Point by point replies

Point by point replies - Jonathan Sanderman

Preface: This MS presents the findings of a spatially explicit implementation of a new IPCC Tier II approach to soil carbon stock changes. The authors use this model to calculate how much SOC has been lost due to agriculture but then go on to run the model annually for the period 1975 – 2010 to produce a dynamic picture of SOC recovery over this modern era of farming. The major takeaway message is that while agriculture has incurred a large SOC debt, recent agronomic improvements have led to 4 Pg C of SOC sequestration over this period of time. This detailed picture of SOC in croplands over the past several decades is of incredible importance to policy makers and as such I believe that paper can be an important contribution to the literature; however, I do have several major concerns that may or may not be addressable with revisions.

Answer to preface:

Dear Dr. Sanderman, thank you for the thorough and helpful review of our manuscript. While checking and revising our data processing in response to your and the other reviewer's feedback, we discovered a bug in the soil model, leading to an overestimation of the C transfer to the soil exclusively for cropland. Additionally, we found unreasonable high forage crop production values in our input data, which are taken from FAO statistics. This made the overall intensification trend in agriculture lead to increasing carbon stocks in cropland. After correcting the bugs, this is no longer the case. Following upon your and your co-reviewers comments we improved additionally the initialization of our SOC estimates by extending the spin-up to a much longer period (1510) and refined the natural litter representation substantially.

Whereas this implies major revisions of our results, discussion and interpretation, we argue that the essence of the paper remains intact, albeit modified. We suggest that our key findings are a) we introduce a soil carbon model that can account for changes in agricultural management and can be applied within integrated assessment frameworks for the first time and b) we show that it is critical to account for management dynamics in SOC assessments. We provide an assessment of how results changed after correcting the bugs as a supplement to our last author's comment.

Here, we respond point-by-point to the reviewers' comments. Please note, that due to the amount of changes introduced within this update, we will refer to the section within the new manuscript rather than include the entire changed paragraph within this replies. We look forward to your response.

1. I had to read the methods section twice and spend an hour with Calvo Buendia et al., 2019 to fully understand what the authors have done. I'm still not 100% confident that I fully understand the methodology. I suggest adding an illustrative example or two graphically demonstrating how the process works. Perhaps starting with a simple case of one lu transition and then showing a more complex case of multiple lu transitions within a pixel.

Answer to 1.:

We added a graphical representation (see Fig. 1) for demonstrating the land-transition accounting to the method part (see Sect. 2.1.2 on "SOC transfer between land-use types").

2. I have not been convinced that this sort of "dynamic" implementation of a state modeling approach is appropriate. I understand that the method was developed by the IPCC as a way to add more nuance into the Tier I emission factor approach but I don't think the method was intended to be applied annually. Why not go all the way to Tier III approach using the

process-based dynamics that are embedded in this simplified model? It appears you have all the data assembled to do this. My main concern with applying a steady-state model to annual changes is we know that the recent past trajectory of SOC (particularly in the slow and passive pools) will greatly influence the short-term model response to improved management – i.e. the model will take years to decades before SOC stocks start to rebuild if the trajectory was negative prior to the change – but this will be completely missed with the steady state application (stocks will start increasing immediately upon change).

Answer to 2.: We would like to clarify that the Tier 2 modeling approach within the refinement of the IPCC Guidelines of 2019 is a first order kinetic approach. It was just named by the IPCC authors as "steady-state method", since it is based on steady-state calculation as an intermediate step. Yet, it does not assume SOC stocks in immediate steady state, and does account - depending on the pool type - for longer than annual transition phase to new steady states. We share your confusion about the naming within the guidelines, and changed the naming in our analysis from "steady-state method" to "Tier 2 modeling approach".

We also highlight in the introduction and conclusions that our Tier-II approach with its reduced complexity has advantages to a more complex process-based Tier-III approach when integrating it into computational intensive Integrated Assessment Models.

3. What time frame are the monthly climate data averaged over to get the rate modifiers for a steady state solution? Given the passive pool has an intrinsic decay rate equivalent to >100 year turnover time, it seems that you need to have a 100+ year average climate to come up with the proper rate modifiers.

