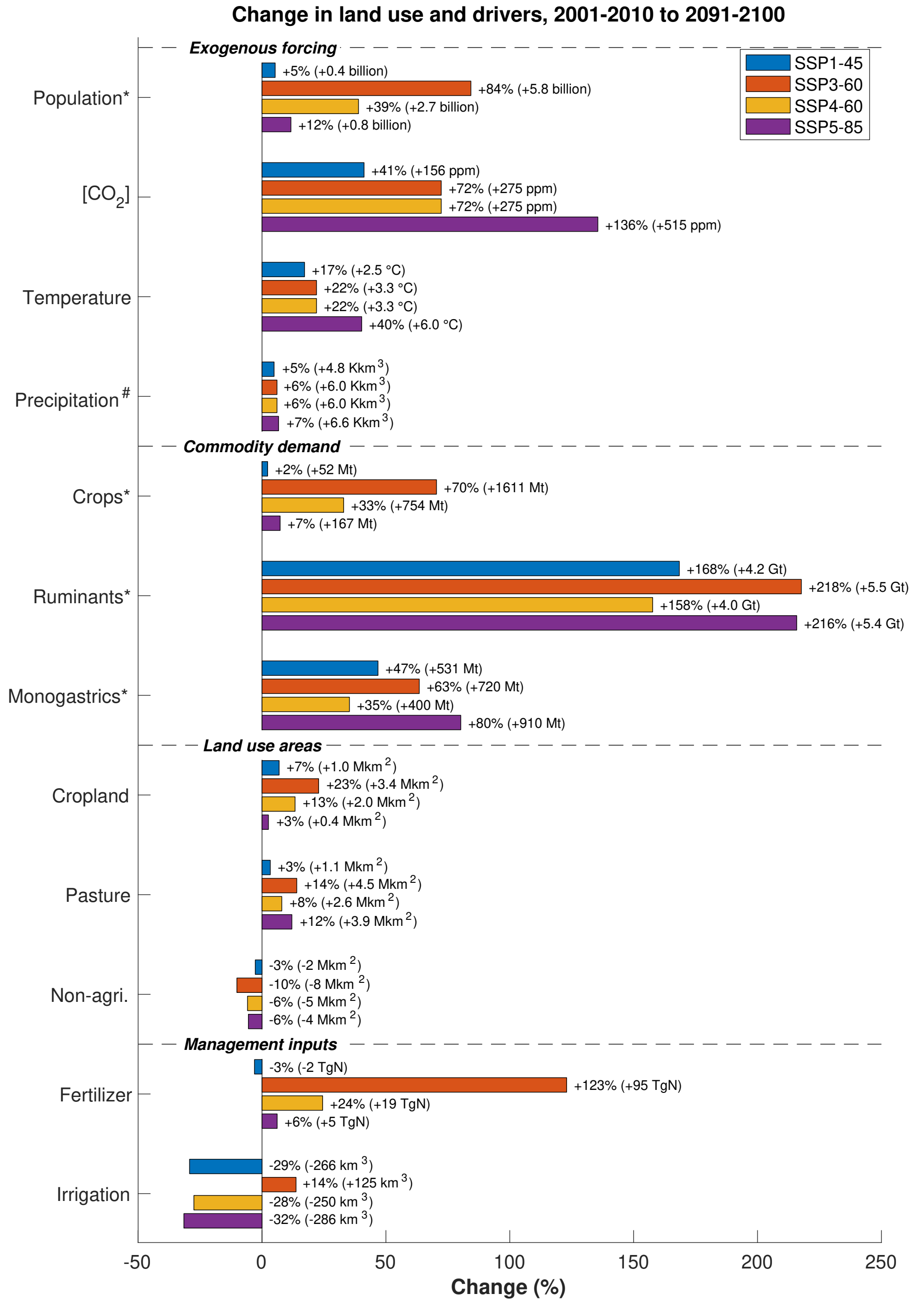


Figure 2. Percent change in global socioeconomic, land management, and atmospheric variables between 2001–2010 and 2091–2100. ~~Whiskers represent interannual variability as standard deviation of the difference (i.e.,~~ $\cancel{\sqrt{\sigma^2_{2001\text{-}2010}+\sigma^2_{2091\text{-}2100}}}$~~).~~ Ruminant demand given in units of feed-equivalent weight. *Asterisk indicates variables whose baseline is 2010 instead of 2001–2010 mean. #The time periods compared for precipitation were 1971–2000 and 2071–2100 due to high interannual variability.

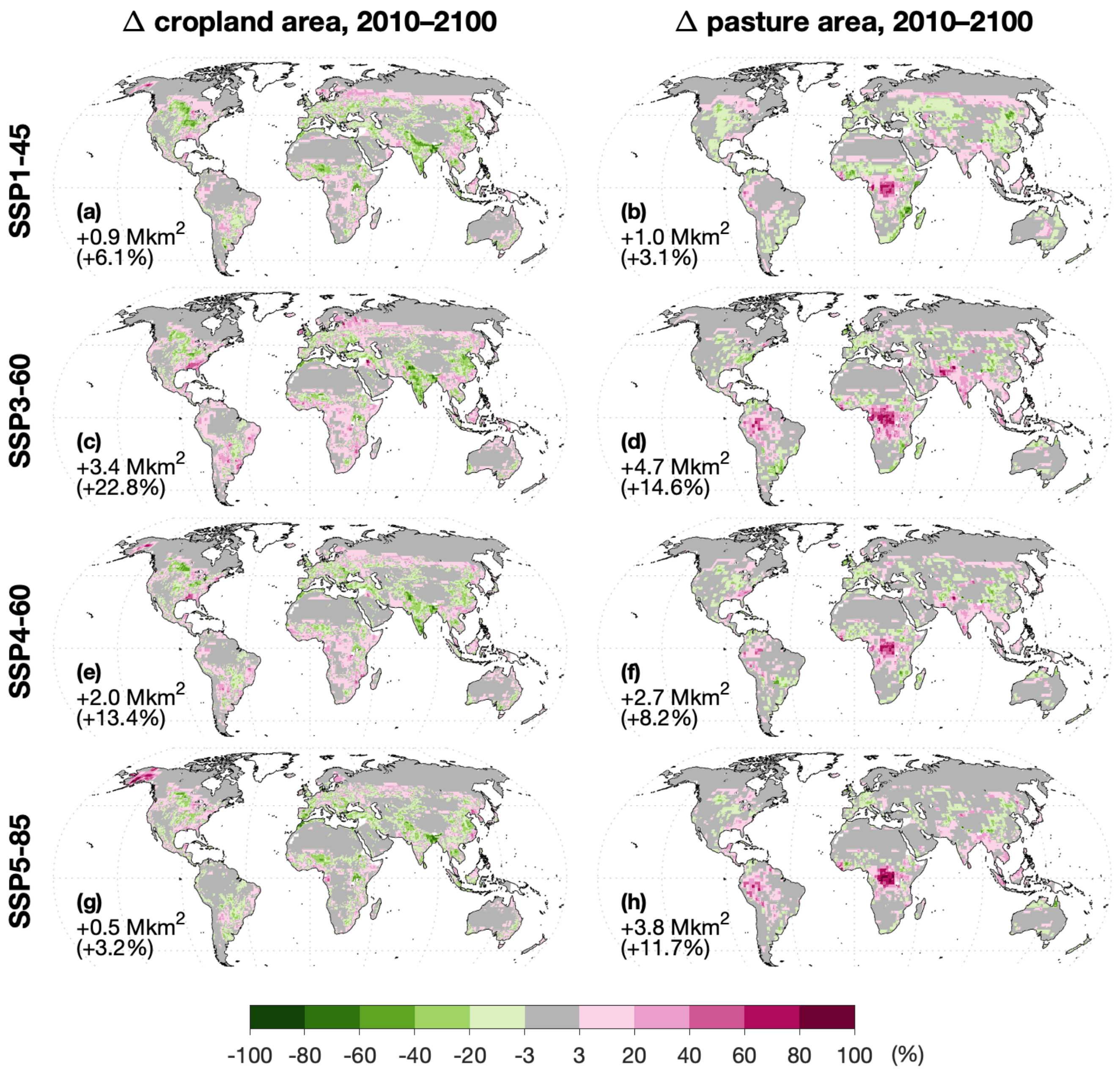


Figure 3. (**Left column**) Change in cropland area (as fraction of gridcell) from 2010 (LUH2) to 2100 (harmonized PLUM) under each SSP-RCP scenario. (**Right column**) As left column, but for pasture. ~~Note that color scales differ between columns.~~

Figure 3 presents the change in cropland and pasture area over 2010–2100 for each scenario after harmonization. It should be noted that the harmonization process, while preserving global changes in net area change for each land use type, produces more *gross* area change. Where relevant, in this section and the rest of the Results, we will point out where apparent strong regional effects of land-use change result from changes that were not present pre-harmonization.

Several regional patterns in crop area ~~stand out~~change stand out in Figure 3:

- North America loses cropland in parts of the Great Plains (mainly C_3 cereals; Fig. ~~SR4~~SR5) and the Midwestern U.S. (mainly oilcrops; Fig. ~~SR5~~SR6) in all scenarios ~~. This is at least partially replaced—to a varying extent among the scenarios—with land in Alaska and the southeastern~~ after harmonization. However, this is exaggerated relative to the original PLUM outputs by ca. 1500%, 1700%, 400%, and 800% for SSPs 1, 3, 4, and 5, respectively. Similarly, harmonization inflates projected cropland expansion in the temperate forests of the eastern U.S. ~~In the latter, new cropland is mostly planted with energy crops (SSP3-60 and SSP4-60; Fig. SR6) and rice (~~and Canada: by ca. 800% for SSP1-45 and 100% for the other scenarios. On the other hand, large-scale cropland expansion in Alaska in all scenarios except SSP3-60 ~~; Fig.SR7). New cropland in Alaska~~ was almost entirely present in the raw PLUM outputs. This new cropland is entirely planted with spring wheat (Fig. SR5) and is most extensive in SSP5-85~~(Fig.SR4)~~, which sees the largest increase in North American cereal demand—nearly 250% by the end of the century (Fig. ~~SR8~~SR7)—but also the largest potential yield increase in Alaska, thanks to ~~RCP8.5's~~ high warming and CO_2 fertilization. Indeed, by ~~2091–2100~~the end of the century, the potential yield of rainfed C_3 cereals there is similar to or exceeds that of the parts of the Great Plains where cropland is lost (Fig. ~~SR9).~~ SR8). It should be noted that while LPJ-GUESS includes several limitations on plant and soil processes based on air and soil temperature, the version used here does not represent permafrost dynamics, and so may be optimistic with regard to the increase in arable land area. However, permafrost extent is expected to decrease across parts of Alaska and the boreal zone as a whole, especially in high-temperature-increase scenarios such as RCP8.5 (Lawrence et al., 2012; Pastick et al., 2015).

- Although crop demand in South Asia (here, India, Sri Lanka, Pakistan, Afghanistan, Bangladesh, Nepal, and Bhutan) increases by more than 100% in SSP5-85 and 170% in SSP3-60 (Fig. ~~SR10~~SR9), that region net loses large amounts of cropland ~~. Imports increase from zero to 5–30~~after harmonization: ca. 30% ~~of South Asia's food and feed crop demand (respectively) by the end of the century, but to satisfy the remaining 70–95, PLUM calculates~~ and 20%, respectively. The raw PLUM outputs saw less loss (8% and 10%, respectively), but the same general pattern. Even so, PLUM projects that the region's crop production would approximately double in both scenarios to satisfy most of the increased demand (Fig. ~~SR11~~SR9). While some of this is accomplished through increased management inputs in a region where the yield gap is large in the baseline, it also depends markedly on yield boosts due to ~~climate change~~ increased rainfall (Fig. SR1) and rising CO_2 concentrations: C_3 cereal yields in the constant land-use experiments ~~triple~~ (~~RCP 6.0~~`rYYclico2`) triple (RCP6.0) or quadruple (~~RCP 8.5~~RCP8.5) across large parts of Pakistan and India. This is mostly due to a CO_2 fertilization effect, especially in ~~RCP 8.5~~RCP8.5, which sees widespread areas of yield decline ~~in the onlyClimate experiment (~~when only varying climate (`r85cli`, Fig. ~~SR12~~SR10).

- ~~Sub-Saharan Africa sees even larger~~ PLUM expects sub-Saharan Africa to see crop production increases even larger than South Asia, ranging from +200% in SSP1-45 to +500% in SSP3-60 (Fig. ~~SR13~~SR11). In contrast to South Asia, nearly the entire region experiences negative yield effects from the changing climate, and the counteracting effect of CO_2 fertilization results in yields that are only net slightly boosted in the constant-LU experiments (~~e.g.,~~ rYYclico2; Fig. ~~SR12~~SR10). The heightened production comes instead from increased management inputs and, to a much smaller degree, cropland expansion.

– China's crop demand peaks by about 2040; by the end of the century, it has either returned to (SSP3-60) or dropped past 2010 levels (by 30%, 40%, and 25% for SSP1-45, SSP4-60, and SSP5-85, respectively; Fig. SR12). Crop imports decrease from 14% of demand to less than 6%. This fits well with apparent net losses of cropland area in all scenarios, but note that harmonization switched SSP1-45's projection from an 8.5% gain to a 15% loss. Moreover, whereas PLUM projected cropland abandonment to occur in the montane shrublands and steppe of the Tibetan Plateau, after harmonization it occurs throughout the eastern temperate and subtropical forests. Slight cropland expansion projected by PLUM in China's subtropical moist forests is increased 300–600% by harmonization in all scenarios except SSP1-45 (+21%).

Pasture area is projected to expand significantly in the western Amazon in ~~SSP3-60 and SSP5-85~~all scenarios (although in SSP1-45 this is strongly exaggerated by harmonization), and even more so in all scenarios in the African rainforest (Fig. 3). This tropical deforestation is largely driven by the increasing consumption of ruminant products in those regions: As incomes increase in developing tropical countries, PLUM ~~forecasts~~ projects greater consumption of commodities such as meat and milk and a reduction in staples such as starchy roots and pulses (Keyzer et al., 2005; Tilman et al., 2011). Depending on the SSP, ruminant products are simulated to account for 23–43% of calories in central Africa by 2100, compared to only 4–7% of calories consumed in 2010 (caloric density derived from FAOSTAT, 2018c). Between 50 and 98% of the ruminant production increase in central Africa goes to this domestic consumption, with the rest being exported.

The African pasture expansion even occurs in SSP1-45, the "Sustainability" scenario (O'Neill et al., 2015), in which LandSyMM simulates a net global pasture expansion of about 1 Mkm^2. For comparison, five other land use models all saw SSP1 pasture area decrease~~,~~: by an average of about 3.4 Mkm^2 (Popp et al., 2017). ~~Some of the discrepancy between the LandSyMMland-use trajectories and those of the other modelsis likely due to inherent differences in how processes are represented~~While we do not expect LandSyMM's results to necessarily match those of other models, such a large, qualitative difference requires explanation. Several factors related to experimental setup and overall model structure ~~may contributeas well~~likely contribute.

First, PLUM makes no assumption about changes in food production needs besides what occurs due to population and GDP changes. The storyline for SSP1, however, with its "low challenges to mitigation," suggests that people will gradually shift to lower-meat diets (O'Neill et al., 2015) than would be expected given GDP levels, at first at least in high-income countries. IMAGE—which simulates a decrease in pasture area of about 7 Mkm^2 by the end of the century (Doelman et al., 2018)—incorporates this dietary shift as a 30% (global) reduction in meat consumption relative to what would have otherwise been simulated, and additionally includes a 33% reduction in food supply chain losses to represent efficiencies from improved management and infrastructure (Doelman et al., 2018). Weindl et al. (2017) use the MAgPIE model to show that, under a scenario like ours where historical differences in livestock production efficiency are maintained or exacerbated, a shift to lower-meat diets can reduce the expansion of pasture in sub-Saharan Africa by over 50%.

Second, the land-use modeling components of most integrated assessment models (IAMs)—for example, all those contributing to the LUH2 trajectories (Hurtt et al., 2011)—include demand for timber and other products. The carbon value of forests (and land more generally) can also be included by some, even if forest products are not explicitly modeled (e.g., MAgPIE; Humpenöder et al., 2014), which could come into play in scenarios with policy-based incentives designed to minimize emissions from deforestation and degradation and/or to maximize carbon sequestration. In contrast, PLUM includes neither forest products nor land carbon value. The only ~~“friction”~~ cost PLUM considers in converting a forest to agriculture is the cost of conversion, with the opportunity cost of lost forest products or services ignored. Similarly, the only incentive to replace existing agricultural land with forest would be to avoid costs associated with production. ~~Work currently underway to include~~ Including forest products, payments for carbon sequestration, and managed forestry into LandSyMM ~~may~~ could result in more forest simulated over the course of the century. This is especially likely for SSP1, whose storyline specifies a gradual improvement in how the global commons are managed (O’Neill et al., 2015). As an example, IMAGE represented this improvement in SSP1-45 by (a) disallowing clearing of forests with carbon density greater than 200 tons ha^{-1} and (b) reforesting half of the world’s degraded or former forest.

The ~~difference~~ spread in land-use area projections between the most extreme scenarios is much higher in this work than in Alexander et al. (2018), by around 500% for cropland and 700% for pasture. The primary reason for this increase in inter-scenario variation is that Alexander et al. (2018) used the SSP2 socioeconomic scenario for all RCPs, whereas here we compare different SSPs paired with appropriate RCPs. ~~Even with~~ The wide variation among the SSPs in population and economic growth trajectories, along with SSP-specific PLUM parameters (Sect. 2.4), contribute to this increased spread~~, however~~. Even so, the LandSyMM trajectories are more closely clustered than those from some other land use models. IMAGE, for example, projects a range of cropland area increase from 0.4 Mkm^2 to 5.3 Mkm^2 across the five SSPs, and pasture trajectories ranging from a decrease of 7.3 Mkm^2 to an increase of 4.4 Mkm^2 (Doelman et al., 2018). As described above, IMAGE makes a number of assumptions (based on the SSP storylines) that PLUM does not regarding future deviations from historical “business-as-usual” trends and relationships, including dietary shifts, ~~infrastructure efficiency improvements~~reductions in food losses during transport, and forest conservation.