Answer to 3.: As we do not use a steady state model, we believe the comment might be based on a misunderstanding and no averaging of climate data is needed.

4. Transfer between lu types is not clear. I do not understand how a "respective share of the SOC is reallocated." My concern is that the per area SOC stock for long-term cultivated land will be much different than the per area SOC stock for newly converted cropland, so I don't see how you can suddenly bin these separate areas into one model component. Perhaps my request for a visual guide will help me (and other readers)understand that I have misunderstood this part of the methods.

Answer to 4.: See answer to 1. The added scheme of land-use transition representation (see Fig.1) illustrates the linearity of the problem; accounting separately for newly converted cropland and existing cropland is - in our equation system - mathematically the same as taking the area-weighted mean.

5. Why was 1901 chosen to spin up to steady state? We know that this was a time of rapid agricultural expansion in several major regions of the world and thus a time of rapid soil carbon loss.

Answer to 5.: We extended the spin-up phase for a much longer period, starting in 1510 (the default spin-up start in LPJmL) to capture the land-use effects. Our initial choice of starting in 1901 was driven by the availability of climate data. To extend our spin-up into the past, we will recycle climate data from 1901 to 1930, as this is done in DGVM simulations as well (Schaphoff et al. 2018a/b, von Bloh et al. 2018). We rewrote the Sect. 2.1.4 on "Initialization of SOC pools" to explain this in more detail.

(References: von Bloh, W. et al. 2018: Implementing the nitrogen cycle into the dynamic global vegetation, hydrology, and crop growth model LPJmL (version 5.0). Geoscientific Model Development 11, 2789–2812. Schaphoff, S. et al. 2018a: LPJmL4 – a dynamic global vegetation model with managed land – Part 1: Model description. Geoscientific Model Development 11, 1343–1375. Schaphoff, S. et al. 2018b: LPJmL4 – a dynamic global vegetation model with managed land – Part 2: Model evaluation. Geoscientific Model Development 11, 1377–1403.)

6. LULUC data – are these data all provided as area within each grid cell (or percent of a grid)? I think so, but please indicate. A supplemental table with the cross-walk between the LUH2 and the 17 crop groups used in this study should be included.

Answer to 6.: Yes, all area data are in million hectares and meant to represent the area within each grid cell. We indicated that more clearly in Sect. 2.3.1 by adding "(given as total land area in million ha)" to our explanation on "physical area" (LN 200). Additionally, we added a mapping table to the data archive at zenodo translating LUH2 croptypes and our crop types into FAO crop categories (see Table "FAO2LUH2MAG_croptypes.csv").

7. Please include units for eq 9 – I don't follow the AGR calculation – it sounds like you are adding biomass and area together. Additionally, HI is usually defined as CP divided by total above ground biomass, so CP x HI is a meaningless number.

Crop residue production (ton dry matter residues) are calculated based on a linear function with positive intercept and a yield-dependent slope. This allometric function accounts for the fact that higher-yielding crops often have a lower harvest index than low-yielding crops. In our revised manuscript, we rewrote the Eq. 9 to make the functional form more visible to

$$\begin{aligned} AGR_{i,t,cg} &= CA_{i,t,cg} \cdot \left(Y_{i,t,cg} \cdot r_{cg}^{\text{ag,prod}} + MCF_{i,t} \cdot r_{cg}^{\text{ag,area}}\right) \qquad \text{and} \\ BGR_{i,t,cg} &= \left(CA_{i,t,cg} \cdot Y_{i,t,cg} + AGR_{i,t,cg}\right) \cdot r_{cg}^{\text{bg}} \quad . \end{aligned} \tag{1}$$

Where one can also simplify $CA \cdot Y$ to a production value, as we did in the last version of the manuscript; keeping area and yield explicit in the new version however makes the dynamic harvest index more explicit.

We also added more detail to Sect. 2.3.2 on "Crop and crop residues production" and units for the three crop-group cg specific ratios: * above-ground residues to harvested biomass $r_{cg}^{\text{ag,prod}}$ in $(tDM\,ha^{-1})(tDM\,ha^{-1})^{-1}$ * above-ground residues to harvested area $r_{cg}^{\text{ag,area}}$ in $tDM\,ha^{-1}$ * below-ground residues to above-ground biomass r_{cg}^{bg} in $tDM\,tDM^{-1}$.

7. Lack of validation. There appears to be no attempt to validate the model or the input data used to drive the model. In general, there is a lack of quantitative evaluation throughout. There are just two qualitative quality assessments – a table comparing calculated stock change factors to Tier I estimates and a discussion on how the map looks similar to other SOC maps.

Answer to 7.:

To address the raised issue, we added to our analysis a whole section on "Model evaluation" (see Sect. 3.4) featuring (next to the assessments done within in the first draft): * a grid level comparison of SOC stock results to SoilGrids 2.0 and LPJmL4 to improve spatial evaluation of our results, * a comparison to uncertainty estimates of SoilGrids 2.0, * a comparison to point-based SOC data.

8. The model itself was developed recently as part of the 2019 Refinements to IPCC Guidelines and those updated guidelines discuss how the model was calibrated to a set of long-term trial sites but do not report any model performance metrics. As pointed out by the authors some areas of central EU and the UK more than double SOC undercurrent agriculture than under native vegetation. This is certainly indicative that there should be some checks against real data (see detailed elaboration on this point further down in this review). It could be argued that point-based validation for a model run at0.5 degree resolution is meaningless but it would be an interesting exercise to see how the model reproduces trends with known SOC histories.

Answer to 8.: Our updated model does not show the surprising behavior of strongly increasing SOC stocks in UK and central EU anymore due to the correction within the model and the corrected fodder production data. Additionally, we added point-based evaluation of the results (see answer to 7).

9. Issues with residue C return. Given that the major takeaway from this paper is that the SOC is being sequestered due to improved yield leading to increased residue return, is there any empirical evidence that C inputs to the mineral soil have nearly doubled (Fig 3)? I think the method for calculating residue return to the soil is potentially flawed leading to this large apparent increase. The authors have assumed that both harvest index (HI) and root-to-shoot(RS) ratios have been constant through time. However, yield improvements over the

last century, and in particular the last 50 years, are a result of improvements in genetics and nutrition. Breeding has resulted in the ability to plant most crops at much higher densities and selection towards more photosynthate being allocated to harvestable organs. Both of these improvements have altered HI and RS ratios. Additionally, there are strong interactions between N fertilization rate and root density. There is a huge literature on crop breeding that support the nonstationarity of these important parameters.

Answer to 9.: We took up this point in the discussion. In particular, we would like to highlight the following: The IPCC methodology, which we use here, tries to capture the effect of a shifting harvest index by making the harvest index a linear function of yield with a positive intercept (see reply to point 7). The parametrization of this dynamic harvest index was not possible for all crops due to a lack of literature estimates, but the most important crops like cereals or soybeans are captured. We rewrote Sect. 2.3.2 "Crop and crop residues production" to be more clear on this aspect and reformulated Eq. 9. See also answer to 7.

10. Units – Gt and Mt are not SI units, please use Pg and Tg

Answer to 10.: To be consistent within the paper, we sticked to tons (t) and hectare (ha) as our two base units as agricultural production is often measured in t per ha, land-use areas in ha and SOC estimates on global scale (e.g. SoilGrids 2.0) sometimes in t per ha as well.

11. L13 (and elsewhere) – "we also find that SOC is very sensitive..." – this is in reference to an unvalidated model result. I'd suggest rewording these sorts of phrases to, "Our model results suggest that SOC is very sensitive...

Answer to 11.: We carefully revised our manuscript on naming and framing of the evaluation related statements.

12. L279 – "we provide the first world map" – no, you did not. All of the global maps that have been developed using a statistical environmental-covariate modeling approach(i.e. soilgrids and similar) implicitly include all historic land management.

Answer to 12.: This is correct. We will rewrite the paragraph.

13. It is great that all the data are provided but I found the Karstens 2020a repository to be confusing. Can you have a description for each file in the repository? The naming convention is not clear. I did not want to download 9 Gb of data to figure it out.

Answer to 13.: We added more detail to the README file including a description for each file. We also changed the output format to the more commonly used NetCDF format.