3.2 Ecosystem service indicators

3.2.1 Carbon storage

Carbon stored in the land system increases for all SSP-RCP scenarios, primarily due to an increase in vegetation carbon (Fig. 4). The increase in each scenario relative to the others depends on both intensity of climate change as well as amount of natural land lost. The large increase of atmospheric CO_2 (and thus greater carbon fertilization) in ~~RCP 8.5 compared to RCP 6.0~~ RCP8.5 compared to RCP6.0 means that SSP5-85 has a much greater increase in terrestrial carbon storage than SSP4-60, despite those scenarios having similar trajectories of natural land area (Fig. 2). SSP3-60, which had the most natural land lost

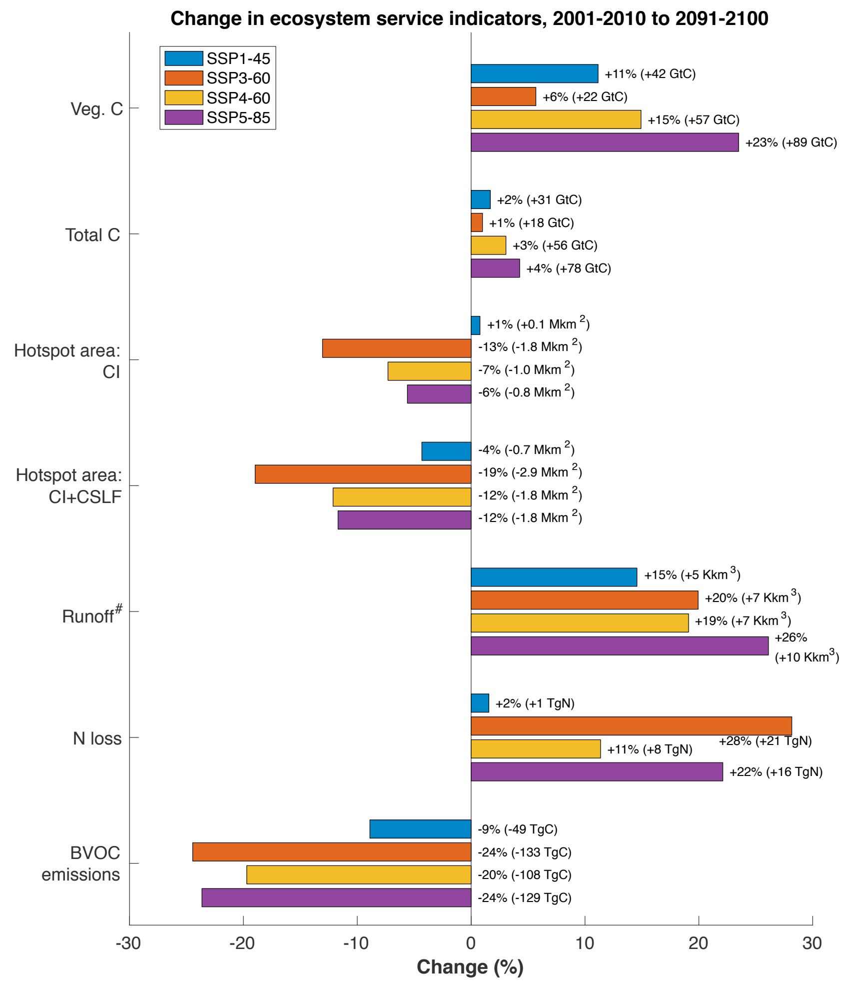


Figure 4. Percent global change in ecosystem service indicators between 2001–2010 and 2091–2100. ~~Whiskers represent standard deviation of the difference (i.e., $\sqrt{\sigma^2_{\text{2001-2010}} + \sigma^2_{\text{2091-2100}}}$).~~ CSLF: Congolian swamp and lowland forests (see Sect. 3.2.5). #The time periods compared for runoff were 1971–2000 and 2071–2100 due to high interannual variability.

but only intermediate carbon fertilization, sees the lowest increase in terrestrial C storage over the century—less than a third that of SSP4-60, which has the same trajectory of changing climate and atmospheric CO_2 concentration but a much smaller population increase.

The contrast between effects of changing climate and atmospheric CO_2 concentration vs. changing land use and management is starker for vegetation carbon than any other indicator variable examined here. In the constant-LU ~~experiment~~experiments (`rYYclico2`), vegetation carbon increased 35%, 43%, 43%, and 54% for SSP1-45, SSP3-60, SSP4-60, and SSP5-85, respectively. The experiments with constant-climate ~~+~~and CO_2 ~~experiment~~(`sXlum`), on the other hand, saw respective *decreases* of 5%, 15%, 8%, and 9%.

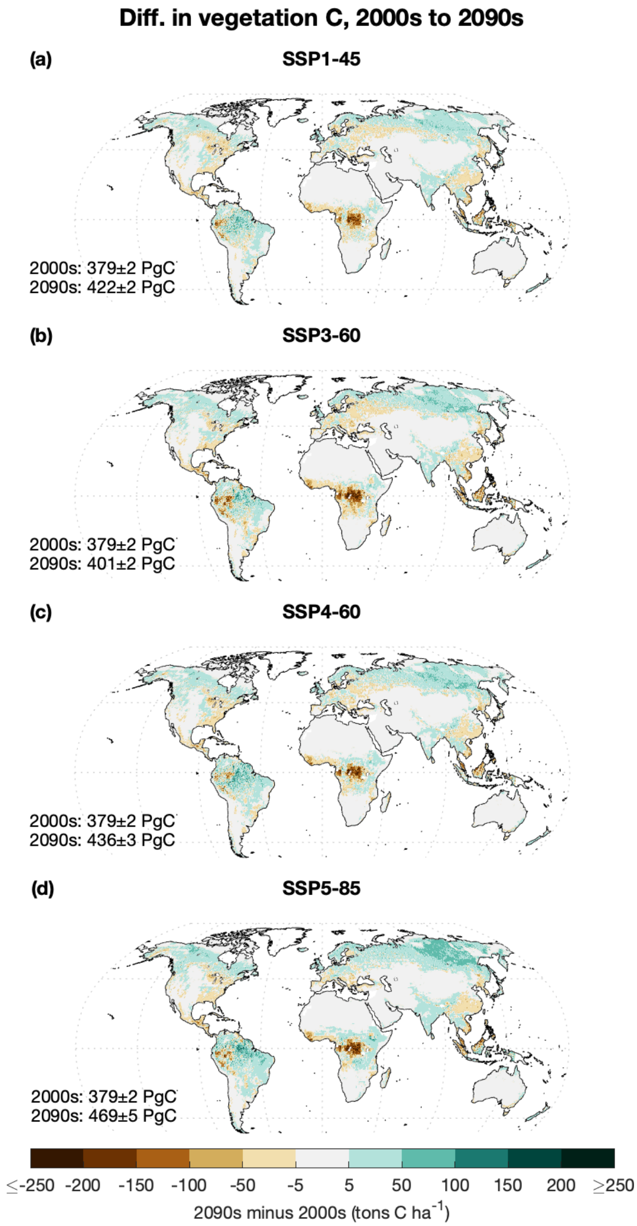


Figure 5. Maps showing difference in mean vegetation carbon between 2001–2010 (“2000s”) and 2091–2100 (“2090s”) for (**a**) SSP1-45, (**b**) SSP3-60, (**c**) SSP4-60, (**d**) SSP5-85. Overlaid text provides decadal means and standard deviations.

Vegetation carbon increases are most pronounced in the tropical and boreal forests (Fig. 5) and are due primarily to CO_2 fertilization, although increasing temperatures and growing season length also contribute in the boreal zone (Fig. ~~SR14~~SR13). Extensive conversion to pasture far outweighs any carbon fertilization effect in the African tropical forest, which loses nearly all of its vegetation carbon and up to half its total carbon by 2100 in all scenarios.

Our results for carbon sequestration fall near the lower end of previously-reported projections. Brovkin et al. (2013) examined the change in land carbon storage over 2006–2100 for a number of climate and land surface models. ~~They found that the~~ 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 with emissions and land use change from RCP8.5, simulated uptake of ~400 GtC. This is much greater than our finding of ~~~78~~ ~89 GtC under SSP5-85 (Fig. ~~5~~4) despite their land use change scenario (from LUH1; Hurtt et al., 2011) having lost about 30% more non-agricultural land. ~~The difference is probably due in large part to our pasture expansion in the central African and western Amazon rainforests (and to a lesser extent, cropland expansion in Alaska) . Most of their agricultural expansion occurs in the northern sub-Saharan African savannas, and even there the change (in units of percentage of the grid cell) only reaches about 30~~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% ~~. Moreover,~~ 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. ~~However, we also see low simulated carbon sequestration under constant land use compared to the models in , even when excluding models that did not simulate nitrogen limitation.~~

Another study with LPJ-GUESS, Krause et al. (2017), used land use trajectories from the IMAGE and MAgPIE IAMs given RCP2.6 and SSP2, finding an increase in total land carbon pools of 34 GtC and 64 GtC, respectively. Land use scenario played an important role in those results and likely contributes to the discrepancy with ours: Their IMAGE pasture area increased from ~35 Mkm^2 to ~40 Mkm^2, whereas their MAgPIE pasture area decreased to ~30 Mkm^2 and our SSP1-45 pasture stays around ~32 Mkm^2.The IMAGE cropland area used in the baseline run of Krause et al. (2017) stayed approximately constant at ~18 Mkm^2, as does our SSP1-45's (although at ~15 Mkm^2), but their MAgPIE cropland area increased to ~20 Mkm^2. Other important differences between the runs in Krause et al. (2017) and ours include our use of different climate forcings and ~~associated photosynthesis scaling parameters.~~ a different photosynthetic scaling parameter (which accounts for real-world reductions in light use efficiency; Haxeltine and Prentice, 1996).

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 (Fig. S3 in Ahlström et al., 2012).

3.2.2 Runoff

Global ~~runoff~~ precipitation increases in all scenarios (Fig. ~~6~~2). Again, SSP3 and SSP4 (the two RCP6.0 scenarios) see similar changes; SSP1-45 sees a smaller increase, and SSP5-85 sees the greatest. ~~Comparison~~ This pattern is roughly equivalent for changes in global runoff (Fig. 4); comparison of the experimental runs shows that climate change is the most important factor in ~~this increase~~ increasing runoff at a global level in all scenarios (e.g., Fig. ~~SR15). Increasing~~ SR14). While the impacts of increasing CO_2 levels ~~generally tend to decrease runoff , a result also~~ on runoff can be strongly regionally dependent (Zhu et al., 2012), we see that overall more CO_2 means less runoff at a global level. A similar result was seen in two global vegetation models analyzed by Davie et al. (2013), although two others in that comparison showed the opposite effect. Land-use change makes the least difference in terms of global annual runoff, but can be important at the regional level. Deforestation in central Africa, for example, is the primary driver of increasing mean annual runoff there because of reduced evapotranspiration relative to existing vegetation. Note, however, that LandSyMM can only represent the effect of land cover change on evapotranspiration and runoff directly—to include the impact of these flux differences on rainfall would require a coupling with a climate model.

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.

Between 1971–2000 and 2071–2100 under SSP5-85, ~~44~~basins comprising 48% of land area saw increasing flood risk, with a mean $P95_{month}$ increase of ~~48~~32% ~~. Drought~~ (Table 2). Basin-aggregated drought risk increased in ~~34~~37% of land area, which saw a mean $P5_{year}$ decrease of ~~59~~58%. At the same time, however, 43% of ~~gridcells~~ land area saw decreasing flood risk (mean $P95_{month}$ decrease ~~50~~42%), and ~~44~~54% saw decreasing drought risk (mean $P5_{year}$ increase ~~88~~49%). Other scenarios saw similar fractions of area affected, but smaller mean magnitude of change in flood or drought metric.

Most of the changes in SSP5-85 result from climate change, with some notable exceptions. Land-use change alone contributes notably to increasing drought risk in ~~Iraq~~eastern China, Pakistan, ~~India's Punjab province, the Great Plains of the U.S. , and eastern China~~ and northwest India (Fig. 6a), although the cropland abandonment driving most of these changes is more densely concentrated pre-harmonization. Agricultural expansion in Alaska and central Africa increase flood risk, while cropland abandonment in southern Pakistan decreases it (Fig. 6~~); it contributes to~~ b). Similar effects in other regions in Fig. 6—for example, increasing drought risk in Iraq and the central United States, and increasing flood risk in ~~northeastern China, India's Rajasthan province, and the central African rainforest.~~northeast China—are driven by land-use changes induced mostly by harmonization. (These land-use changes would of course be happening *somewhere*, and thus could still affect runoff similarly, but in a different and potentially more concentrated region.) Land-use change can also serve to counteract the impacts of climate change on runoff. For example, the severity of very low runoff events increases in central America, but it would have increased more if not for the expansion of agriculture there. The effects of land-use change on runoff might be stronger and more widespread if LPJ-GUESS were run coupled with a climate model, which would account for associated changes in land-atmosphere water and energy fluxes that can have similar impacts on the hydrological cycle as greenhouse gas emissions (Quesada et al., 2017).

~~The LandSyMM~~ Our results for SSP5-85 are compared with the RCP8.5 ensemble from Asadieh and Krakauer (2017) in Table 2. In all categories, LandSyMM finds a mean effect of stronger magnitude. LandSyMM finds less land in basins with increasing drought risk and more with decreasing drought risk than the Asadieh and Krakauer (2017) ensemble. Our results for fraction of land in ~~each class are roughly similar for most classes, except for increasing drought risk , more of which is found in~~ basins with increasing or decreasing flood risk are more similar (within 6 percentage points) to those from the Asadieh and Krakauer (2017) ensemble. ~~In all categories, LandSyMM finds a mean effect of stronger magnitude. Some differences between~~

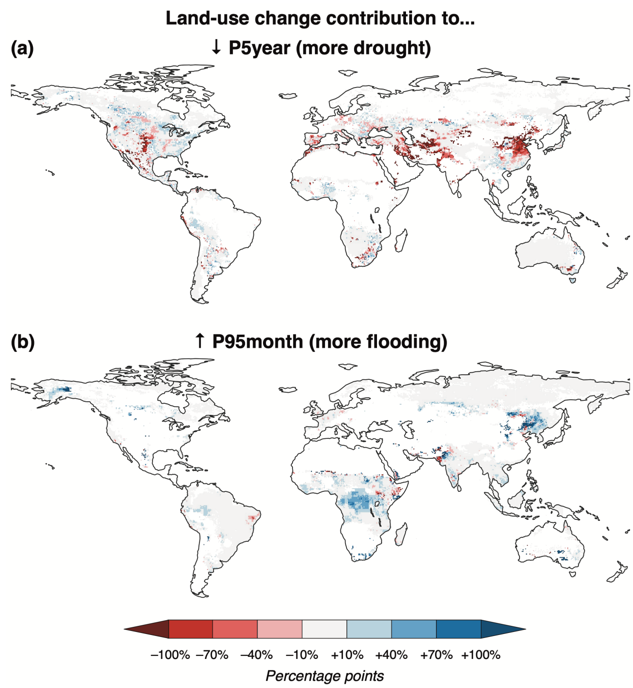


Figure 6. Contribution of land-use change in SSP5-85 to (**a**) decreasing $P5_{year}$ (drought) and (**b**) increasing $P95_{month}$ (flooding) between 1971–2000 and 2071–2100. White areas either did not have decreasing $P5_{year}$ or increasing $P95_{month}$, respectively, or were excluded due to low baseline runoff (after Asadieh and Krakauer, 2017). Contribution calculated as difference between full run and constant-LU run (i.e., sXlum_rYYclico2 − rYYclico2).

~~our results and those of might be expected because we used a single climate model and monthly instead of~~ 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 ~~. Perhaps more importantly, however~~as Asadieh and Krakauer (2017) did, instead of the LPJ-GUESS-output monthly values. We also expect that the average magnitude of change in those areas would have been closer to zero.

Another important difference between Asadieh and Krakauer (2017) and our analysis is that, whereas that study used five climate models, we used only one. Specifically, compared to 18 other models examined in Ahlström et al. (2012, their Fig. S2), IPSL-CM5A-MR in RCP8.5 simulates a much larger precipitation increase around the Equator, where we see the largest increase in runoff (Fig. SR14a). Finally, LPJ-GUESS is not a full hydrological model: e.g., it does not include river routing. ~~The effect of this can be somewhat explored by aggregating our results to the basin scale, at which LPJ has been shown to compare well to dedicated hydrological models . Aggregating to the basins with which PLUM determines irrigation water availability does not change the fraction of included land area in any class by more than two percentage points, except for decreasing drought risk, which goes from 49to 58of included land. The mean change within every class is reduced in absolute terms, bringing our results closer to those from~~ Land surface and hydrological models that include river routing, such as those

Table 2. Fraction of ~~included~~ land ~~in each group of~~ with changing drought and/or flood risk between the last three decades of the 20th and 21st centuries in SSP5-85. Numbers in parentheses give each group's mean percent change in runoff. ~~Increasing and decreasing drought risks correspond respectively to decreasing and increasing P5; increasing and decreasing flood risks correspond respectively~~ LandSyMM results have been aggregated to ~~increasing and decreasing P95~~basin scale. AK2017: Asadieh and Krakauer (2017).

Class	~~By gridcell By basin~~ LandSyMM	AK2017
↑ drought risk ~~38(–59~~(↓ P5)	~~39~~37% (–58%)	~~57~~43% (–51%)
↓ drought risk ~~49(+88~~(↑ P5)	~~58~~54% (+49%)	~~43~~33% (+30%)
↑ flood risk ~~49(+48~~(↑ P95)	~~51~~48% (+32%)	~~48~~37% (+25%)
↓ flood risk ~~47(–50~~(↓ P95)	~~46~~43% (–42%)	~~52~~40% (–21%)

included in the Asadieh and Krakauer (2017) ensemble, are needed to fully explore how changing precipitation, transpiration, and evaporation actually translate into changes in streamflow and surface water levels.

3.2.3 Nitrogen losses

While the evolution of total global nitrogen loss is fairly similar for all scenarios over the first two decades of the simulation, there are notable differences by the end of the century. N loss showed little net change over the century in SSP1-45, whereas SSP3-60 saw a large increase in total N losses. SSP4-60 did not see a notable change in total N losses. SSP5-85 saw an increase in total N losses, although less than that of SSP3-60.

Our N loss at the end of the historical period was similar to that of Krause et al. (2017), but whereas their runs estimated an increase in N losses of 60–80% under RCP2.6, ours under SSP1-45 increased only 2%. Krause et al. (2017) used fertilizer information from IMAGE and MAgPIE ~~, which they note~~ . Strong increases in fertilizer in those models resulted in strongly increased yields, but nitrogen limitation is alleviated at much lower levels in LPJ-GUESS. IMAGE and MAgPIE fertilization rates thus often exceeded what plants in LPJ-GUESS could actually take up, resulting in high amounts of N loss. Coupling LPJ-GUESS with PLUM provides for a more internally consistent estimate of future N losses, while still reproducing historical fertilizer application well (Alexander et al., 2018).

One interesting pattern is that climate and management changes can have similar effects on N losses. SSP3-60 sees global fertilizer application more than double by the end of the century, while SSP5-85 fertilizer application at end of the run is slightly lower than in 2011 (Fig. 2). This is reflected in the N losses for the ~~constant-climate+CO_2 experiment~~`sXlum` experiments, which increase 25% by the 2090s ~~in SSP3-60~~ with SSP3 but only 7% ~~in SSP5-85~~with SSP5. However, in the full ~~run~~runs (`sXlum_rYYclico2`), SSP3-60's N losses increase only about 27% more than SSP5-85's (Fig. 4). This is because the latter experiences higher average global temperatures (increasing gaseous losses) and a greater increase in runoff (increasing dissolved losses), due to the extreme RCP8.5 climate change scenario; in the constant-LU ~~experiment~~(`rYYclico2`) experiments, N losses ~~in SSP3-60 and SSP5-85~~ with RCP6.0 and RCP8.5 increase by 15% and 24%, respectively. In either case—but especially under SSP3-60—these increases in fertilizer usage and concomitant nitrogen pollution would exacerbate humanity's already unsustainable impacts on nutrient cycling (Rockström et al., 2009).

3.2.4 BVOCs

Global combined BVOC emissions over 2001–2010 totaled ~546 TgC yr^{-1} (~503 and ~43 TgC yr^{-1} for isoprene and monoterpenes, respectively), which compares well with estimates from LPJ-GUESS ~~(without PLUM) and other models~~ using different land use scenarios (Arneth et al., 2008; Hantson et al., 2017; Szogs et al., 2017) and the MEGAN model (Sindelarova et al., 2014). Emissions decline in all scenarios: by the most in SSP3-60 and SSP5-85, slightly less in SSP4-60, and the least in SSP1-45 (Fig. 4). This reflects a combination of the effects of land-use change and [CO_2] increases. In the ~~"constant-climate+CO_2" experiment~~`sXlum` experiments, declines in combined BVOC emissions closely reflect declines in non-agricultural land area (most decrease ~~in SSP3-60, less in SSP4-60 and SSP5-85~~with SSP3, less with SSP4 and SSP5, and least ~~in SSP1-45~~with SSP1; Fig. ~~SR16~~SR15). This is a function of the much higher BVOC emissions potential of forests relative to cropland and pasture, as also seen in results from Hantson et al. (2017) and Szogs et al. (2017). In the full ~~experiment~~runs (`sXlum_rYYclico2`), BVOC emissions decline more in SSP5-85 than in SSP4-60 because the former has higher atmospheric [CO_2], which suppresses BVOC formation (Arneth et al., 2007a). 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).