14. Fig 1 – perhaps it is just the spatial scale of these small maps (and I haven't downloaded the results to explore in more detail) but it looks like there is zero intact forest in the Congo Basin and very little intact forest in the Amazon. Also, I would have likedto see a map showing the trend in SOC spatially – how are the 4 Pg C that has been sequestered been spread across the globe? Is it all in Central EU and UK?

Answer to 14.: Fig. 2 (a) showed cropland SOC for every grid cell that contains cropland, without giving an information on the extent of cropland. We will mask out cells with very low cropland area (less than 1000 ha on a 0.5 degree grid), as they might give to the impression of greater cropland extent within large parts of intact forested area. We now also provide a total SOC debt and SOC debt trend map (see Fig. 3).

15. Fig 2 – this is a really interesting way of summarizing the model results

Answer to 15.: Thank you!

16. Fig 4 – the finding presented here is very counterintuitive to me. Why is the SOC debt halved when the model is initialized with natural vegetation? Shouldn't the 1975 SOC debt be much greater if the 1901 starting point was natural vegetation instead of actual land use? Perhaps I am just misunderstanding this sensitivity analysis.

Answer to 16.: The initialization analysis was meant to help the reader understand the potential maximal error of underestimating on-going emissions in cropland, that where converted to cropland before 1901. We have now extended the spin-up phase from 1510 to now. We therefore omitted the initialization analysis in the revised manuscript.

17. Discussion section – in general, there is very little discussion of how these results fit into the large literature on SOC. There are many places were a reference or two would great increase the credibility of the statements that are being made.

Answer to 17.: We rewrote the whole discussion section and included more connections to existing literature here.

18. L358-360 – the finding that northern temperate zones (particularly in EU and UK) now have SOC levels up to twice that of native state yet tropical soils have lost 40-70% of their SOC is problematic and, as the authors point out in relation to the EU example,likely points to issues with getting C input to soil correct. The EU has the perfect data set to test this model finding – the EU JRC LUCAS survey was a balanced sampling design between forested and agricultural land uses. In the tropics, it has been fairly well documented that already infertile tropical soils do not lose nearly as much SOC as their fertile temperate zone soil counterparts. While there are issues and large scale generalizations in the IPCC Tier I default factors, they do represent the consensus literature on the topic. The updated meta-analysis between the 2006 and 2019 IPCCC guidelines when this emission factor for the tropics changed dramatically (see Table 4in this MS) points to this new knowledge.}

Answer to 18.: After our bugfix, SOC stocks for the EU no longer gain SOC compared to natural vegetation. In general, all carbon stocks are much smaller and show much higher losses compared to the Tier 1 approach for both temperate and tropical soils. This may be indicative of gaps in the accounting of carbon inputs to the soil and will be discussed in more detail within the discussion. The comparison to point data is however challenging, also as point measurements do not well capture the landscape average and will likely show a very high variance. On the one hand the LUCAS database is given as soil carbon density and would need consistent bulk density data to be comparable to our results. However, we included point measurement comparison using the data provided in by Sanderman et al., 2017 for their model evaluation. Additionally, we hope that our additional comparison to SoilGrids 2.0 might help fill the evaluation gap here as SoilGrids 2.0 is based on point measurements.

19. L396 – how is this validation? It is just a comparison.

Answer to 19.: The naming and framing of the evaluation of our results was improved. We included a whole result section on "Model evaluation" (see Sect. 3.4).

20. L400-406 – there is a large literature that can be drawn upon to support some of the claims made in this section.

Answer to 20.: We rewrote the paragraph on the discussion on the stock change factors compared to the IPCC, as also some of the results have changed.

It now reads:

LN 481-495: "In comparison with default stock change factors of the IPCC guidelines, our model estimates a stronger decline of SOC stocks (Table 3) for almost all regions. Tropical soils might suffer from low C input rates due to large yield gaps (Global Yield Gap and Water Productivity Atlas. Available URL: www.yieldgap.org (accessed on: 03/01/2022)) and high shares of residue removal and burning in lower-income countries (Smil, 1999a; Williams et al., 1997; Jain et al., 2014). Yet, even when comparing our etimates to the low-input stock change factors of the IPCC, our SOC loss is roughly twice as large as the revised 2019 IPCC default values, while it shows very good agreement with the older default values from IPCC (2006). Don et al. (2011) estimated SOC losses for tropical soils of around 25% on average corresponding to a stock change factor of 0.75, but also reported a wide range of measured SOC changes from -80% to +58%. Fujisaki et al. (2015) however found much lower loss rates of around 9%, attributing the difference to the different time period length since the conversion to cropland. As our results do not specifically account for cropland age and most of the cropland is older than 20 years (as assumed for the default IPCC Tier 1 stock change factors) our stock change factors have to be lower by definition following the steady-state assumption that cropland will continue to approach a new equilibrium. For the same reason, our estimates for temperate regions might be lower than both IPCC (2006) and IPCC (2019) default values. With the production-increasing impact of irrigation and fertilization on carbon-poor dryland soils, SOC under cropland can also be higher than under PNV with stock change factors above 1 (see Fig. 2(d)), but these areas are much smaller than where the stock change factors are well below unity."

21. Section 4.4 – I do not think this is a valid comparison because SoilGrids explicitly tried to capture high carbon density soils well while your model explicitly excludes organic soils. I suggest applying an agriculture mask to all of these data sets and then redo the analysis. Additionally, ISRIC released an update to SoilGrids >6 months ago that focuses primarily on mineral soil carbon stocks. This update is probably a better comparison.

Answer to 21.: We included the recommended SoilGrids 2.0 data for a more spatially detailed evaluation of our results especially for cropland soils. We aso applied a cropland mask to most f our results inclusing only grid cells with more than 1000 ha of cropland.

22. L425-428 – this seems out of place.

Answer to 22.: We removed the sensitivity analysis on plant parameterization, because we greatly improved litter parameterization of the natural vegetation.

23. L453 – comparison to 4p1000 is not really fair because your model is really just the business-as-usual scenario with SOC gains simply because yields are improving glob-ally. 4p1000 is about intentional management shifts to increase SOC.

Answer to 23.: We agree with the reviewer that the comparison is a bit misleading. We still think that it is fruitful to compare the observed rates with ambitious targets on the global scale and reframed it to: LN 514-517:"Annual C loss rates of 0.2 per 1000 C still have the opposite trend as the promoted 4 per 1000 C sequestration rate target (Minasny et al., 2017). Dedicated efforts to increase cropland SOC are thus necessary, as management improvements at historical rates are not enough to counteract ongoing SOC degradation on cropland."

Point by point replies - second reviewer

0. The authors have conducted a study evaluating the influence of management on soil organic carbon in global croplands. This is an important topic for consideration of greenhouse gas mitigation with natural solutions for climate change policy and pro-grams such as the 4 per mille initiative. As the authors mention, there are few studies that have evaluated cropland management effects on soil organic carbon, and possibly none that have addressed the influence at a global scale. The result that increased residue return to soils is the leading driver of carbon changes over the past few decades in croplands is an important finding. As the authors note, the 4 Gt C increase in carbon is less than the goals of the 4 per mille initiative, which some have argued is not realistic. I have a few concerns about the study after review of the IPCC documentation on the method that the authors selected for this analysis. I would suggest that the authors make revisions before the manuscript is accepted.

Answer to 0.: Dear reviewer, thank you for the thorough and helpful review of our manuscript. While checking and revising our data processing in response to your and the other reviewer's feedback, we discovered a bug in the soil model, leading to an overestimation of the C transfer to the soil exclusively for cropland. Additionally, we found unreasonable high forage crop production values in our input data, which are taken from FAO statistics. This made the overall intensification trend in agriculture lead to increasing carbon stocks in cropland. After correcting the bugs, this is no longer the case. Following upon your and your co-reviewers comments we improved additionally the initialization of our SOC estimates by extending the spin-up to a much longer period (1510) and refined the natural litter representation substantially.

Whereas this implies major revisions of our results, discussion and interpretation, we argue that the essence of the paper remains intact, albeit modified. We suggest that our key findings are a) we introduce a soil carbon model that can account for changes in agricultural management and can be applied within integrated assessment frameworks for the first time and b) we show that it is critical to account for management dynamics in SOC assessments. We provide an assessment of how results changed after correcting the bugs as a supplement to our last author's comment.