Decreases in isoprene emissions are primarily driven by tropical deforestation for agriculture, especially the expansion of pasture in central Africa and South America, and to a lesser extent by the expansion of cropland ~~for bioenergy~~ in the southeastern U.S (Fig. ~~SR17)~~SR16), although the latter is exaggerated by harmonization. The suppressive effect of increasing [CO_2] is mostly counteracted in all RCPs by rising temperatures, which increase BVOC volatility. Monoterpene emissions in what is now tundra increase as woody vegetation expands there, but present-day boreal forests are the main ~~drivers~~ areas of declining monoterpene emissions (Fig. SR17). This is primarily due to the BVOC-suppressing effect of increasing [CO_2], but land-use change also contributes, especially in Alaska~~(Fig.SR18).~~.

It is important to keep in mind that the implications of changing BVOC emissions depend on complex, regionally-varying atmospheric chemistry that governs their effects on existing species (e.g., methane) and the formation of secondary products (e.g., ozone and aerosols). The LandSyMM framework, lacking as it does an atmospheric chemistry model, can thus inform only a surface-level discussion of the possible effects of changing BVOCs. 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.

The globally decreased BVOC emissions in all scenarios could contribute a cooling effect in the future, due to expected lower tropospheric ozone concentrations, shorter methane lifetime, and enhanced photosynthesis thanks to more diffuse radiation. This ~~would~~ could be counteracted somewhat by warming arising from the reduced formation of secondary aerosols, and it is important to note that the effects on climate are likely to vary from region to region (Rosenkranz et al., 2014). Southeast Asia and the southeastern U.S. are populous areas that could see public health benefits from the deforestation-induced reduction of isoprene emissions and associated ozone levels. However, ~~note that in the latter (and to a lesser extent the~~

~~former),~~ a sizable portion of that agricultural expansion is ~~mainly~~ for growing bioenergy crops simulated in LPJ-GUESS as ~~Miscanthus~~*Miscanthus*; BVOC levels would be reduced much less (or perhaps even increased) if woody bioenergy crops were grown instead on the same area (Rosenkranz et al., 2014), but that possibility is not yet included in LandSyMM. 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.

~~It is important to keep in mind that the implications of changing BVOC emissions depend on complex, regionally-varying atmospheric chemistry that governs their effects on existing species (e.g., methane) and the formation of secondary products (e.g., ozone and aerosols). The LandSyMMframework, lacking as it does an atmospheric chemistry model, can thus inform only a surface-level discussion of the possible effects of changing BVOCs~~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.

3.2.5 Biodiversity hotspots

The large expansion of agricultural land in SSP3-60 has direct consequences for habitat in biodiversity hotspots, which lose over 13% of their non-agricultural land in that scenario (Fig. 4). No other scenario lost more than 8%, and SSP1-45 actually saw a slight gain. However, note that the central African rainforest is not included in the CI hotspots, since that region did not meet the criterion regarding how much of its primary vegetation had been lost (Myers et al., 2000; Mittermeier et al., 2004). The amount of deforestation projected there in all ~~scenarios would undoubtedly~~ scenarios—ranging from more than 50% in SSP1-45 to 77% in SSP3-60—could result in great impacts to regional ~~and global~~ biodiversity. We thus checked how much ~~hotspot~~ area is lost if we include the five ecoregions classified by Olson et al. (2001) as Congolian swamp and lowland forests (CSLF), which together roughly correspond to the area of pasture expansion common to all scenarios, into a new "CI+CSLF" hotspot map. This paints a worse picture in all scenarios, nearly doubling hotspot area loss in SSP3-60, more than doubling it in SSP4-60 and SSP5-85, and changing the 1% gain of SSP1-45 to a 9% loss.

~~Previously, performed a similar analysis using the same hotspots, but with the LUH1 land-use scenarios over 2005–2100. Although they considered only primary land as habitat—i.e., any land that had once been agriculture or experienced wood harvest was "uninhabitable"—their results can still provide a useful perspective. found losses of primary natural hotspot land of 25in RCP4.5, 40in RCP6.0, and 58in RCP8.5. However, they found a smaller effect of land-use change using species-area curves to estimate the number of endemic species driven to extinction: 0.2–25between 2005 and 2100. This is in part due to the shape of the species-area curves used in their calculations, in which the rate of extinctions per hectare lost is high at the beginning of land clearance in a region but falls as more area is cleared. This nonlinear effect is important to consider, especially considering how much land has (by definition) been cleared already in the hotspots, but such an analysis is beyond the scope of the present study.~~

Hof et al. (2018) considered the effects of both climate and land-use change under RCPs 2.6 and 6.0 on species distribution models of amphibians, birds, and mammals. They found that the area of land impacted by these combined threats was approximately equal between the two scenarios for birds and mammals (with more area affected for amphibians under RCP6.0),

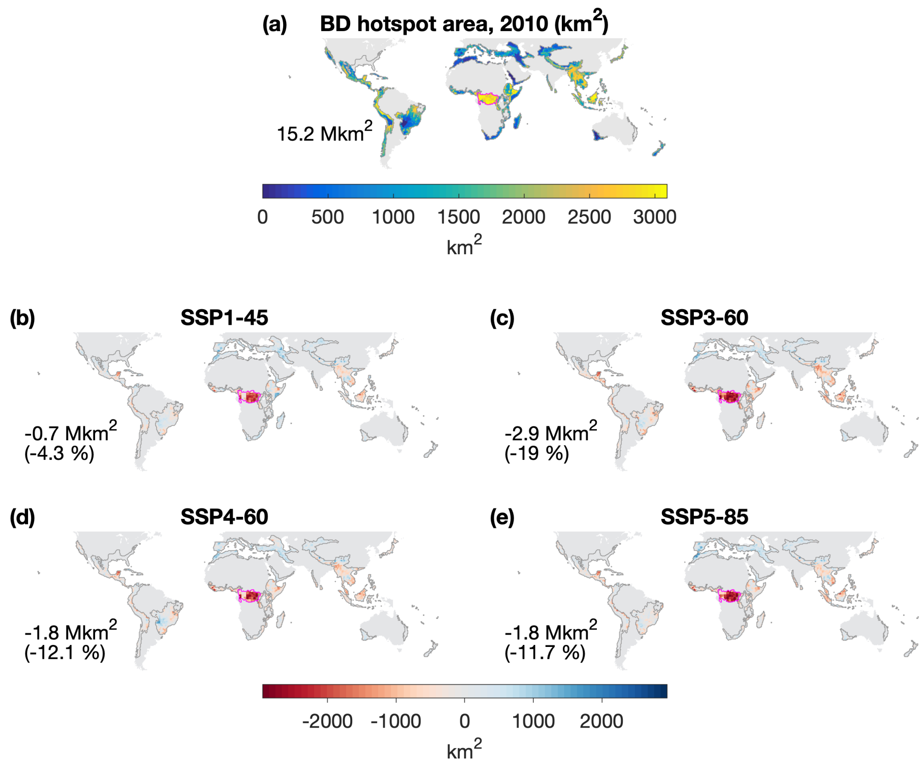


Figure 7. (**a**) Area (from LUH2) of non-agricultural land in "CI+CSLF" hotspots in 2010; (**b–e**) change in non-agricultural land area there by 2100 for each scenario. Black outlines indicate CI hotspots; magenta outline indicates CSLF region.

because although climate change was less detrimental under RCP2.6, to meet such an ambitious climate change target, that scenario required more land devoted to growing bioenergy crops. We ~~may~~ see a similar effect~~between SSP4-60 and SSP5-85: if ignoring cropland planted with Miscanthus, we see losses~~: If ignoring *Miscanthus* area, loss of natural land ~~area in the~~ in CI+CSLF hotspots ~~of approximately 1.1 and 1.5 Mkm^2, respectively, compared to about 1.8 Mkm^2 lost in both scenarios with Miscanthus cropland included. Ignoring Miscanthus cropland results in a CI+CSLF hotspot area net gain of <0.1 Mkm^2 for SSP1-45 and a lossof 2.0 Mkm^2 for~~ 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 means less land 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.

It should be noted that area loss in biodiversity hotspots will not necessarily correspond to linear decreases in species richness. Jantz et al. (2015) considered the losses of primary non-agricultural land in the LUH1 land use trajectories (Hurtt et al., 2011), which between 2005 and 2100 were 25% in RCP4.5, 40% in RCP6.0, and 58% in RCP8.5. (Note that Jantz et al. considered only primary land as habitat: i.e., any land that had once been agriculture or experienced wood harvest was "uninhabitable.") However, they translated those values into 0.2–25% of hotspot-endemic species driven to extinction by habitat loss. This is smaller than the fraction of land area because Jantz et al. (2015) used species-area curves, which model the rate of extinctions per hectare lost as high at the beginning of land clearance in a region but falling as more area is cleared. This nonlinear effect is important to consider, especially considering how much land has (by definition) been cleared already in the hotspots, but such an analysis is beyond the scope of the present study. Thus, our numbers for fraction of habitat lost (or gained) should not be considered to translate directly into extinction estimates.

4 Conclusions

This work is among the first to comprehensively consider the impacts of future land use and land management change on a suite of ecosystem services under different ~~scenarios~~ possible futures of climate and socioeconomic change. ~~We show that storylines~~ Using a uniquely spatially-detailed, process-based coupled model system, we show that scenarios with high socioeconomic challenges to climate change mitigation—SSP3 and SSP5—consistently have some of the most severe consequences for the natural world and the benefits it provides humanity via carbon sequestration, biodiversity, and water regulation. These two scenarios also most strongly affect biogeochemical cycling of nitrogen and BVOCs; while increases in nitrogen losses are generally detrimental, the impact of decreased BVOC emissions is likely to vary regionally. However, various elements of uncertainty—related to PLUM parameter values, global climate model selection, and model design—affect these results and remain to be explored.

Policymakers and other stakeholders need options for how we can meet the needs of a growing and changing society while achieving climate and sustainable development goals (Benton et al., 2018). Some progress has already been made in this regard at landscape and global scales (Eitelberg et al., 2016; Verhagen et al., 2018). LandSyMM, and analyses it enables such as the ones presented here, can be another powerful tool in this aspect of the science-policy interface.

Code availability. The code for harmonizing land use and management is available for download on GitHub (Rabin, 2019).

Author contributions. All authors contributed to the conceptual design of LandSyMM. S. Rabin composed most of this manuscript, although all authors contributed to its editing. S. Rabin performed most analyses, with R. Henry helping to interpret PLUM results. P. Alexander and R. Henry managed PLUM code and performed PLUM runs. S. Rabin made changes as described to LPJ-GUESS code and performed LPJ-GUESS runs.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. A.A. and S.R. acknowledge funding from the Helmholtz Association Impulse and Networking fund and the HGF ATMO programme. P. Alexander and R.H. were supported by the UK's Global Food Security Programme project "Resilience of the UK food system to Global Shocks" (RUGS, BB/N020707/1). In addition, the authors would like to thank Jonathan Doelman for sharing data about the IMAGE scenarios (Doelman et al., 2018). This is paper number FILL ON ACCEPTANCE of the Birmingham Institute of Forest Research.

References

Ahlström, A., Schurgers, G., Arneth, A., and Smith, B.: Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections, Environmental Research Letters, 7, 044 008–10, 2012.

Alexander, P., Rabin, S. S., Anthoni, P., Henry, R., Pugh, T. A. M., Rounsevell, M. D. A., and Arneth, A.: Adaptation of global land use and management intensity to changes in climate and atmospheric carbon dioxide, Global Change Biology, 24, 2791–2809, 2018.

Arneth, A., Miller, P. A., Scholze, M., Hickler, T., Schurgers, G., Smith, B., and Prentice, I. C.: CO_2 inhibition of global terrestrial isoprene emissions: Potential implications for atmospheric chemistry, Geophysical Research Letters, 34, 73–5, 2007a.

Arneth, A., Niinemets, U., Pressley, S., Bäck, J., Hari, P., Karl, T., Noe, S., Prentice, I. C., Serça, D., Hickler, T., Wolf, A., and Smith, B.: Process-based estimates of terrestrial ecosystem isoprene emissions: incorporating the effects of a direct CO_2-isoprene interaction, Atmospheric Chemistry and Physics, 7, 31–53, 2007b.

Arneth, A., Monson, R. K., Schurgers, G., Niinemets, U., and Palmer, P. I.: Why are estimates of global terrestrial isoprene emissions so similar (and why is this not so for monoterpenes)?, Atmospheric Chemistry and Physics, 8, 4605–4620, 2008.

Arthur, W. B.: Complexity, the Santa Fe approach, and non-equilibrium economics, The Economic Journal, 18, 149–166, 2010.

Asadieh, B. and Krakauer, N. Y.: Global change in streamflow extremes under climate change over the 21st century, Hydrology and Earth System Sciences, 21, 5863–5874, 2017.

Ashmore, M. R.: Assessing the future global impacts of ozone on vegetation, Plant, Cell and Environment, 28, 949–964, 2005.

Bayer, A. D., Lindeskog, M., Pugh, T. A. M., Anthoni, P. M., Fuchs, R., and Arneth, A.: Uncertainties in the land-use flux resulting from land-use change reconstructions and gross land transitions, Earth System Dynamics, 8, 91–111, 2017.

Benton, T. G., Bailey, R., Froggatt, A., King, R., Lee, B., and Wellesley, L.: Designing sustainable landuse in a 1.5°C world: the complexities of projecting multiple ecosystem services from land, Current Opinion in Environmental Sustainability, 31, 88–95, 2018.

Brovkin, V., Boysen, L., Arora, V. K., Boisier, J. P., Cadule, P., Chini, L., Claussen, M., Friedlingstein, P., Gayler, V., van den Hurk, B. J. J. M., Hurtt, G. C., Jones, C. D., Kato, E., de Noblet-Ducoudré, N., Pacifico, F., Pongratz, J., and Weiss, M.: Effect of anthropogenic land-use and land-cover changes on climate and land carbon storage in CMIP5 projections for the twenty-first century, Journal of Climate, 26, 6859–6881, 2013.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R., Piao, S., and Thornton, P.: 6: Carbon and Other Biogeochemical Cycles, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge, United Kingdom and New York, NY, USA, 2013.

Dankers, R., Arnell, N. W., Clark, D. B., Falloon, P. D., Fekete, B. M., Gosling, S. N., HEINKE, J., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble, Proceedings of the National Academy of Sciences of the United States of America, 111, 3257–3261, 2014.

Davie, J. C. S., Falloon, P. D., Kahana, R., Dankers, R., Betts, R. A., Portmann, F. T., Wisser, D., Clark, D. B., Ito, A., Masaki, Y., Nishina, K., Fekete, B., Tessler, Z., Wada, Y., Liu, X., Tang, Q., Hagemann, S., Stacke, T., Pavlick, R., Schaphoff, S., Gosling, S. N., Franssen, W., and Arnell, N.: Comparing projections of future changes in runoff from hydrological and biome models in ISI-MIP, Earth System Dynamics, 4, 359–374, 2013.

Di Marco, M., Harwood, T. D., Hoskins, A. J., Ware, C., Hill, S. L. L., and Ferrier, S.: Projecting impacts of global climate and land-use scenarios on plant biodiversity using compositional-turnover modelling, Global Change Biology, 1, 173–16, 2019.

Doelman, J. C., Stehfest, E., Tabeau, A., van Meijl, H., Lassaletta, L., Gernaat, D. E. H. J., Hermans, K., Harmsen, M., Daioglou, V., Biemans, H., van der Sluis, S., and van Vuuren, D. P.: Exploring SSP land-use dynamics using the IMAGE model: Regional and gridded scenarios of land-use change and land-based climate change mitigation, Global Environmental Change-Human and Policy Dimensions, 48, 119–135, 2018.

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J. P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J. Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M. P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, 2013.

Ebi, K. L. and McGregor, G.: Climate Change, Tropospheric Ozone and Particulate Matter, and Health Impacts, Environmental Health Perspectives, 116, 1449–1455, 2008.

Eitelberg, D. A., van Vliet, J., Doelman, J. C., Stehfest, E., and Verburg, P. H.: Demand for biodiversity protection and carbon storage as drivers of global land change scenarios, Global Environmental Change-Human and Policy Dimensions, 40, 101–111, 2016.

Elliott, J., Deryng, D., Müller, C., Frieler, K., Konzmann, M., Gerten, D., Glotter, M., Flörke, M., Wada, Y., Best, N., Eisner, S., Fekete, B. M., Folberth, C., Foster, I., Gosling, S. N., Haddeland, I., Khabarov, N., Ludwig, F., Masaki, Y., Olin, S., Rosenzweig, C., Ruane, A. C., Satoh, Y., Schmid, E., Stacke, T., Tang, Q., and Wisser, D.: Constraints and potentials of future irrigation water availability on agricultural production under climate change., Proceedings of the National Academy of Sciences, 111, 3239–3244, 2014.

Engström, K., Olin, S., Rounsevell, M. D. A., Brogaard, S., van Vuuren, D. P., Alexander, P., Murray-Rust, D., and Arneth, A.: Assessing uncertainties in global cropland futures using a conditional probabilistic modelling framework, Earth System Dynamics, 7, 893–915, 2016a.

Engström, K., Rounsevell, M. D. A., Murray-Rust, D., Hardacre, C., Alexander, P., Cui, X., Palmer, P. I., and Arneth, A.: Applying Occam's razor to global agricultural land use change, Environmental Modelling and Software, 75, 212–229, 2016b.

FAOSTAT: Commodity Balances/Crops Primary Equivalent (2018-09-24), Food and Agriculture Organization of the United Nations, 2018a.

FAOSTAT: Commodity Balances/Livestock and Fish Primary Equivalent (2018-09-24), Food and Agriculture Organization of the United Nations, 2018b.

FAOSTAT: Food Supply (2018-09-24), Food and Agriculture Organization of the United Nations, 2018c.

Feng, Z. and Kobayashi, K.: Assessing the impacts of current and future concentrations of surface ozone on crop yield with meta-analysis, Atmospheric Environment, 43, 1510–1519, 2009.

Fowler, D., Pilegaard, K., Sutton, M. A., Ambus, P., Raivonen, M., Duyzer, J., Simpson, D., Fagerli, H., Fuzzi, S., Schjoerring, J. K., Granier, C., Neftel, A., Isaksen, I. S. A., Laj, P., Maione, M., Monks, P. S., Burkhardt, J., Daemmgen, U., Neirynck, J., Personne, E., Wichink-Kruit, R., Butterbach-Bahl, K., Flechard, C., Tuovinen, J. P., Coyle, M., Gerosa, G., Loubet, B., Altimir, N., Gruenhage, L., Ammann, C., Cieslik, S., Paoletti, E., Mikkelsen, T. N., Ro-Poulsen, H., Cellier, P., Cape, J. N., HorvAth, L., Loreto, F., Niinemets, U., Palmer, P. I., Rinne, J., Misztal, P., Nemitz, E., Nilsson, D., Pryor, S., Gallagher, M. W., Vesala, T., Skiba, U., BrUggemann, N., Zechmeister-Boltenstern, S., Williams, J., Dowd, C. O., Facchini, M. C., de Leeuw, G., Flossman, A., Chaumerliac, N., and Erisman, J. W.: Atmospheric composition change: Ecosystems-Atmosphere interactions, Atmospheric Environment, 43, 5193–5267, 2009.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., and Sitch, S.: Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model, Journal of Hydrology, 286, 249–270, 2004.

Haddeland, I., HEINKE, J., Biemans, H., Eisner, S., Flörke, M., Hanasaki, N., Konzmann, M., Ludwig, F., Masaki, Y., Schewe, J., Stacke, T., Tessler, Z. D., Wada, Y., and Wisser, D.: Global water resources affected by human interventions and climate change, Proceedings of the National Academy of Sciences of the United States of America, 111, 3251–3256, 2014.

Hantson, S., Knorr, W., Schurgers, G., Pugh, T. A. M., and Arneth, A.: Global isoprene and monoterpene emissions under changing climate, vegetation, CO2 and land use, Atmospheric Environment, 155, 35–45, 2017.