Here, we respond point-by-point to the reviewers' comments. Please note, that due to the amount of changes introduced within this update, we will refer to the section within the new manuscript rather than include the entire changed paragraph within this replies. We look forward to your response.

1. The Tier 2 method is in a croplands chapter of the IPCC report. The documentation in the report states that the model would need to be parameterized for other land uses.Did the authors parameterize the model for other land uses that would be considered natural vegetation? If not, the estimation of soil organic carbon for natural vegetation may not be valid. The authors seem to suggest that this is a possibility in Section 4.4 when stating the soil organic carbon and debt from land use change have to be interpreted with caution. If the model has not been parameterized for natural vegetation, I would suggest that the authors focus on cropland model results, and remove the carbon debt results. The results for the cropland alone are important, and deserve publication even if the natural vegetation estimates are not valid with this model.

Answer to 1.: This is an important point. No we did not reparameterized the model for other land-use types. While our analysis focuses on cropland, estimating the natural soil carbon stocks is necessary to account for the C entering the cropland budget via land conversion. To improve and evaluate the parametrization of natural soil carbon, we will include the following model updates: * We improved the litterfall parameterization in natural vegetation, accounting for plant-functional type and plant-organ specific parameterization of nitrogen and lignin contents. * We compared our results on soil stock under natural vegetation with the results of a model parametrized for natural vegetation (LPJmL4). * We compared our results for natural vegetation with a dataset of point measurements

2. Is it possible to estimate uncertainty with this method? IPCC methods often have large uncertainty but does this method have less uncertainty because it is a Tier 2method. If it is not possible to estimate uncertainty could the authors speculate on the level of uncertainty in the predictions. Knowing something about uncertainty would be helpful in comparisons to the modeled results from other studies that are shown in the manuscript.

Answer to 2.: The quantitative assessment of the uncertainty of our projections unfortunately exceeds the scope of this article and would likely require a study in itself. The model includes a high number of parameters, and for most of these the uncertainty distributions have not been quantified so far. Moreover, we think that beyond parameter uncertainty, the structural uncertainty from the model design is also very high.

A good way to account for uncertainty, also accounting for structural uncertainty, is the comparison of different modeling approaches. Our manuscript now includes our own model, a comparison to the IPCC stock change factors, a comparison of natural vegetation estimates to the DGVM LPJmL, a comparison to a machine-learning approach based on observation data (SoilGrids 2.0), and to point measurement data. The high ranges that emerge from this comparison reveal that it is likely too early to meaningfully quantify the total uncertainty connected to global SOC estimates.

We moreover discuss sources of uncertainty within the discussion section e.g. * LN 406-410: "It is important to evaluate the validity of our results as modeling management effects on SOC dynamics at the global scale comes with large uncertainties. The model includes a large number of parameters, and for most of these the uncertainty distributions have not been quantified so far. Moreover, we think that beyond parameter uncertainty, the structural uncertainty from the model design is high. The management data itself is prone to uncertainties as well, as most of it is only indirectly calculated from reported data." * LN 423-427: "For most of the parameters used in our management estimates no uncertainty estimates exist. This is why, in our view, most of the uncertainty with respect to the impacts of SOC management is included in the management data itself, and especially in the residue and manure production and application numbers, as these are only indirectly derived from crop and livestock production, feed and area data (FAOSTAT, 2016; Weindl et al., 2017). The uncertainty of recycling shares adds on top of the uncertain total numbers of manure and residue biomass."

3. The authors state that a sensitivity analysis presented in Figure 4 shows that management impact is robust to the initialization of the soil organic carbon stocks at the beginning of the spin-up phase. But, the stocks and change in stocks almost halves the values if the initialization is done with natural vegetation. The initialization does make a difference, and needs further explanation.

Answer to 3.: It is correct that the SOC stocks are highly dependent on the legacy of management. However, Figure 4 shows that the SOC gap - the difference between a baseline scenario and a counterfactual scenario with constant management - was not strongly dependent on the initialization. We therefore concluded that the initialization does not affect our central finding.