Haxeltine, A. and Prentice, I. C.: BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types, Global Biogeochemical Cycles, 10, 693–709, 1996.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S.: Global flood risk under climate change, Nature Climate Change, pp. 1–6, 2013.

Hof, C., Voskamp, A., Biber, M. F., Böhning-Gaese, K., Engelhardt, E. K., Niamir, A., Willis, S. G., and Hickler, T.: Bioenergy cropland expansion may offset positive effects of climate change mitigation for global vertebrate diversity, Proceedings of the National Academy of Sciences of the United States of America, 10, 201807 745–6, 2018.

Humpenöder, F., Popp, A., Dietrich, J. P., Klein, D., Lotze-Campen, H., Bonsch, M., Bodirsky, B. L., Weindl, I., Stevanovic, M., and Müller, C.: Investigating afforestation and bioenergy CCS as climate change mitigation strategies, pp. 1–14, 2014.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos, A., Jones, C. D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D. P., and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, Climatic Change, 109, 117–161, 2011.

IIASA: SSP Database (version 0.93), International Institute for Applied Systems Analysis, 2014.

IPBES: Summary for Policymakers of the Regional Assessment Report on Biodiversity and Ecosystem Services for Europe and Central Asia, Tech. rep., Bonn, Germany, 2018a.

IPBES: Summary for policymakers of the regional assessment report on biodiversity and ecosystem services for the Americas of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Tech. rep., BONN, GERMANY, 2018b.

Ito, A., Nishina, K., Reyer, C. P. O., François, L., Henrot, A.-J., Munhoven, G., Jacquemin, I., Tian, H., Yang, J., Pan, S., Morfopoulos, C., Betts, R. A., Hickler, T., Steinkamp, J., Ostberg, S., Schaphoff, S., Ciais, P., Chang, J., Rafique, R., Zeng, N., and Zhao, F.: Photosynthetic productivity and its efficiencies in ISIMIP2a biome models: benchmarking for impact assessment studies, Environmental Research Letters, 12, 085 001–14, 2017.

Jantz, S. M., Barker, B., Brooks, T. M., Chini, L. P., Huang, Q., Moore, R. M., Noel, J., and Hurtt, G. C.: Future habitat loss and extinctions driven by land-use change in biodiversity hotspots under four scenarios of climate-change mitigation, Conservation Biology, 29, 1122–1131, 2015.

Kaldor, N.: The irrelevance of equilibrium economics, The Economic Journal, 82, 1237–1255, 1972.

KC, S. and Lutz, W.: The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100, Global Environmental Change-Human and Policy Dimensions, 42, 181–192, 2017.

Kesselmeier, J. and Staudt, M.: Biogenic Volatile Organic Compounds (VOC): An Overview on Emission, Physiology and Ecology, Journal of Atmospheric Chemistry, 33, 23–88, 1999.

Keyzer, M. A., Merbis, M. D., Pavel, I. F. P. W., and van Wesenbeeck, C. F. A.: Diet shifts towards meat and the effects on cereal use: can we feed the animals in 2030?, Ecological Economics, 55, 187–202, 2005.

Koirala, S., Hirabayashi, Y., Mahendran, R., and Kanae, S.: Global assessment of agreement among streamflow projections using CMIP5 model outputs, Environmental Research Letters, 9, 064 017–12, 2014.

Krause, A., Pugh, T. A. M., Bayer, A. D., Doelman, J. C., Humpenöder, F., Anthoni, P., Olin, S., Bodirsky, B. L., Popp, A., Stehfest, E., and Arneth, A.: Global consequences of afforestation and bioenergy cultivation on ecosystem service indicators, Biogeosciences, 14, 4829–4850, 2017.

Lawrence, D. M., Slater, A. G., and Swenson, S. C.: Simulation of Present-Day and Future Permafrost and Seasonally Frozen Ground Conditions in CCSM4, Journal of Climate, 25, 2207–2225, 2012.

Lawrence, P. J., Lawrence, D. M., and Hurtt, G. C.: Attributing the Carbon Cycle Impacts of CMIP5 Historical and Future Land Use and Land Cover Change in the Community Earth System Model (CESM1), Journal of Geophysical Research: Biogeosciences, 123, 1732–1755, 2018.

Lindeskog, M., Arneth, A., Bondeau, A., Waha, K., Seaquist, J., Olin, S., and Smith, B.: Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa, Earth System Dynamics, 4, 385–407, 2013.

Mitra-Kahn, B. H.: Debunking the Myths of Computable General Equilibrium Models, Tech. rep., 2008.

Mittermeier, R. A., Gil, P. R., Hoffmann, M., Pilgrim, J., Brooks, T. M., Mittermeier, C. G., Lamoreux, J., and da Fonseca, G. A. B.: Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions: Conservation International, Conservation International, 2004.

Molotoks, A., Stehfest, E., Doelman, J., Albanito, F., Fitton, N., Dawson, T. P., and Smith, P.: Global projections of future cropland expansion to 2050 and direct impacts on biodiversity and carbon storage, Global Change Biology, 24, 5895–5908, 2018.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., and Kent, J.: Biodiversity hotspots for conservation priorities, Nature, 403, 853–858, 2000.

Myers, S. S., Gaffikin, L., Golden, C. D., Ostfeld, R. S., Redford, K. H., Ricketts, T. H., Turner, W. R., and Osofsky, S. A.: Human health impacts of ecosystem alteration., Proceedings of the National Academy of Sciences, 110, 18 753–18 760, 2013.

Myhre, G., Shindell, D. T., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., and Zhang, H.: 8: Anthropogenic and natural radiative forcing, pp. 659–740, Cambridge, United Kingdom and New York, NY, USA, 2013.

Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Lysenko, I., Senior, R. A., Börger, L., Bennett, D. J., Choimes, A., Collen, B., Day, J., De Palma, A., Díaz, S., Echeverria-Londoño, S., Edgar, M. J., Feldman, A., Garon, M., Harrison, M. L. K., Alhusseini, T., Ingram, D. J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Correia, D. L. P., Martin, C. D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H. R. P., Purves, D. W., Robinson, A., Simpson, J., Tuck, S. L., Weiher, E., White, H. J., Ewers, R. M., Mace, G. M., Scharlemann, J. P. W., and Purvis, A.: Global effects of land use on local terrestrial biodiversity, Nature, 520, 45–50, 2015.

Olin, S., Lindeskog, M., Pugh, T. A. M., Schurgers, G., Wårlind, D., Mishurov, M., Zaehle, S., Stocker, B. D., Smith, B., and Arneth, A.: Soil carbon management in large-scale Earth system modelling: implications for crop yields and nitrogen leaching, Earth System Dynamics, 6, 745–768, 2015a.

Olin, S., Schurgers, G., Lindeskog, M., Wårlind, D., Smith, B., Bodin, P., Holmér, J., and Arneth, A.: Modelling the response of yields and tissue C : N to changes in atmospheric CO_2 and N management in the main wheat regions of western Europe, Biogeosciences, 12, 2489–2515, 2015b.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., DAmico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., and Kassem, K. R.: Terrestrial Ecoregions of the World: A New Map of Life on Earth, BioScience, 51, 933–938, 2001.

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., and van Vuuren, D. P.: A new scenario framework for climate change research: the concept of shared socioeconomic pathways, Climatic Change, 122, 387–400, 2014.

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., and Solecki, W.: The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century, Global Environmental Change-Human and Policy Dimensions, pp. 1–12, 2015.

Pastick, N. J., Jorgenson, M. T., Wylie, B. K., Nield, S. J., Johnson, K. D., and Finley, A. O.: Distribution of near-surface permafrost in Alaska: Estimates of present and future conditions, Remote Sensing of Environment, 168, 301–315, 2015.

Popp, A., Calvin, K., Fujimori, S., Havlik, P., Humpenöder, F., Stehfest, E., Bodirsky, B. L., Dietrich, J. P., Doelmann, J. C., Gusti, M., Hasegawa, T., Kyle, P., Obersteiner, M., Tabeau, A., Takahashi, K., Valin, H., Waldhoff, S., Weindl, I., Wise, M., Kriegler, E., Lotze-Campen, H., Fricko, O., Riahi, K., and van Vuuren, D. P.: Land-use futures in the shared socio-economic pathways, Global Environmental Change-Human and Policy Dimensions, 42, 331–345, 2017.

Portmann, F. T., Siebert, S., and Döll, P.: MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, Global Biogeochemical Cycles, 24, n/a–n/a, 2010.

Quesada, B., Arneth, A., and de Noblet-Ducoudré, N.: Atmospheric, radiative, and hydrologic effects of future land use and land cover changes: A global and multimodel climate picture, Journal of Geophysical Research: Atmospheres, 122, 5113–5131, 2017.

Rabin, S. S.: Harmonizing LandSyMM with historical data, p. 3, 2019.

Rap, A., Scott, C. E., Reddington, C. L., Mercado, L., Ellis, R. J., Garraway, S., Evans, M. J., Beerling, D. J., MacKenzie, A. R., Hewitt, C. N., and Spracklen, D. V.: Enhanced global primary production by biogenic aerosol via diffuse radiation fertilization, Nature Publishing Group, pp. 1–7, 2018.

Rasulov, B., Talts, E., and Niinemets, Ü.: Spectacular Oscillations in Plant Isoprene Emission under Transient Conditions Explain the Enigmatic CO2 Response., PLANT PHYSIOLOGY, 172, 2275–2285, 2016.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schnellnhuber, H. J., Nykvist, B., de Wit, C. A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P. K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R. W., Fabry, V. J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., and Foley, J. A.: A safe operating space for humanity, Nature, 461, 2009.

Rogelj, J., Shindell, D. T., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshagi, H., Kobayashi, S., Kriegler, E., Mundaca, L., Séférian, R., and Vilariño, M. V.: 2: Mitigation pathways compatible with 1.5ºC in the context of sustainable development, in: Global warming of 1.5ºC: An IPCC Special Report on the impacts of global warming of 1.5ºC above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, edited by Masson-Delmotte, V., Zhai, P., Pörtner, H. O., ROBERTS, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J. B. R., Chen, Y., Zhou, X., Gomis, M. I., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T., pp. 93–174, Cambridge, United Kingdom and New York, NY, USA, 2018.

Rosenkranz, M., Pugh, T. A. M., Schnitzler, J.-P., and Arneth, A.: Effect of land-use change and management on biogenic volatile organic compound emissions - selecting climate-smart cultivars, Plant, Cell and Environment, 38, 1896–1912, 2014.

Schurgers, G., Arneth, A., Holzinger, R., and Goldstein, A. H.: Process-based modelling of biogenic monoterpene emissions combining production and release from storage, Atmospheric Chemistry and Physics, 9, 3409–3423, 2009.

Shcherbak, I., Millar, N., and Robertson, G. P.: Global metaanalysis of the nonlinear response of soil nitrous oxide (N2O) emissions to fertilizer nitrogen., Proceedings of the National Academy of Sciences, 111, 9199–9204, 2014.

Simpson, R. D., Sedjo, R. A., and Reid, J. W.: Valuing Biodiversity for Use in Pharmaceutical Research, Journal of Political Economy, 104, 163–185, 1996.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Müller, J. F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years, Atmospheric Chemistry and Physics, 14, 9317–9341, 2014.

Smith, B., Prentice, I. C., and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space, Global Ecology and Biogeography, 10, 621–637, 2001.

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, Biogeosciences, 11, 2027–2054, 2014.

Sporre, M. K., Blichner, S. M., Karset, I. H. H., Makkonen, R., and Berntsen, T. K.: BVOC–aerosol–climate feedbacks investigated using NorESM, Atmospheric Chemistry and Physics, 19, 4763–4782, 2019.

Szogs, S., Arneth, A., Anthoni, P., Doelman, J. C., Humpenöder, F., Popp, A., Pugh, T. A. M., and Stehfest, E.: Impact of LULCC on the emission of BVOCs during the 21st century, Atmospheric Environment, 165, 73–87, 2017.

Tang, Q. and Lettenmaier, D. P.: 21st century runoff sensitivities of major global river basins, Geophysical Research Letters, 39, n/a–n/a, 2012.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, 2012.

Tilman, D., Balzer, C., Hill, J., and Befort, B. L.: Global food demand and the sustainable intensification of agriculture, Proceedings of the National Academy of Sciences, 108, 20 260–20 264, 2011.

Tilman, D., Isbell, F., and Cowles, J. M.: Biodiversity and Ecosystem Functioning, Annual Review of Ecology, Evolution, and Systematics, 45, 471–493, 2014.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K.: The representative concentration pathways: an overview, Climatic Change, 109, 5–31, 2011.

van Vuuren, D. P., Stehfest, E., Gernaat, D. E. H. J., Berg, M., Bijl, D. L., Boer, H. S., Daioglou, V., Doelman, J. C., Edelenbosch, O. Y., Harmsen, M., Hof, A. F., and Sluisveld, M. A. E.: Alternative pathways to the 1.5 °C target reduce the need for negative emission technologies, Nature Publishing Group, pp. 1–10, 2018.

Verhagen, W., van der Zanden, E. H., Strauch, M., van Teeffelen, A. J. A., and Verburg, P. H.: Optimizing the allocation of agri-environment measures to navigate the trade-offs between ecosystem services, biodiversity and agricultural production, Environmental Science and Policy, 84, 186–196, 2018.

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., Schlesinger, W. H., and Tilman, D. G.: Human Alteration of the Global Nitrogen Cycle: Sources and Consequences, Ecological Applications, 7, 737–750, 1997.

Weindl, I., Popp, A., Bodirsky, B. L., Rolinski, S., Lotze-Campen, H., Biewald, A., Humpenöder, F., Dietrich, J. P., and Stevanovic, M.: Livestock and human use of land: Productivity trends and dietary choices as drivers of future land and carbon dynamics, Global and Planetary Change, 159, 1–10, 2017.

Wheater, H. and Evans, E.: Land use, water management and future flood risk, Land Use Policy, 26, S251–S264, 2009.

Wilhite, D. A. and Glantz, M. H.: Understanding: the drought phenomenon: the role of definitions, Water International, 10, 111–120, 1985.

Yang, W. and Omaye, S. T.: Air pollutants, oxidative stress and human health, Mutation Research/Genetic Toxicology and Environmental Mutagenesis, 674, 45–54, 2009.

Young, P. J., Arneth, A., Schurgers, G., Zeng, G., and Pyle, J. A.: The CO_2 inhibition of terrestrial isoprene emission significantly affects future ozone projections, Atmospheric Chemistry and Physics, 9, 2793–2803, 2009.

Zhang, B., Tian, H., Lu, C., Dangal, S. R. S., Yang, J., and Pan, S.: Global manure nitrogen production and application in cropland during 1860–2014: a 5 arcmin gridded global dataset for Earth system modeling, Earth System Science Data, 9, 667–678, 2017.

Zhu, Q., Jiang, H., Peng, C., Liu, J., Fang, X., Wei, X., Liu, S., and Zhou, G.: Effects of future climate change, CO2 enrichment, and vegetation structure variation on hydrological processes in China, Global and Planetary Change, 80-81, 123–135, 2012.

Impacts of future agricultural change on ecosystem service indicators

Supplementary Methods

SM1 ~~Run~~ LPJ-GUESS run setup

~~The first run performed in this study generated the calibration factors described in Sect.SM2. In the calibration run for~~ This section provides further details and rationale for the types of model runs used in LandSyMM (summarized in Table SM1). Figure SM1 gives an overview of how information flows between the different runs. Tables SM3–SM6 describe the input data used for each run type.

LPJ-GUESS simulates only four crop functional types (CFTs): spring wheat, winter wheat, maize, and rice. These must be translated to the seven PLUM crop types, as described in Section SM2. This process uses crop-specific values called calibration factors. In this study, the "calibration" was used in generating these calibration factors, with mostly the same methodology detailed in Alexander et al. (2018). However, in that previous work, only four crop stand types were simulated in LPJ-GUESS, corresponding to the four LPJ-GUESS CFTs (Alexander et al., 2018). This meant that starchy roots, oilcrops, and pulses all received the same management inputs (i.e., fertilizer and irrigation). In the work presented here, we have separated these three into distinct stand types that all use the TeSW CFT but with different management inputs based on crop-specific historical datasets (Table SM8). This change results in different calibration factors being used here than in Alexander et al. (2018); the new calibration factors can be found in Fig. SM3.

The calibration run was forced with "CRU-NCEP v7 CRUp" climate data (Table SM3), which use forcings from CRU-NCEP version 7 (Le Quéré et al., 2016; Viovy, 2016, [1]) except with CRU TS3.24 precipitation (Harris et al., 2014) due to problems discovered in the CRU-NCEP precipitation data.[2] The MIRCA2000 dataset (Portmann et al., 2010) provided crop type distributions for the year 2000, which were used for all historical years in the calibration run. Some mapping between MIRCA, LPJ-GUESS, and PLUM crop types was required, details of which can be found in Section SM2. Fertilizer application for the calibration run was taken from the dataset prepared for the Global Gridded Crop Model Intercomparison exercise (Elliott et al., 2015) of the Agricultural Model Intercomparison Project (AgMIP; Rosenzweig et al., 2013). In the historical period of other runs, nitrogen fertilizer application rates were taken from LUH2 (Hurtt et al., in prep). 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. Simulation years outside the dataset's 1860–2014 range used 1860 and 2014 values, respectively.

A set of "yield-generating" experiments were then performed to produce potential crop yields and pasture grass production for input to PLUM. These consist of two phases: an initial and an alternating phase. The initial phase runs from 1850 to 2000, and is intended to reproduce historical land uses and crop yields in a way that is consistent with previously developed land-use histories. Historical land use areas, irrigation, and synthetic nitrogen fertilizer application levels were taken from the Land Use Harmonization v2 dataset (LUH2; Hurtt et al., in prep.). Historical manure application rates (simplified upon import to LPJ-GUESS as pure nitrogen addition) come from Zhang et al. (2017b).

The alternating phase, which begins in 2001, iterates between "potential runs" and "actual runs," the latter of which exist only to provide initial soil conditions to the potential runs (Fig. SM1). In the potential runs, the non-barren land in every gridcell is converted to 50% cropland and 50% pasture, with homogenized soil based on the state after the previous actual run (Fig. SM1; Table SM5). Cropland is subdivided into 36 equally-sized stands in a factorial experiment with the six crop stand types (excluding Miscanthus), three nitrogen fertilizer treatments (0, 200, and 1000 kgN ha^{-1} yr^{-1}, and two irrigation treatments (rainfed or fully irrigated). Potential runs begin every five years, with each lasting ten years. Only the last five years' yields are passed to PLUM, with the first five years being used to give LPJ-GUESS time to spin up crop phenological parameters. Thus, for example, the potential run covering 2006–2015 generates output for 2011–2015, which is used in PLUM to determine land uses and managements for 2016–2020 (Fig. SM1). Actual runs each last five years, with the land system state being saved after each for input to the po-

[1] http://dods.extra.cea.fr/store/p529viov/cruncep/V7_1901_2015, accessed 30 June 2016

[2] 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.

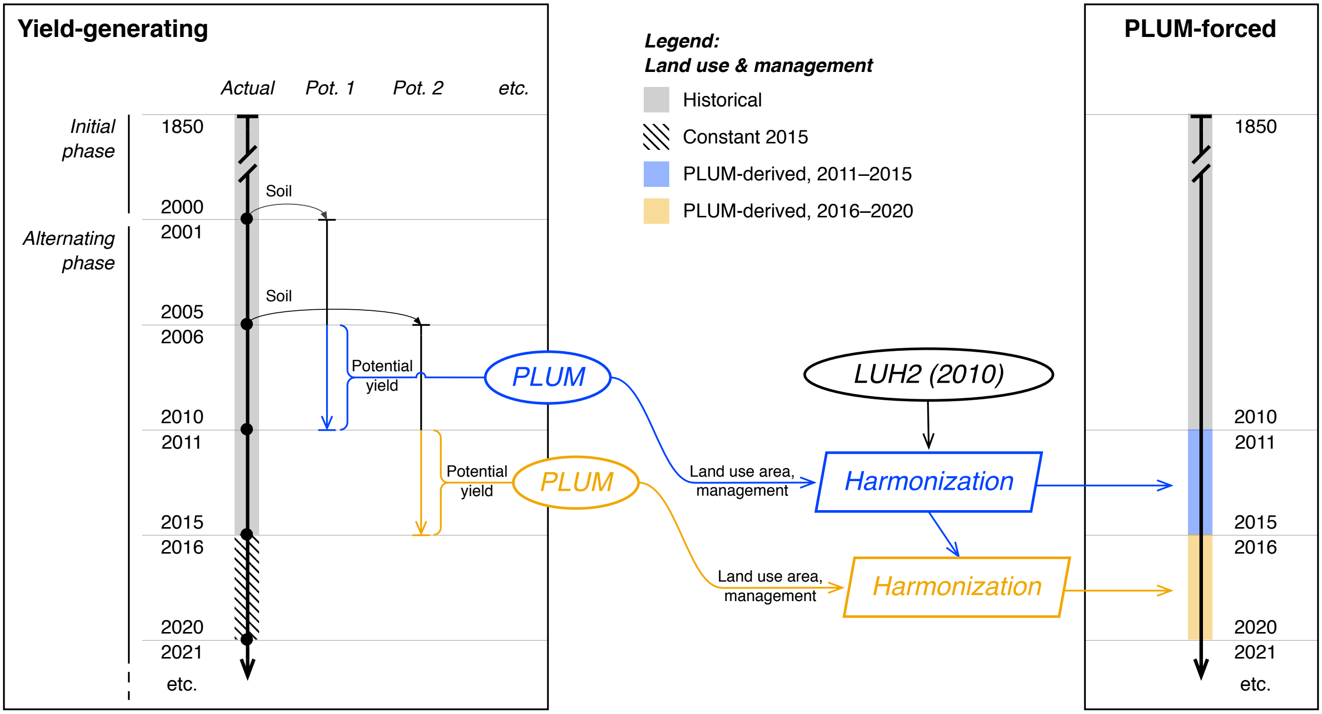


Figure SM1. Information flow between LPJ-GUESS and PLUM. Historical land use and management is time-varying for land use fractions through 2015 but constant for crop mix, fertilizer application, and per-crop irrigated fraction (see Sect. 3.3.1). Adapted from Figure SI-1 in Alexander et al. (2018).

tential runs. Land use and management in the actual runs uses LUH2 data through 2015, after which values are held constant (Table SM4).

All "yield-generating" runs used the atmospheric CO_2 concentrations and climate forcings from the Fifth Coupled Model Intercomparison Project (CMIP5; Taylor et al., 2012): specifically, the IPSL-CM5A-MR forcings (Dufresne et al., 2013), which were bias-corrected (Ahlström et al., 2012) to the 1961–1990 observation-based climate used by the calibration runs (Tables SM4, SM5). Because not all SSP-RCP combinations are equally plausible, the PLUM-forced runs used future climate forcings corresponding to the most likely RCP for each SSP, based on the SSP-RCP probability matrix from Engström et al. (2016): i.e., RCP4.5 for SSP1, RCP6.0 for SSP3 and SSP4, and RCP8.5 for SSP5. Three instances of the yield-generating runs were performed: one with each of the RCP climate scenarios.

Finally, the "PLUM-forced" runs combined the PLUM outputs of land use areas and management (harmonized as described in the main text, Sect. XXX) with the same climate and atmospheric CO_2 concentrations used in the yield-generating runs.

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.

SM2 Mapping of MIRCA, LPJ-GUESS, and PLUM crop types

MIRCA crop types are mapped to the LPJ-GUESS crop stand types as described in Table SM8. Some MIRCA categories were excluded: sugarcane, citrus, date palm, grapes/vine, cotton, cocoa, coffee, other annuals, other perennials, and fodder grasses. A dummy crop type, ExtraCrop, was created to par-

Table SM1. Guide to the runs performed.

Run name	Years	Purpose	Number
Calibration	1901–2005	Simulate 1995–2005 crop yields for calibrating against FAOSTAT data.	1
Yield-generating, actual	1850–2100	1850–2010: Historical simulation for comparison with PLUM-forced future runs. All: Provide soil state for yield-generating potential runs.	1850–2010: 1 2011–2100: 4 (SSP-RCP combinations)
Yield-generating, potential	1850–2100	Generate potential yield for each crop type under different fertilization and irrigation treatments, for use in PLUM.	1850–2010: 1 2011–2100: 4 (SSP-RCP combinations)
PLUM-forced	2011–2100	Simulate terrestrial vegetation and ecosystem service indicators under land-use and -management trajectories specified by PLUM.	~~17: (standard, constant-CO_2 + climate)*(4 PLUM outputs) + constant-LU*(3 RCPs) + onlyCO_2*3 RCPs + onlyClimate*3 RCPs~~ 25: `sXlum_rYYclico2`, `sXlum`*(4 PLUM outputs), (`rYYclico2` + `rYYco2` + `rYYcli`)*(3 RCPs), (`sXlum_rYYco2` + `sXlum_rYYcli`)*(4 SSP-RCP combinations)

tition this unmapped fraction of cropland away from cropland considered in PLUM. Additionally, 10.3% of mapped crop types were moved to ExtraCrop, corresponding to the cropland fraction not harvested due, e.g., to crop failures or fallow periods (FAOSTAT, 2018c, b). In all, approximately 38% of cropland was in ExtraCrop over 2001–2010. LPJ-GUESS plants ExtraCrop, which receives no irrigation or fertilizer, with either winter wheat (TeWW) or spring wheat (TeSW) based on sowing constraints derived from long-term climate history in each gridcell. By not applying management inputs to ExtraCrop, we likely underestimate the effects of future land use and management changes on water use and nitrogen losses in absolute terms; however, this allows us to focus solely on the ecosystem services impacts of the crops explicitly included in PLUM.

To generate yields of a crop not included in LPJ-GUESS (e.g., oilcrops), a separate stand is simulated and planted with spring wheat, and a calibration factor then later multiplied onto the resulting yields to generate a wider range of crop yields as input to PLUM (Table SM8). (For consistency, calibration factors are also used for crops such as rice that are included in both LPJ-GUESS and PLUM.) These calibration factors are derived from a comparison of the LPJ-GUESS simulated yields with the crop yields reported in the FAOSTAT database (FAOSTAT, 2018c, a). A historical "calibration run" from 1901–2005 was used to generate gridded yields for 1995–2005; these were aggregated to the country level, with simulated country-year data points being regressed (with Y-intercept set to zero) against the values derived from FAOSTAT.

SM3 Harmonization of future and historical land use data

~~Information flow between LPJ-GUESS and PLUM. Historical land use and management is time-varying for land use fractions through 2015 but constant for crop mix, fertilizer application, and per-crop irrigated fraction (see Sect.3.3.1). Adapted from Figure SI-1 in .~~ The harmonization code (Rabin, 2019) is based on the code published for LUH1 (Hurtt et al., 2011, http://luh.umd.edu/code.shtml), but extended to harmonize the area of pasture and *each crop* rather than just pasture and cropland. It begins with land use from LUH2 in 2010, with cropland subdivided based on MIRCA as described in Section SM1, then attempts to apply changes (deltas) in land use area between PLUM's 2010 and 2011 outputs.

Grid cells can reach limits: The deltas might specify loss of cropland when the grid cell is already 0% cropland, or likewise the deltas might specify expansion of cropland when it's already 100% cropland. In such cases, the algorithm looks for space to apply the remaining "unmet" deltas in the grid cells bordering the cell in question. It expands the radius of this search until all deltas are satisfied. This process is repeated for every cell with unmet deltas. Once complete, the algorithm moves on to the next year: The PLUM changes for 2011–2012 are applied to the harmonized 2011 land use map, and so on.

Our algorithm has another feature not present in the LUH1 harmonization, which is to harmonize fertilizer and irrigation (the latter in arbitrary units of intensity*area). The way it does this is analogous to how it treats land use area. Limits for irrigation in a cell are 0 and 1. The lower limit for fertilizer for any given crop is also zero, but the upper limit varies. It is either the maximum seen for the crop in any gridcell in any PLUM output thus far (i.e., if we're working on deltas for 2020–2021, consider PLUM outputs from 2021 and before), or in LUH2 during or before the base year (here, 2010), or in any harmonized output thus far (although, because of the first two rules, this should never come into play).

Technical details on other differences between the LUH1 method and our algorithm, as well as our harmonization code itself, are available on GitHub (Rabin, 2019).

While harmonization preserves global area totals, our implementation greatly increases the area of land experiencing land use change (Fig. SM2). This behavior stems from our version working on area of specific crops, not just (as the LUH1 harmonization did) on total cropland area. If we collapse all crops into one type, gross land use changes are barely affected by harmonization (Table SM2).

Why does this happen? Consider a gridcell that, in 2011, PLUM tells to lose 100 km^2 of rice. If not considering individual crops, PLUM's land use map for 2010 would only have had to match the historical data in terms of total cropland area. But in our implementation, they need to agree on *rice* area specifically. Per-crop harmonization thus increases the probability of disagreement between PLUM and the historical data for land use area in 2010, thereby increasing the probability that the harmonization needs to make up the difference in another gridcell.

Although our harmonization procedure is based on a well-regarded algorithm, the discrepancies introduced by considering separate crop types complicate interpretation of results. In a gridcell that experiences land-use change only after harmonization, impacts on ecosystem services may have been displaced from where PLUM anticipated that land-use change to occur. The extent to which overall regional and global scale ecosystem service provision is affected by the disagreement between the historical data and PLUM in 2010 will depend on whether the land-use change is coming from a cell of similar biome, land-use history, etc. Increasing the amount of land experiencing land-use transitions also decreases mean ecosystem age, with potential implications for carbon storage and ecosystem function.

Where necessary, in our results we note geographic patterns that may have been more a result of the harmonization process than PLUM-specified land-use change. However, reducing the impacts of harmonization on land use area and management maps would allow a more direct interpretation, with fewer confounding factors, of how changing demand, climate, and technology affect ecosystem services. Given the inherent complexities of modelling human societies, harmonization is likely to remain a fundamental component of such investigations for the foreseeable future; as such, further development of our algorithm will aim to minimize harmonization effects.

Table SM2. Harmonization-induced change in 2010–2100 gross gain and loss of non-agricultural area.

Scenario	As described Δ gain	As described Δ loss	Combined crops Δ gain	Combined crops Δ loss
SSP1-45	+275%	+124%	+7%	+3%
SSP3-60	+25%	+168%	+2%	< +0.5%
SSP4-60	+159%	+39%	–1%	< –0.5%
SSP5-85	+126%	+36%	+1%	< +0.5%

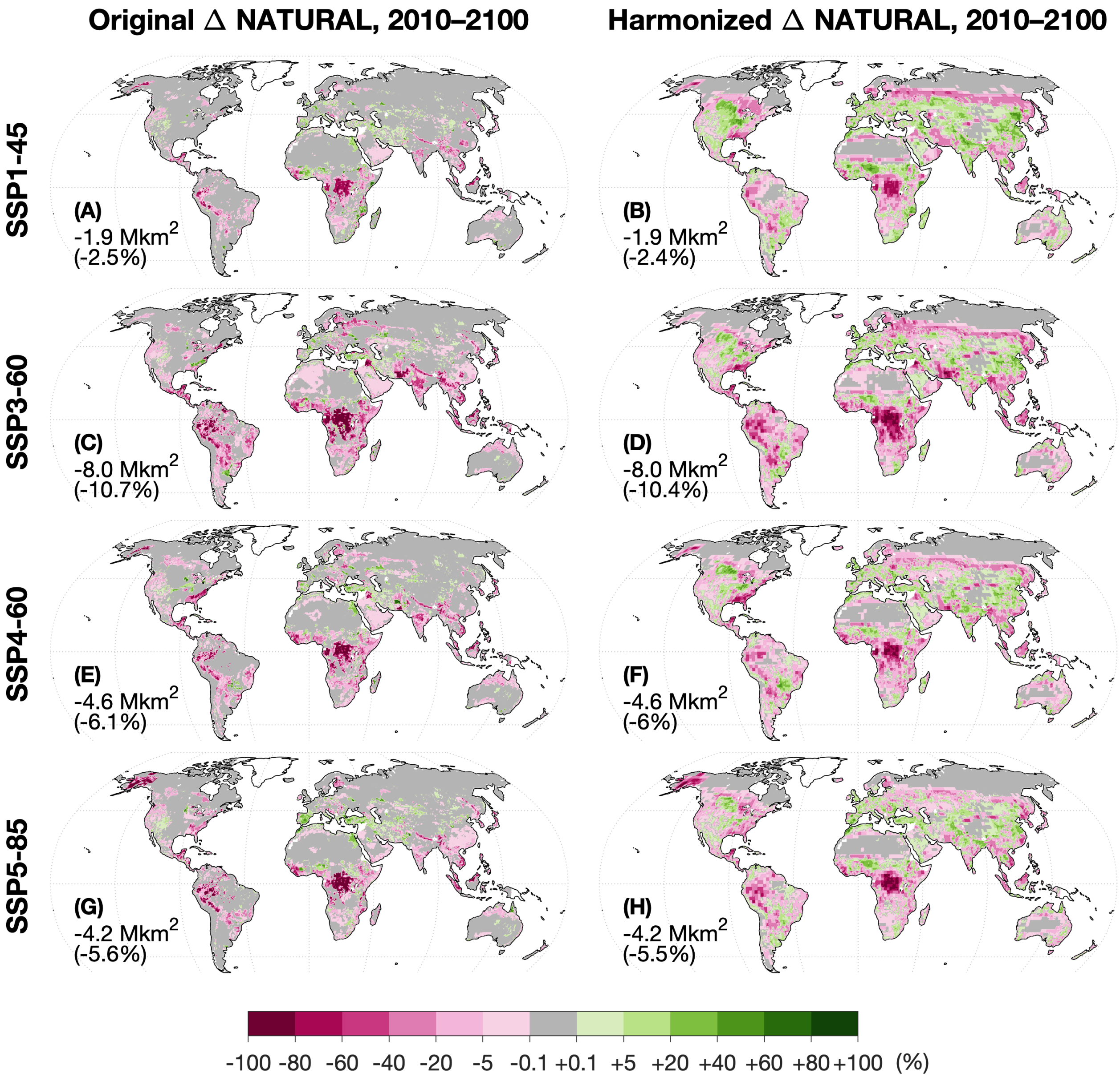


Figure SM2. Change in non-agricultural land area (as fraction of gridcell) over 2010–2100 from raw PLUM outputs (left column) and after harmonization (right column) for each scenario (rows). Inconsistencies in relative net global change (percentages) are due to different baseline (2010) land use area maps between the two columns.

Table SM3. Input data used for each portion of calibration run. Zhang et al. (2017a) manure data extended as specified in Methods.

Years	Climate	LU	Crop fractions	Fert.	Irrig.
1850–1900	–	–	–	–	–
1901–2005	CRU-NCEP v7 CRUp	LUH2	MIRCA @2000	AgMIP @2000 + Zhang et al. @2000	MIRCA @2000*
2006–2010	–	–	–	–	–
2011–2015	–	–	–	–	–
2016–2100	–	–	–	–	–

–: Time period not simulated in given run.
DATASET @YYYY: Using value from DATASET at year YYYY.
*: Irrigation specified by fraction of crop fully rainfed or fully irrigated.

Table SM4. Input data used for each portion of "actual" yield-generating runs. Zhang et al. (2017a) manure data extended as specified in Methods.

Years	Climate	LU	Crop fractions	Fert.	Irrig.
1850–1900	CMIP5 IPSL-CM5A-MR	LUH2	MIRCA @2000	LUH2 + Zhang et al.	LUH2*
1901–2005	↓	↓	↓	↓	↓
2006–2010	↓	↓	↓	↓	↓
2011–2015	↓	↓	↓	↓	↓
2016–2100	↓	LUH2 @2015	↓	LUH2 @2015 + Zhang et al. @2015	↓

DATASET @YYYY: Using value from DATASET at year YYYY.
*: Irrigation specified by fraction of crop fully rainfed or fully irrigated.

Table SM5. Input data used for each portion of "potential" yield-generating runs.

Years	Climate	LU	Crop fractions	Fert.	Irrig.
1850–1900	–	–	–	–	–
1901–2005	–	–	–	–	–
2006–2010	CMIP5 IPSL-CM5A-MR	Ice/water from LUH2; vegetated 50-50 cropland and pasture	Even crop × fertilizer × irrigation factorial stands	0, 200, or 1000 kgN ha^{-1}	Rainfed or fully irrigated
2011–2015	↓	↓	↓	↓	↓
2016–2100	↓	↓	↓	↓	↓

–: Time period not simulated in given run.

Table SM6. Input data used for each portion of PLUM-forced runs. Zhang et al. (2017a) manure data extended as specified in Methods.

Years	Climate	LU	Crop fractions	Fert.	Irrig.
1850–1900	CMIP5 IPSL-CM5A-MR	LUH2	MIRCA @2000	LUH2 + Zhang et al.	MIRCA @2000*
1901–2005	↓	↓	↓	↓	↓
2006–2010	↓	↓	↓	↓	↓
2011–2015	↓	PLUM	PLUM	PLUM	PLUM†
2016–2100	↓	PLUM	PLUM	PLUM	PLUM†

DATASET @YYYY: Using value from DATASET at year YYYY.
*: Irrigation specified by fraction of crop fully rainfed or fully irrigated.
†: irrigation specified as fraction of maximum irrigation demand fulfilled.

Table SM7. Parameters used in PLUM. For more information, see Alexander et al. (2018) main text and supplement.

Parameter	SSP1	SSP3	SSP4	SSP5
Irrigation cost, w_{cost} (USD m^{-2})	0.000440	0.000232	0.000350	0.000232
Fertilizer cost, f_{cost} (USD t^{-1})	2.2	1.5	1.8	1.1
Other intensity cost, m_{cost} (USD at max management input)	0.8	0.7	0.7	0.6
Land cover change cost, lc_{change}: Natural to agricultural (USD ha^{-1})	107	31	54	38
Land cover change cost, lc_{change}: Pasture to cropland (USD ha^{-1})	290	205	232	161
Land cover change cost, lc_{change}: Cropland to pasture (USD ha^{-1})	575	366	432	300
Minimum natural or managed forest cover	19.5%	4.5%	6.2%	4.5%
Technology yield change rate, δ, above that from intensification of production	0.44%	0.00%	0.20%	0.30%
International market price sensitivity, λ	0.4	0.4	0.4	0.4
International import tariff, i_{tariff}	–20%	43%	19%	–36%
Transport costs, t_{cost} (USD t^{-1})	63	43	57	37
Annual change in imports allowed	2.2%	1.4%	1.7%	2.6%

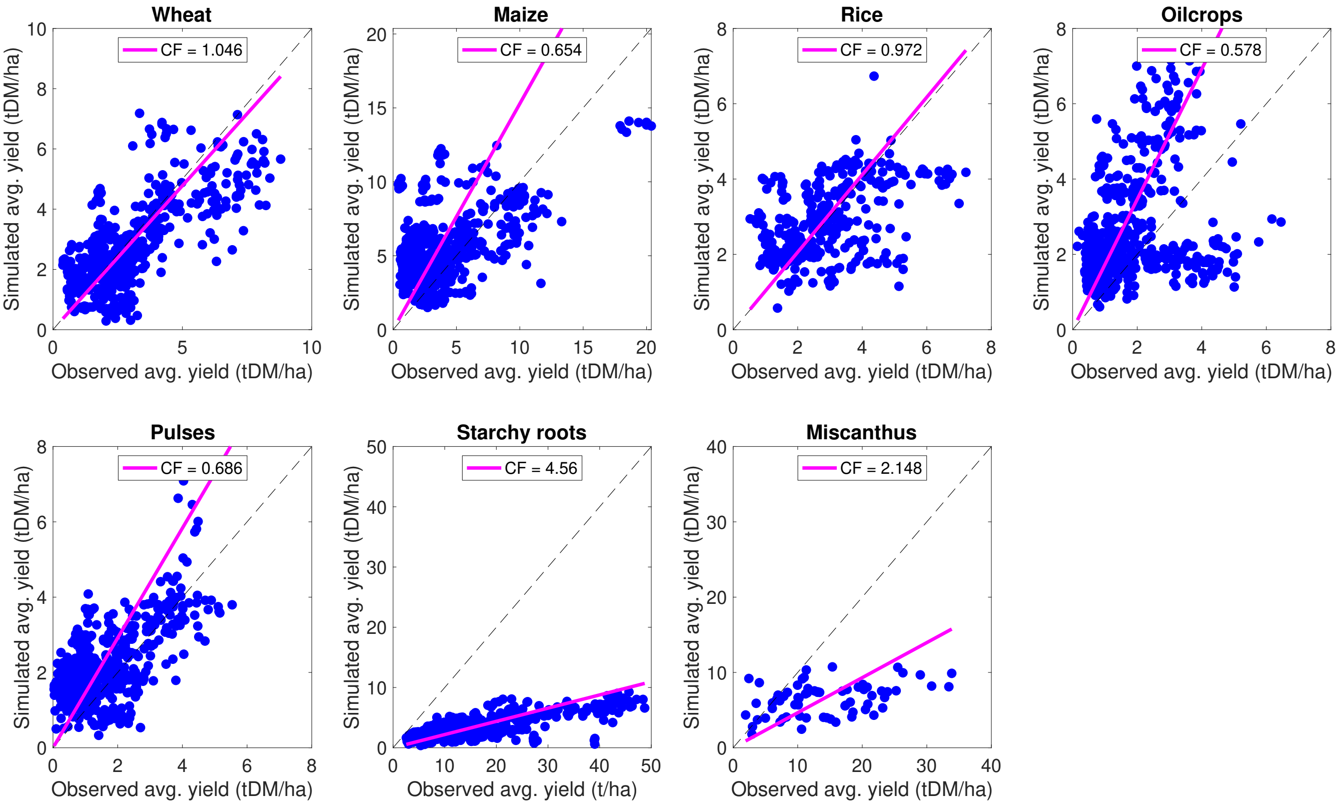


Figure SM3. Scatter plots between observed and LPJ-GUESS yield , with regression line used to determine calibration factors. Each point represents one country's yield in a single year. Corresponds to Fig. SI-2 in Alexander et al. (2018).