Still, in order to improve our estimates also for the absolute SOC stock, we now extended the length of the spin-up phase, starting in 1510 (default spin-up start for introducing land use in simulations with LPJmL, see e.g. Schaphoff et al. 2018a/b, von Bloh et al. 2018). We describe this initialization in section 2.1.4, and updated the results section accordingly. This new feature makes our sensitivity analysis and the connected discussion obsolete and shortened the manuscript.

(von Bloh, W. et al. 2018: Implementing the nitrogen cycle into the dynamic global vegetation, hydrology, and crop growth model LPJmL (version 5.0). Geoscientific Model Development 11, 2789–2812. Schaphoff, S. et al. 2018a: LPJmL4 – a dynamic global vegetation model with managed land – Part 1: Model description. Geoscientific Model Development 11, 1343–1375. Schaphoff, S. et al. 2018b: LPJmL4 – a dynamic global vegetation model with managed land – Part 2: Model evaluation. Geoscientific Model Development 11, 1377–1403.)

4. Good to see that the authors have made a comparison to another approach to confirm the Tier 2 results. The Tier 1 method provided by the IPCC has been used for this purpose. In section 2.2, the authors present a method estimating stock change factors instead of soil organic carbon changes. But, the results in Table 4 for the stock change factors are not convincing that the methods are consistent, and the text seems unclear with discussion about larger differences with the IPCC 2019 values, which were updated by the authors and should be more accurate – I would think. Why not estimate the change in soil organic carbon for a direct comparison with the Tier 2 method instead of the stock change factors? Also, the placement of these results after the discussion seems odd, and conventionally would be presented in the results section before discussion.

Answer to 4.: Thank you for the recommendation. We placed the comparison within the result section. For a Tier 1 approach the change in soil organic carbon must be calculated based on a reference stock. The default stocks given by the IPCC have a very low spatial resolution (42 coarse climate zones and soil type specific values), which lead to additional uncertainty when directly comparing changes in soil organic carbon. As our aim is to isolate the management impact on SOC stocks, the stock change factors seem to us more informative than the absolute levels (which mix the uncertainties of the SOC stocks and the SOC change by management). We added more detail to the discussion of the strong deviation between the 2006 and 2019 default stock change factors.

It now reads:

LN 481-495: "In comparison with default stock change factors of the IPCC guidelines, our model estimates a stronger decline of SOC stocks (Table 3) for almost all regions. Tropical soils might suffer from low C input rates due to large yield gaps (Global Yield Gap and Water Productivity Atlas. Available URL: www.yieldgap.org (accessed on: 03/01/2022)) and high shares of residue removal and burning in lower-income countries (Smil, 1999a; Williams et al., 1997; Jain et al., 2014). Yet, even when comparing our etimates to the low-input stock change factors of the IPCC, our SOC loss is roughly twice as large as the revised 2019 IPCC default values, while it shows very good agreement with the older default values from IPCC (2006). Don et al. (2011) estimated SOC losses for tropical soils of around 25% on average corresponding to a stock change factor of 0.75, but also reported a wide range of measured SOC changes from -80% to +58%. Fujisaki et al. (2015) however found much lower loss rates of around 9%, attributing the difference to the different time period length since the conversion to cropland. As our results do not specifically account for cropland age and most of the cropland is older than 20 years (as assumed for the default IPCC Tier 1 stock change factors) our stock change factors have to be lower by definition following the steady-state assumption that cropland will continue to approach a new equilibrium. For the same reason, our estimates for temperate regions might be lower than both IPCC (2006) and IPCC (2019) default values. With the production-increasing impact of irrigation and fertilization on carbon-poor dryland soils, SOC under cropland can also be higher than under PNV with stock change factors above 1 (see Fig. 2(d)), but these areas are much smaller than where the stock change factors are well below unity."

5. Figure 3 shows results from making certain practices constant from 1975 to 2010.The authors state around line 315 that the effect of no-till has been strong since 1990,but the effect seems minor and may not differ statistically from the histManagement with uncertainty. The conclusion about the importance of residue seems most important here.

Answer to 5.: Indeed, the effect of residues management seems most important. We now highlight this at several places in the discussion and removed the statement about tillage, e.g.