Table SM8. Mapping between different crop types used in this study. The first column gives the crop types used by PLUM, which are also the stand types simulated by LPJ-GUESS. The second column shows which CFTs LPJ-GUESS simulates in those stands, and the calibration factor for adjusting the yield for use in PLUM. Table adapted from Alexander et al. (2018).

Stand type = PLUM crop	CFT (calibration factor)	MIRCA (crop fraction) or AgMIP (calibration-run fertilizer) crop types	LUH2 crops (actual-run fertilizer and irrigated fraction)	FAO crops (calibration area; from FAOSTAT's "Production: Crops")	FAO crops (calibration production tonnage; from FAOSTAT's "Commodity Balance: Crops primary equivalent")
CerealsC3	TeWW or TeSW (1.056)	Wheat, Barley, Rye	All: C3ann	Wheat, Barley, Rye	Wheat and products, Barley and products, Rye and products
CerealsC4	TeCo (0.738)	Maize, Millet, Sorghum	All: C4ann	Maize, Millet, Sorghum	Maize and products, Millet and products, Sorghum and products
Rice	TrRi (1.052)	Rice	C3ann	Rice paddy	Rice (paddy equivalent)
Oilcrops	TeSW (0.687)	Sunflower, Soybeans, Groundnuts/ Peanuts, Rapeseed/Canola, Oilpalm	Area-weighted average of C3nfx (Soybeans, Groundnuts/Peanuts) and C3ann (others)	Coconuts, Seed cotton, Groundnuts with shell, Karite nuts (sheanuts), Castor oil seed, Tung nuts, Jojoba seed, Safflower seed, Poppy seed, Melonseed, Tallowtree seed, Kapok fruit, Linseed, Hempseed, Oilseeds nes, Olives, Oil palm fruit, Rapeseed, Mustard seed, Sesame seed, Soybeans, Sunflower seed	Oilcrops
Pulses	TeSW (0.865)	MIRCA: Pulses; AgMIP: Groundnuts, Soybeans	All: C3nfx	Pulses total	Pulses
Starchy roots	TeSW (5.443)	Potatoes, Sugarbeet, Cassava	All: C3ann	Roots and tubers total	Starchy roots
Miscanthus	TeCo (2.148)	n/a (only simulated in PLUM-forced runs)	n/a (only simulated in PLUM-forced runs)	n/a (calibrated against BetyDB data)	n/a (calibrated against BetyDB data)
ExtraCrop	n/a	MIRCA: Sugarcane, citrus, date palm, grapes/vine, cotton, cocoa, coffee, other annuals, other perennials, fodder grasses	n/a (no fertilizer or irrigation)	n/a (not calibrated)	n/a (not calibrated)

References

Ahlström, A., Schurgers, G., Arneth, A., and Smith, B.: Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections, Environmental Research Letters, 7, 044 008–10, 2012.

Alexander, P., Rabin, S. S., Anthoni, P., Henry, R., Pugh, T. A. M., Rounsevell, M. D. A., and Arneth, A.: Adaptation of global land use and management intensity to changes in climate and atmospheric carbon dioxide, Global Change Biology, 24, 2791–2809, 2018.

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J. P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J. Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M. P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Climate Dynamics, 40, 2123–2165, 2013.

Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., and Sheffield, J.: The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0), Geoscientific Model Development, 8, 261–277, 2015.

Engström, K., Olin, S., Rounsevell, M. D. A., Brogaard, S., van Vuuren, D. P., Alexander, P., Murray-Rust, D., and Arneth, A.: Assessing uncertainties in global cropland futures using a conditional probabilistic modelling framework, Earth System Dynamics, 7, 893–915, 2016.

FAOSTAT: Commodity Balances/Crops Primary Equivalent (2018-09-24), Food and Agriculture Organization of the United Nations, 2018a.

FAOSTAT: Inputs/Land Use (2018-09-24), Food and Agriculture Organization of the United Nations, 2018b.

FAOSTAT: Production/Crops (2018-09-24), Food and Agriculture Organization of the United Nations, 2018c.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, International Journal of Climatology, 34, 623–642, 2014.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos, A., Jones, C. D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D. P., and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, Climatic Change, 109, 117–161, 2011.

Le Quéré, C., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Peters, G. P., Manning, A. C., Boden, T. A., Tans, P. P., Houghton, R. A., Keeling, R. F., Alin, S., Andrews, O. D., Anthoni, P., Barbero, L., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Currie, K., Delire, C., Doney, S. C., Friedlingstein, P., Gkritzalis, T., Harris, I., Hauck, J., Haverd, V., Hoppema, M., Klein Goldewijk, K., Jain, A. K., Kato, E., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Melton, J. R., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-i., O amp apos Brien, K., Olsen, A., Omar, A. M., Ono, T., Pierrot, D., Poulter, B., Rödenbeck, C., Salisbury, J., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Sutton, A. J., Takahashi, T., Tian, H., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2016, Earth System Science Data, 8, 605–649, 2016.

Portmann, F. T., Siebert, S., and Döll, P.: MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, Global Biogeochemical Cycles, 24, n/a–n/a, 2010.

Rabin, S. S.: Harmonizing LandSyMM with historical data, p. 3, 2019.

Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P., Antle, J. M., Nelson, G. C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., and Winter, J. M.: The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies, Agricultural and Forest Meteorology, 170, 166–182, 2013.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, 2012.

Viovy, N.: CRUNCEP Version 7: Atmospheric Forcing Data for the Global Carbon Budget 2016, 2016.

Zhang, B., Tian, H., Lu, C., Dangal, S. R. S., Yang, J., and Pan, S.: Manure nitrogen production and application in cropland and rangeland during 1860–2014: A 5-minute gridded global data set for Earth system modeling, Earth System Science Data Discussions, pp. 1–35, 2017a.

Zhang, B., Tian, H., Lu, C., Dangal, S. R. S., Yang, J., and Pan, S.: Global manure nitrogen production and application in cropland during 1860–2014: a 5 arcmin gridded global dataset for Earth system modeling, Earth System Science Data, 9, 667–678, 2017b.

Impacts of future agricultural change on ecosystem service indicators

Supplementary Results

~~This page intentionally left blank.~~

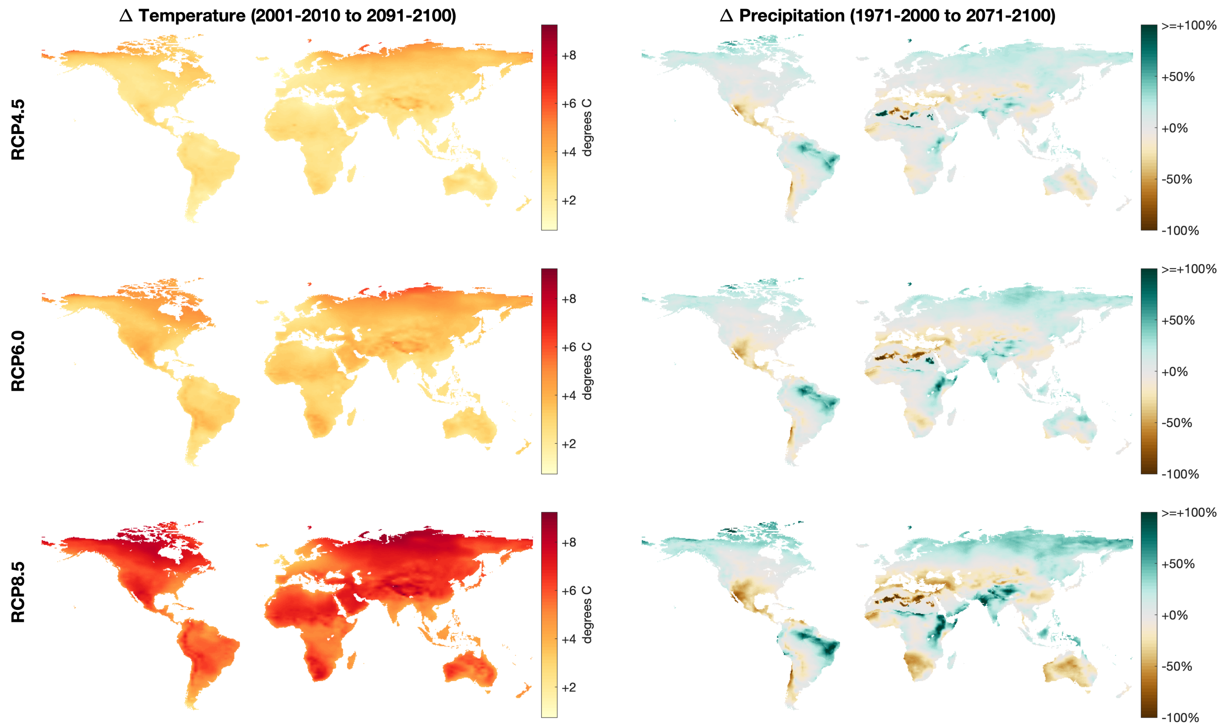


Figure SR1. Change in mean global temperature (degrees Celsius; left column) and precipitation (percent, right column) between historical and future periods for each RCP (rows). Temperature given as bias-corrected by LPJ-GUESS. 30 years used for precipitation because of high interannual variability, consistent with main text Fig. 2.

~~As Figure SR5, but for *Miscanthus*.~~

~~As Figure SR5, but for Rice.~~

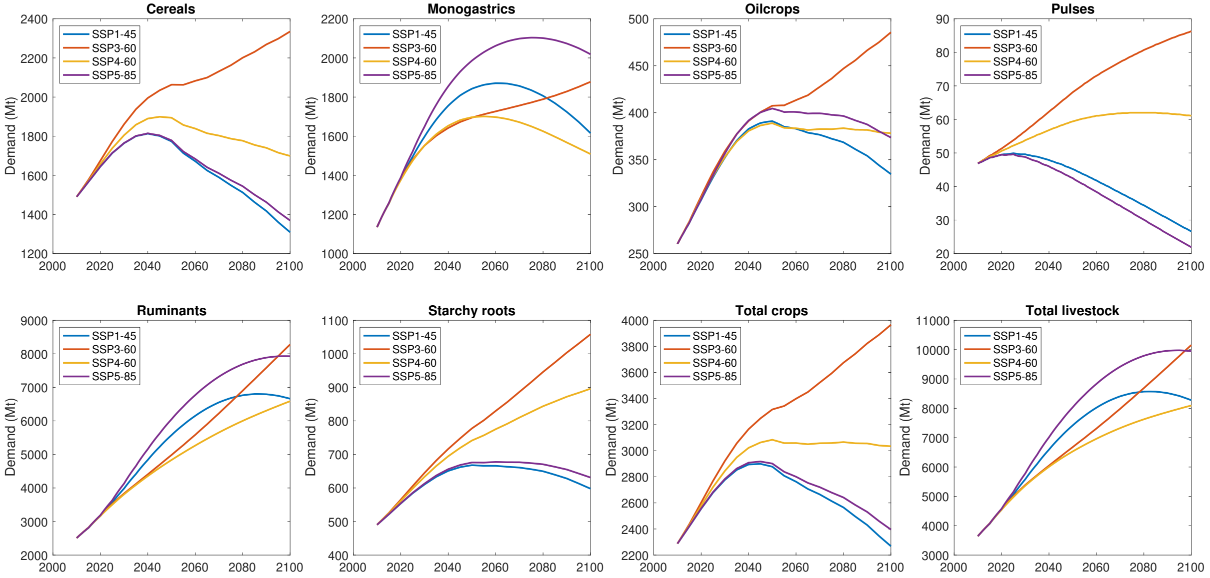


Figure SR2. Demand calculated by PLUM for each commodity in each scenario. Demand for livestock (monogastrics, ruminants, and total livestock) is given in terms of feed equivalent.

.

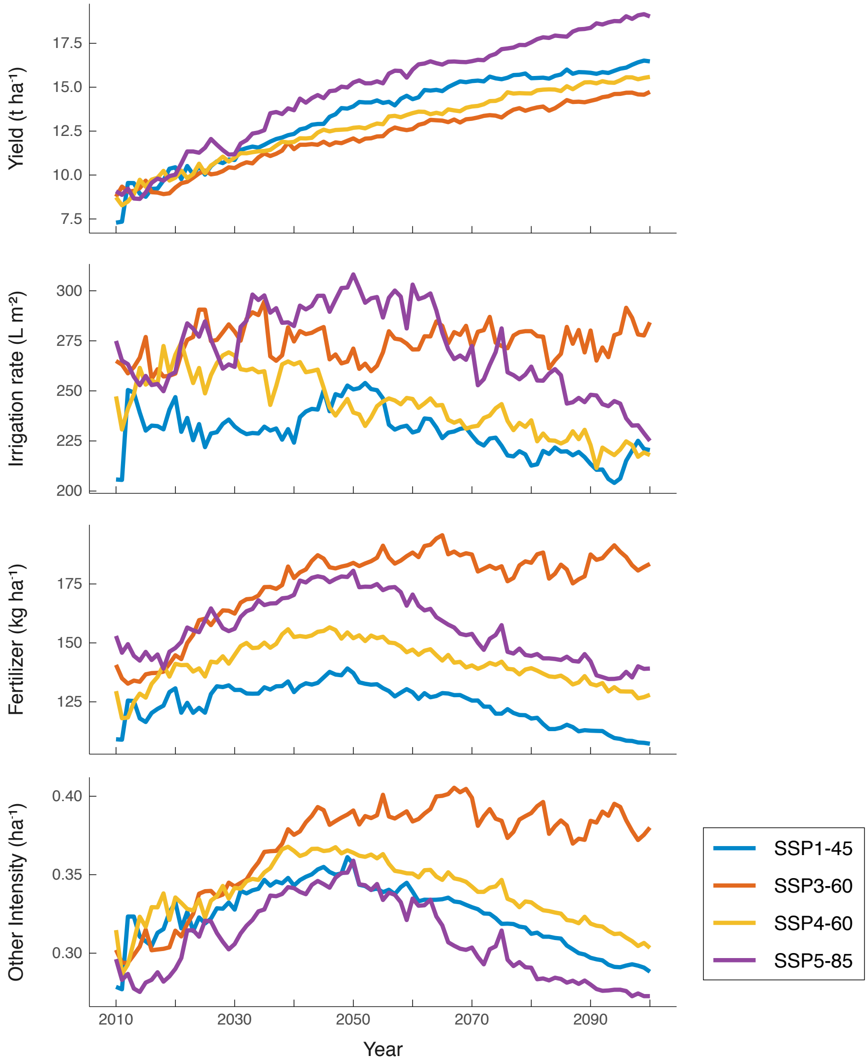


Figure SR3. ~~Global~~ Expected global management inputs and ~~expected~~ yields calculated by PLUM (before harmonization) for each scenario. Averages per area of cropland. Non-harmonized values at beginning of period do not align because of scenario-specific parameters in PLUM.

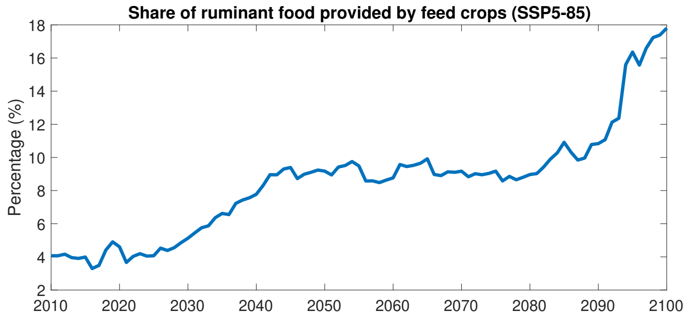


Figure SR4. Fraction of SSP5-85 ~~**(top)** livestock~~ ruminant food demand ~~and **(bottom)** crop usage~~that is satisfied by feed crops, as opposed to pasture grasses.

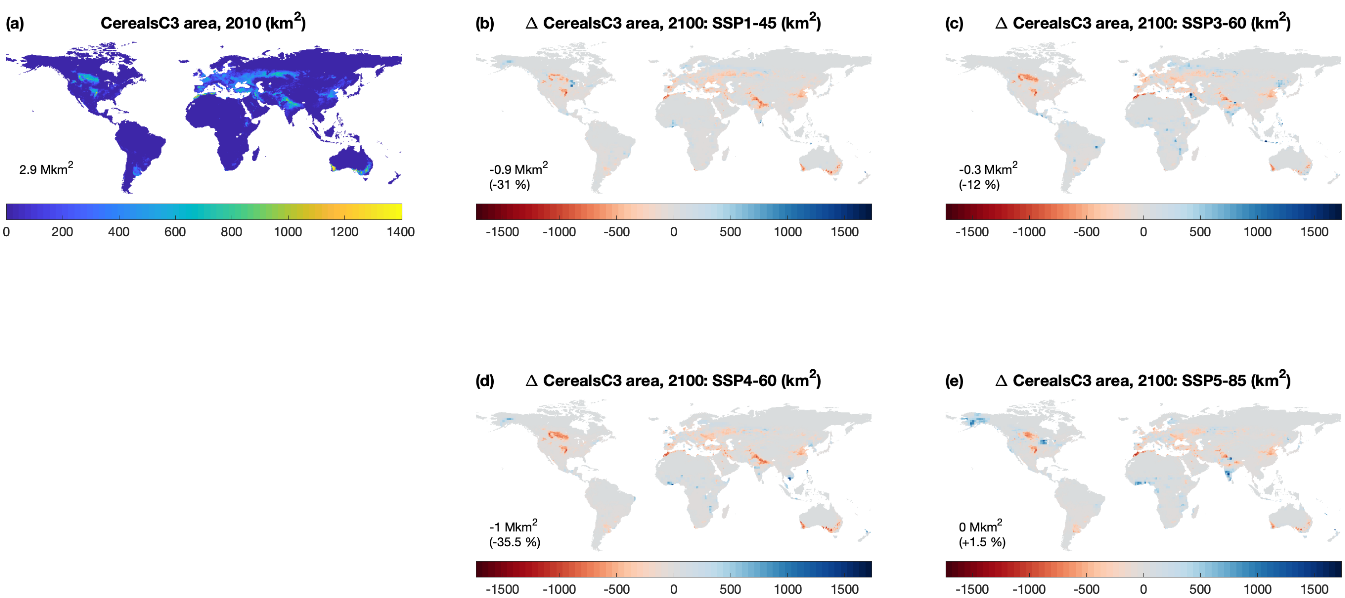


Figure SR5. (**a**) Area of cropland planted with CerealsC3 in 2010 from LUH2. (**b–e**) Difference between (**a**) and area in 2100 from PLUM for (**b**) SSP5-85, (**c**) SSP3-60, (**d**) SSP1-45, (**e**) SSP4-60.

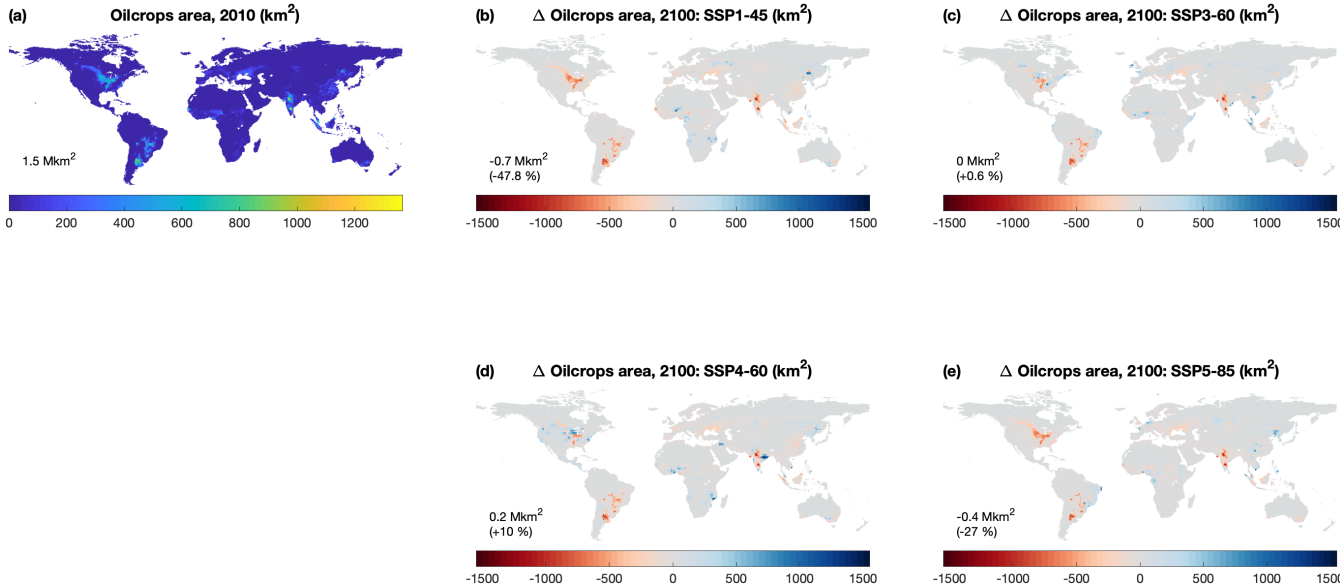


Figure SR6. As Figure SR5, but for Oilcrops: (**a**) Area of cropland planted with Oilcrops in 2010 from LUH2. (**b–e**) Difference between (**a**) and area in 2100 from PLUM for (**b**) SSP5-85, (**c**) SSP3-60, (**d**) SSP1-45, (**e**) SSP4-60.

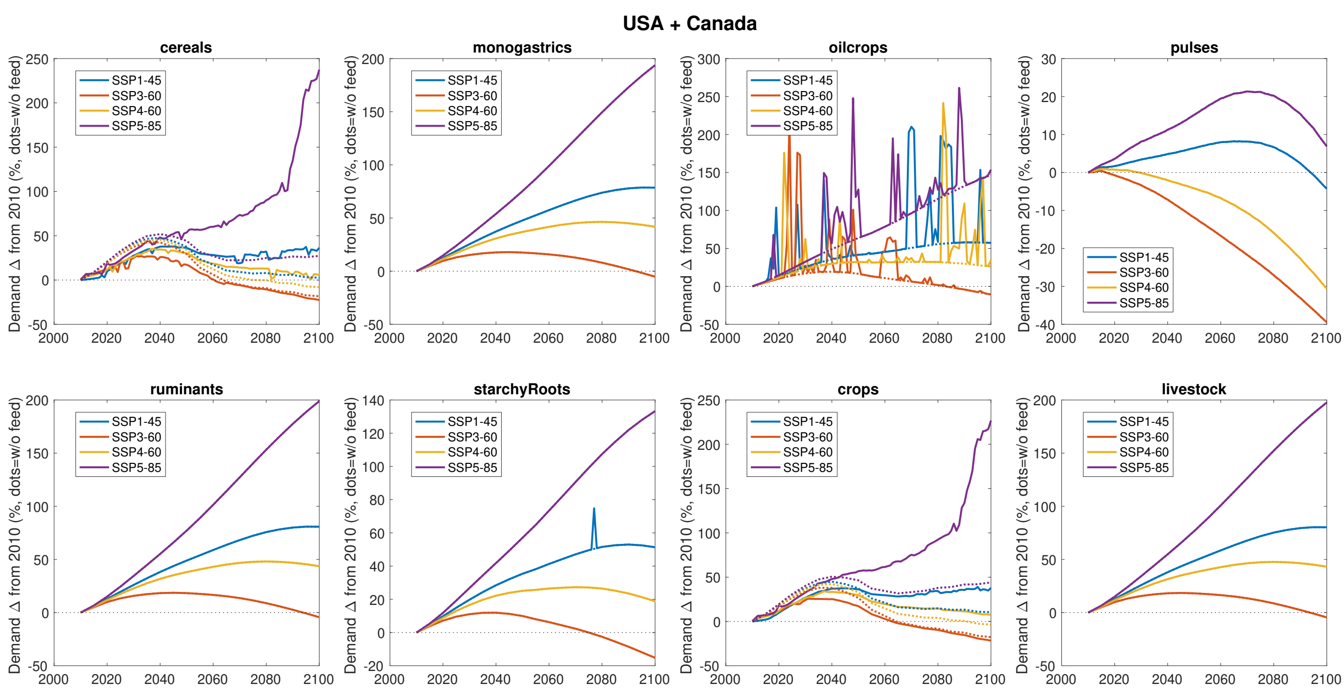


Figure SR7. Percentage change in demand in North America (United States and Canada) for commodities and commodity groups in each scenario. Solid lines include all uses; dotted lines exclude feed.

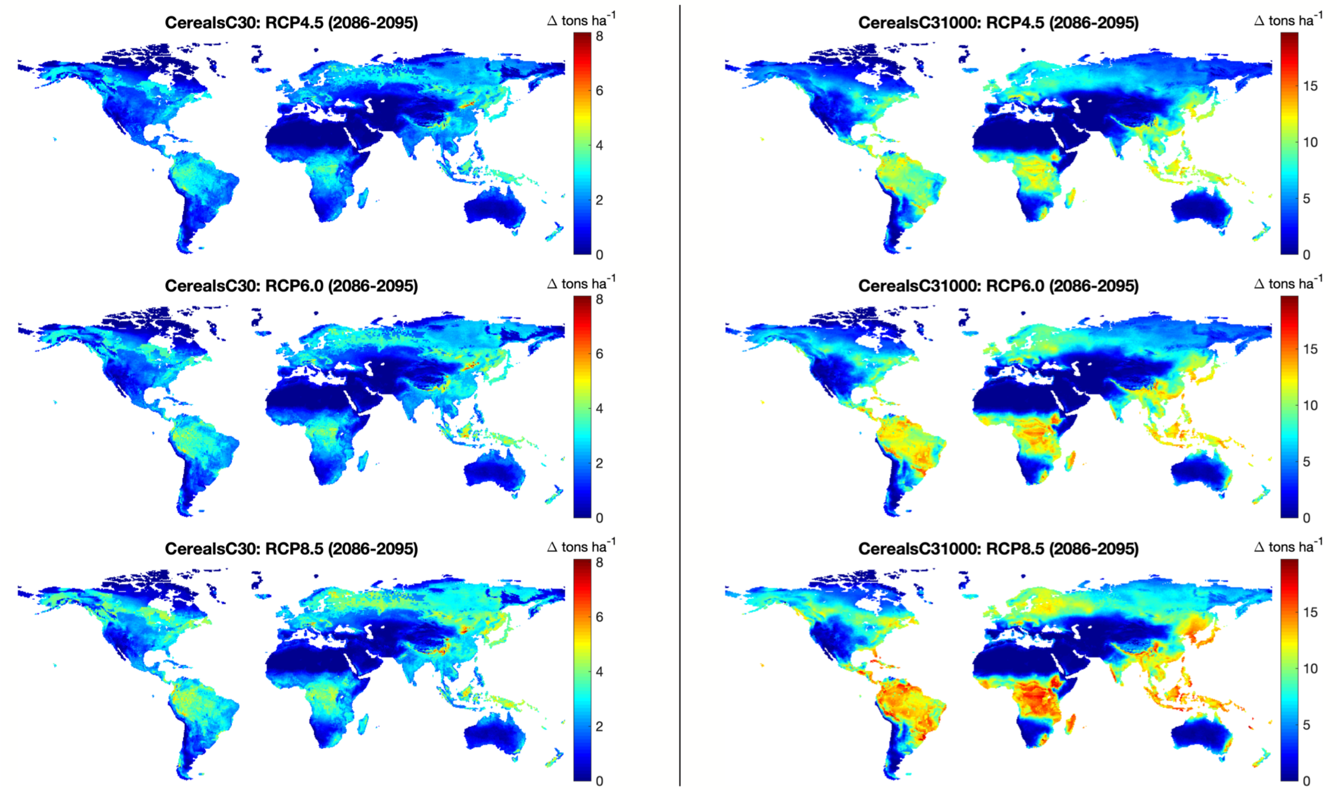


Figure SR8. LPJ-GUESS simulated mean yield in 2086–2095 (not including PLUM calibration factors) for rainfed CerealsC3 with (**columns**) 0 and 1000 kgN ha^{-1} (**rows**) in each climate scenario, from yield-generating potential runs. Note different color scales between columns.

~~As Fig.SR7 but for South Asia (India, Sri Lanka, Pakistan, Afghanistan, Bhutan, and Nepal).~~

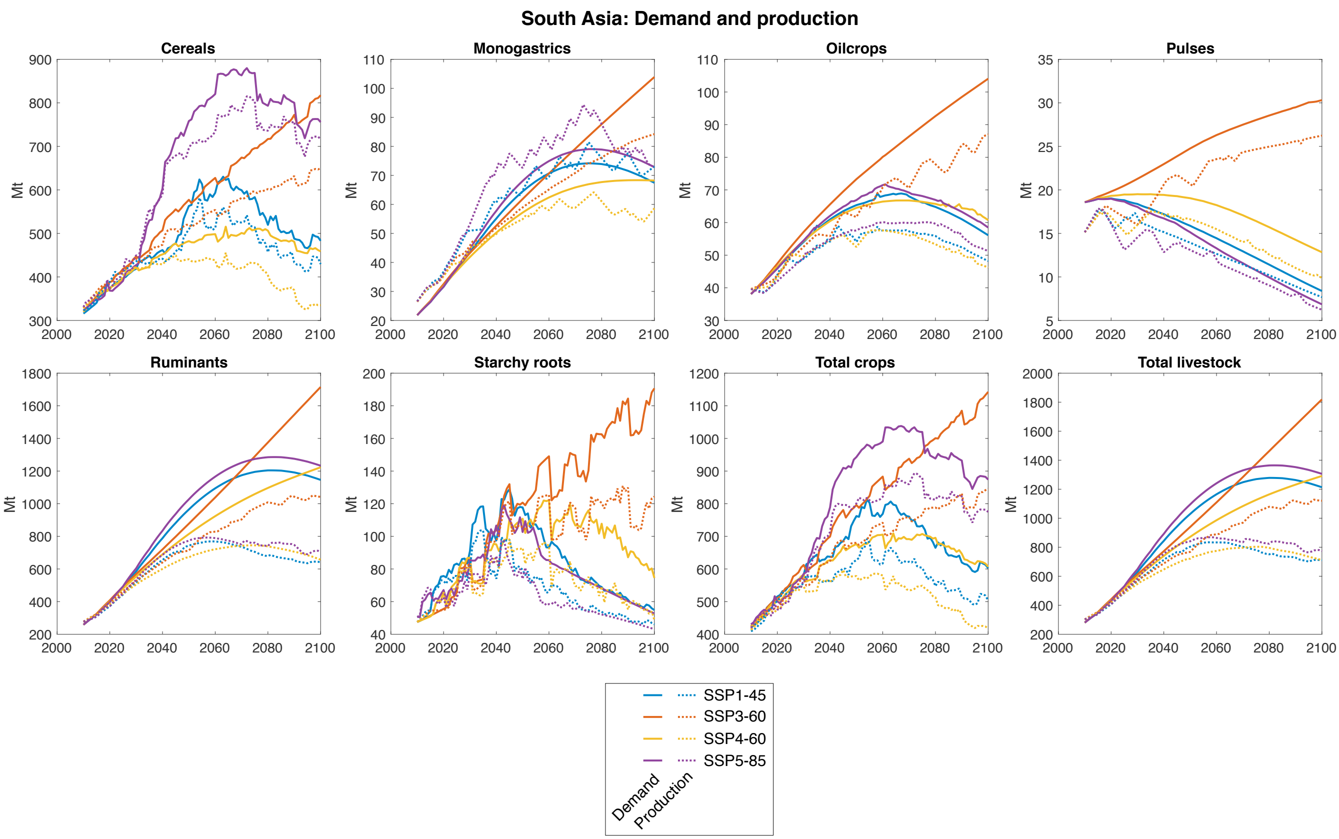


Figure SR9. ~~Percentage change in PLUM-expected~~ Demand and PLUM-expected domestic production ~~in~~ for South Asia (India, Sri Lanka, Pakistan, Afghanistan, Bhutan, and Nepal)~~for commodities and commodity groups in each scenario. Solid lines include production for domestic use and exports; dotted lines include only exports~~.

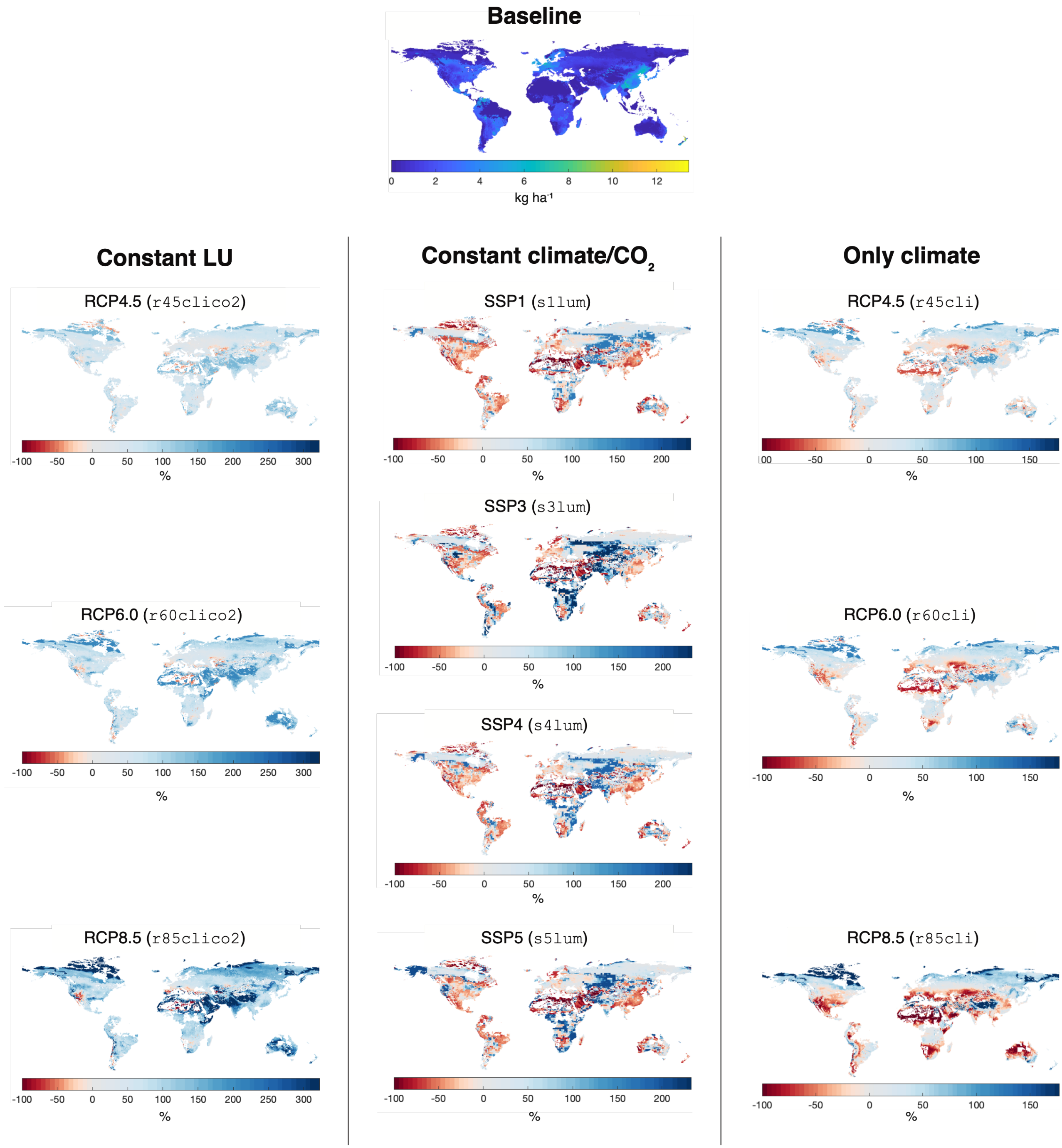


Figure SR10. Percent change in mean yield (kg ha^{-1} yr^{-1}) of CerealsC3 from 2001–2010 ("Baseline") to 2091–2100 in Constant-LU (rYYclico2; left), Constant Climate~~+CO2~~ /CO_2 (sXlum; center), and Only Climate (rYYcli; right) experiments for each scenario. Note that color scales differ between columns.

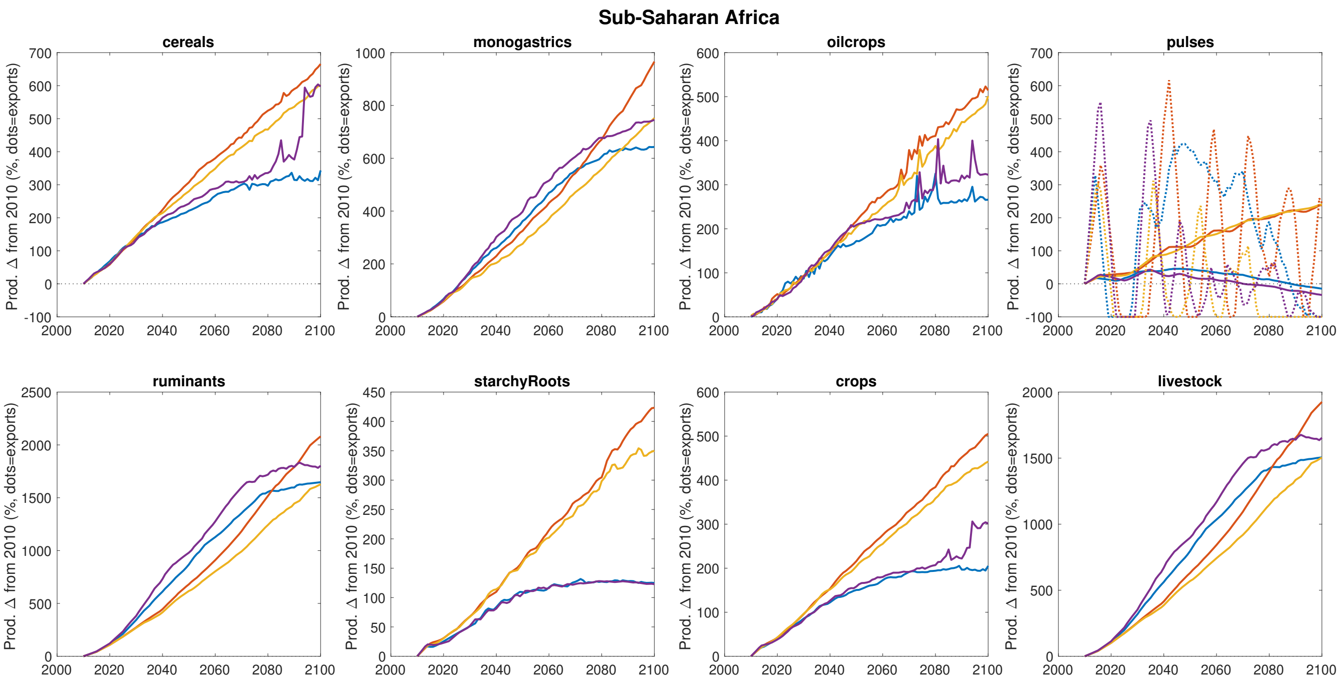


Figure SR11. ~~As Fig.?? but~~ Percentage change in production for commodities and commodity groups in each scenario for Sub-Saharan Africa (Madagascar plus all continental African countries except Algeria, Djibouti, Egypt, Morocco, Libya, and Tunisia). Solid lines include all production; dotted lines represent exports.

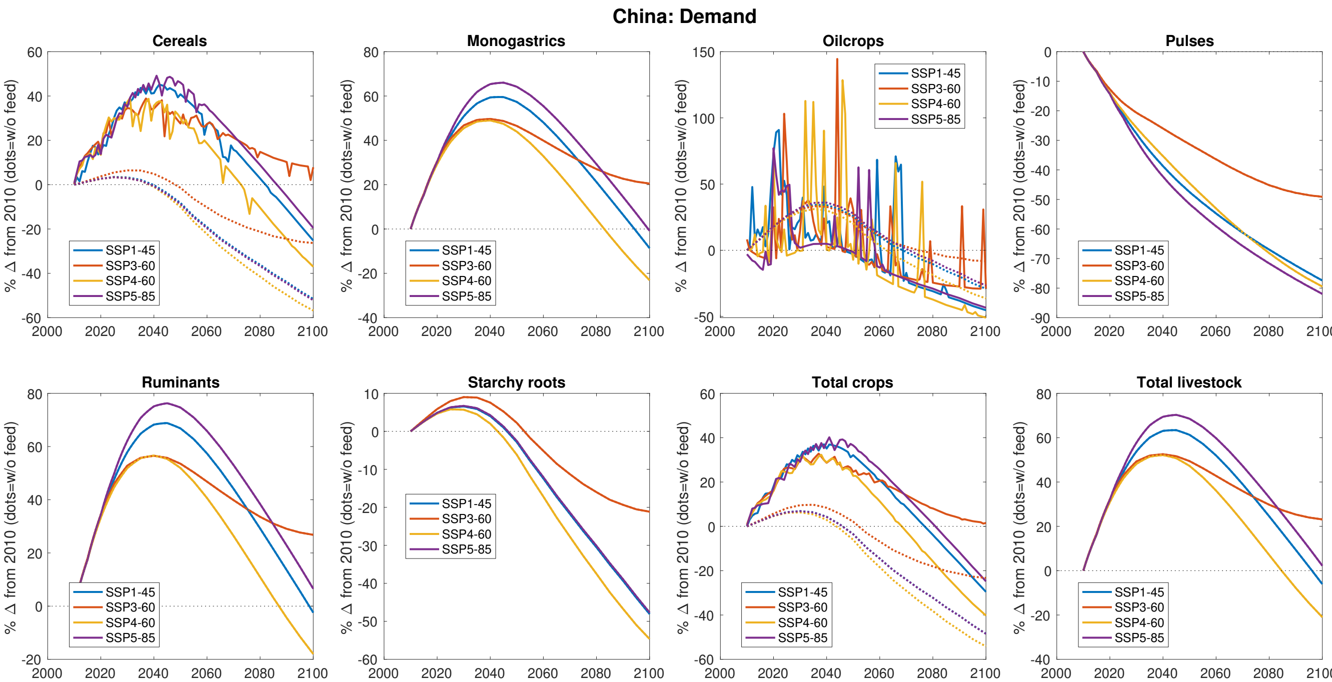


Figure SR12. Demand trajectories for China.

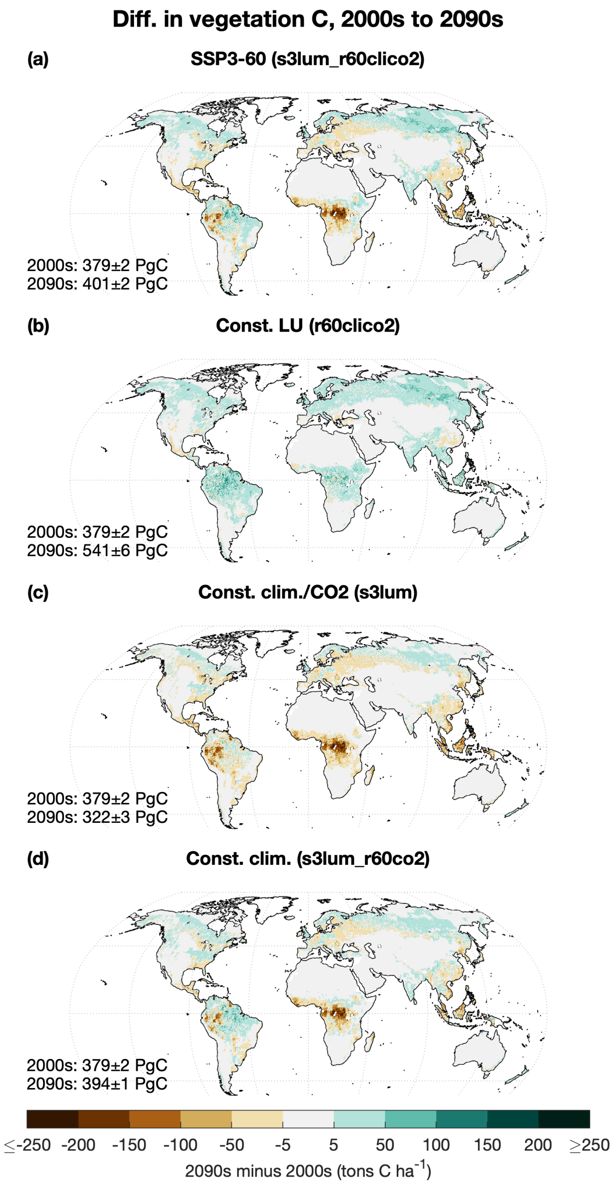


Figure SR13. Maps showing difference in mean vegetation carbon between 2001–2010 ("2000s") and 2091–2100 ("2090s") for (**a**) SSP3-60, (**b–e**) related experiments with land use, climate, and/or CO_2 held constant. Overlaid text provides decadal means and standard deviations.

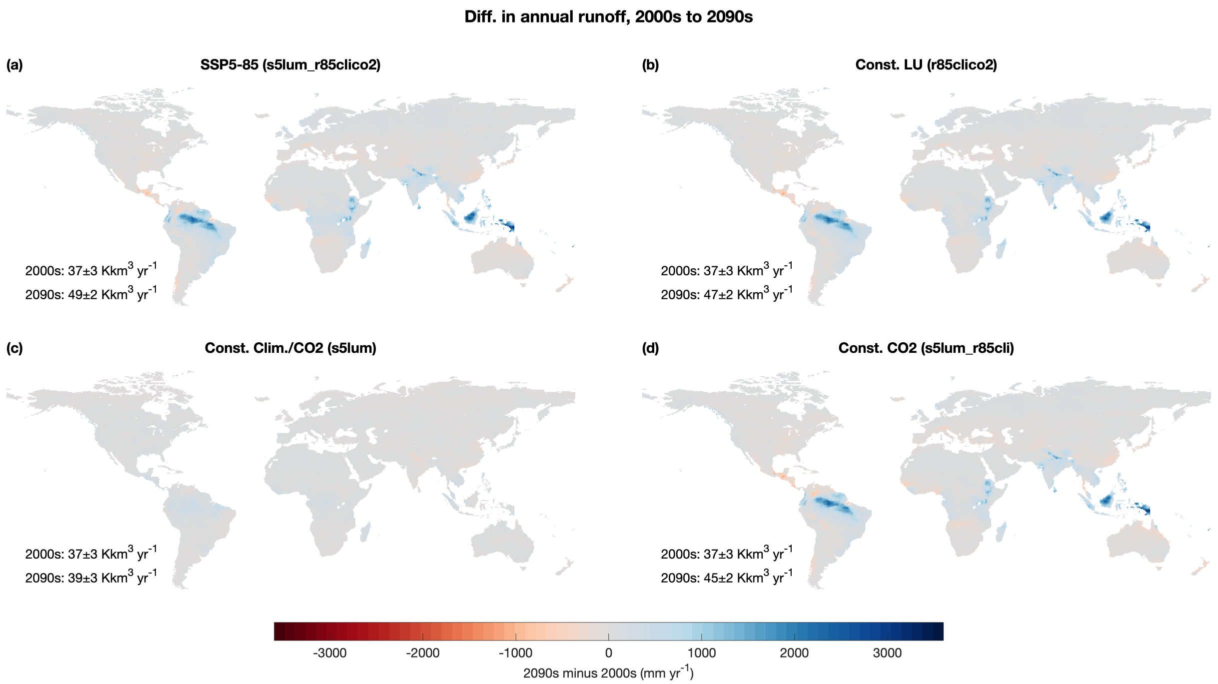


Figure SR14. Maps showing difference in mean annual runoff between 2001–2010 ("2000s") and 2091–2100 ("2090s") for (**a**) ~~SSP3-60~~SSP5-85, (**b–e**) related experiments with land use, climate, and/or CO_2 held constant. Overlaid text provides decadal means and standard deviations.

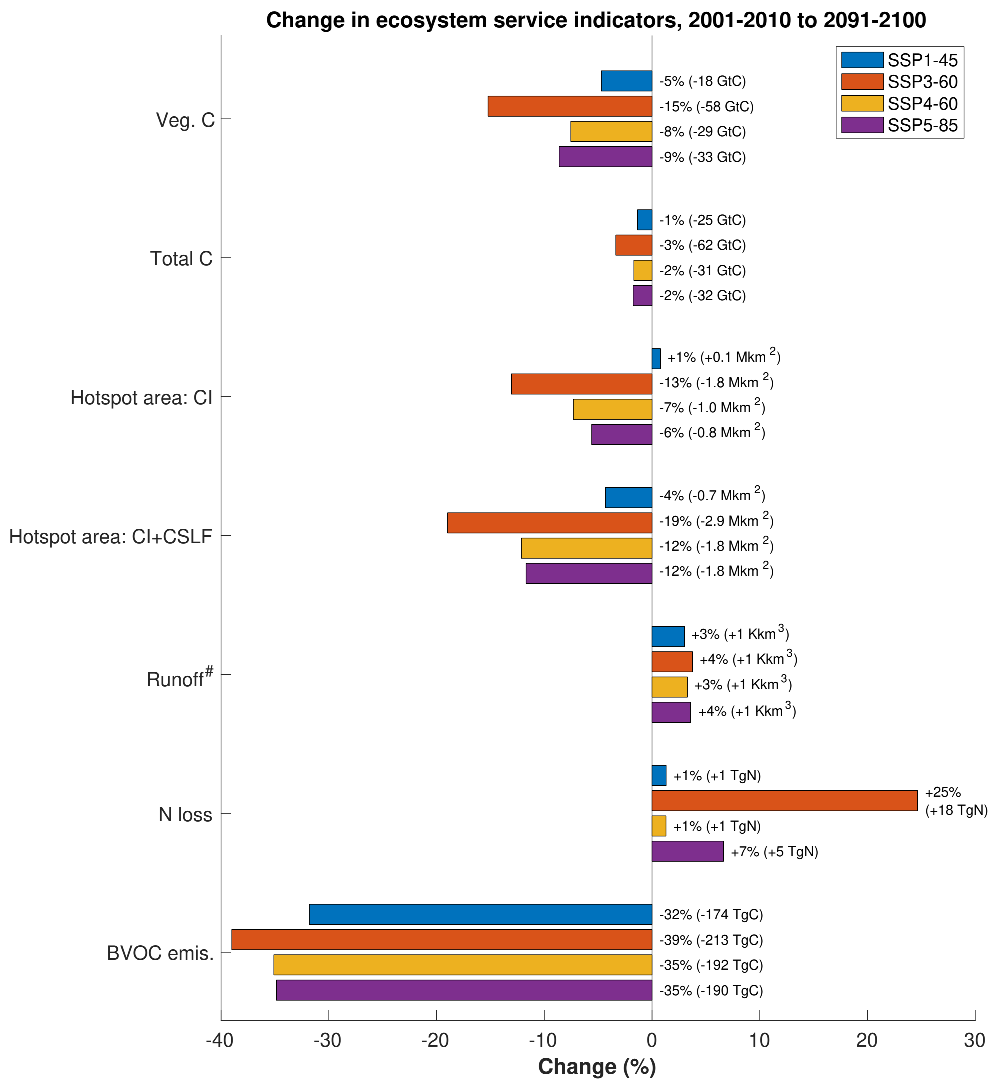


Figure SR15. As Figure ~~3~~4 in main text, but for ~~the Constant Climate+CO2 experiment~~experiments with climate and CO_2 held constant (sXlum): Percent global change in ecosystem service indicators between 2001–2010 and 2091–2100. CSLF: Congolian swamp and lowland forests (see Sect. 3.2.5). #The time periods compared for runoff were 1971–2000 and 2071–2100 due to high interannual variability.

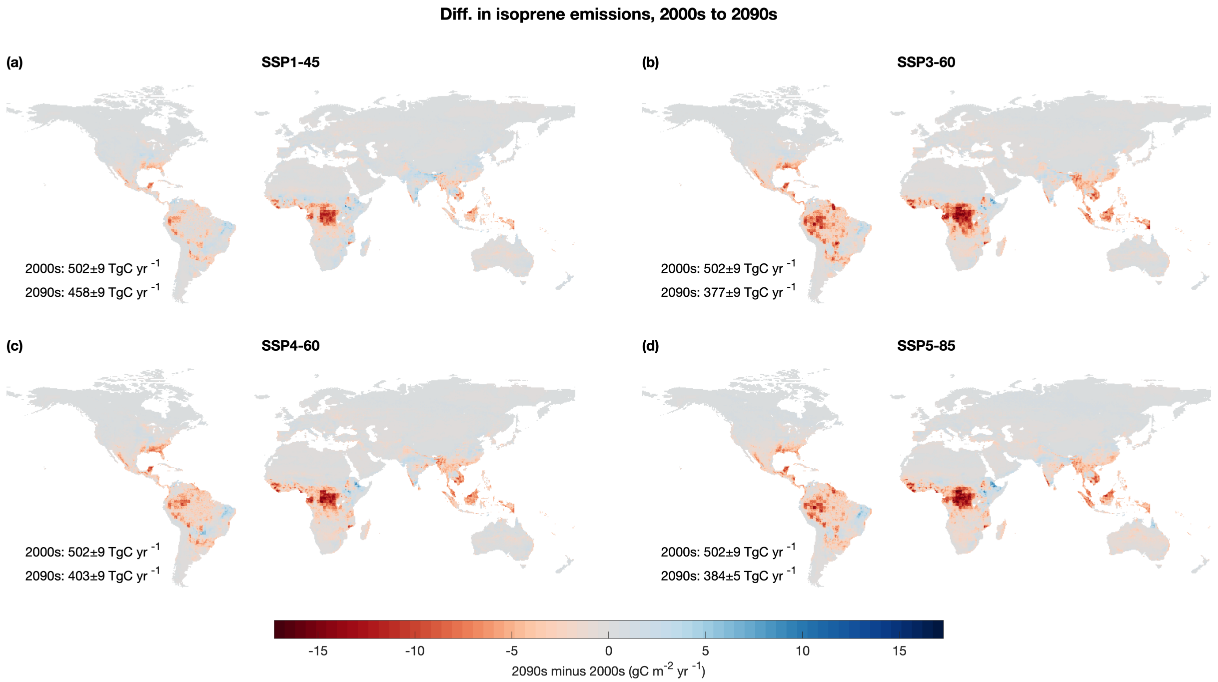


Figure SR16. Difference in isoprene emissions (gC m^{-2} yr^{-1}) between 2001–2010 and 2091-2100 in each scenario.

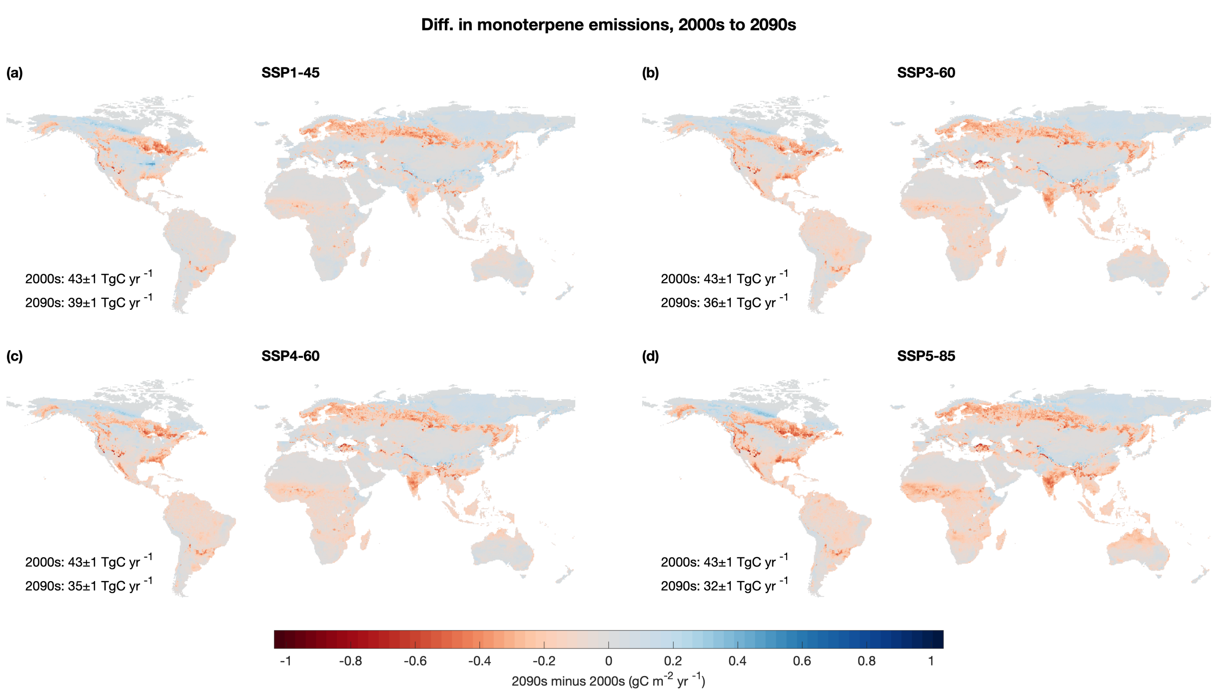


Figure SR17. Difference in monoterpene emissions (gC m^{-2} yr^{-1}) between 2001–2010 and 2091-2100 in each scenario.