- LN 349-350: "Both the constManure and constTillage scenarios show only small deviations from the historical agricultural management values with $0.15\,\mathrm{GtC\,yr^{-1}}$. The effect of no-tillage only becomes discernible from 2000 onwards."
- LN 448-452: "The substantial impact of changing management practices through time is indicated by the development of our estimated stock change factors (see Table 3)) as well as by the time trend of the SOC debt (see Fig. 2(a)). Residue management has changed over the last decades, especially with the phasing out of residue burning practices in several regions and increased general productivity, showing a clear impact on SOC dynamics — underlining the importance to account for these effects in soil carbon modeling."
- LN 533-538: "Herzfeld et al. (2021) also consider historical management trends for fertilizer and manure inputs as well as on residue removal rates and tillage systems, but cannot reproduce the substantial increase in agricultural productivity over the last decades. Still, they find that compared to no-tillage systems, residue management has much larger potential to affect the strength of their projected future global cropland SOC decline. This is consistent with our finding that increasing SOC inputs from above-ground residues had the strongest effect on the slowing-down of the SOC debt increase (Fig. 5)."

6. The authors suggest that there needs to be a circular flow with food supply chain back to soils. They assumed that none of the waste from supply chains are returned to soils (near line 300) but this seems incorrect. Municipal waste and materials are amended to soils in many regions of the world although maybe there are no data on these amendments? If this is the issue, the authors could mention that they are making a conservative assumption due to lack of data.

Answer to 6.: Indeed. We now write:

LN 430-434: "While our data set, by including crop residues and manure, likely the largest carbon inputs to soils, it does not account for a list of minor carbon inputs from cover crops, agroforestry, green manure, weed biomass as well as application of human excreta, sewage sludge, processing wastes, forestry residues or biochar. Including these sources would correct our estimates upwards and bring our estimates closer to the IPCC stock change factors (see Sect. 3.4.1). Unfortunately, data on the quantity of these inputs is very scarce and does not exist with global coverage."

7. The authors evaluate the sensitivity of the Tier 2 model for tree litter with methods in Section 2.4.3. The Tier 2 model divides litter into metabolic and structural components,and the authors have averaged lignin to nitrogen across tree components as input to the model. But forest also include deadwood and should be separated from other forest litter to model decomposition. Did the authors add a deadwood pool? I question if this model is appropriate

for forest if deadwood is not modeled separately.

Answer to 7.: We follow this suggestion and added more detail on the litterfall of the natural vegetation. Using additional information from LPJmL4, we use plant-functional type specific parameterization for the natural litterfall and split up the litterfall into a wood, leaf and root fraction adding different parameterizations for these (see Sect. 2.1.1 and Table 1). We are not be able to add a deadwood pool, since that would require additional parameterization of turnover dynamics for this new pool. Deadwood pools are also not treated explicitly in many DGVMs but are considered part of litter pools, which distinguish woody from non-wood litter pools. The separation of litterfall into wood and soft tissue fluxes will thus add similar stock detail as in DGVMs.

8. Recommend that the authors provide more explanation for Equation 9, which determines the residue amount of C, and is a key driver of the carbon change. Harvest index is the proportion of plant biomass that is harvested, but the authors are multiplying the harvested crop product by the harvest index. But the conventional approach is 'harvested crop production divided by the harvest index' to determine the total biomass and then subtract the harvested amount to estimate the residue. The authors are accounting for double harvesting and fallow in this calculation, which I agree is important, but some further explanation is needed about the calculation to understand how residue carbon is estimate from crop production, harvest index and area.

Answet to 8.:

Crop residue production (ton dry matter residues) are calculated based on a linear function with positive intercept and a yield-dependent slope. This allometric function accounts for the fact that higher-yielding crops often have a lower harvest index than low-yielding crops. In our revised manuscript, we rewrote the Eq. 9 to make the functional form more visible to

$$\begin{aligned} AGR_{i,t,cg} &= CA_{i,t,cg} \cdot \left(Y_{i,t,cg} \cdot r_{cg}^{\text{ag,prod}} + MCF_{i,t} \cdot r_{cg}^{\text{ag,area}}\right) \qquad \text{and} \\ BGR_{i,t,cg} &= \left(CA_{i,t,cg} \cdot Y_{i,t,cg} + AGR_{i,t,cg}\right) \cdot r_{cg}^{\text{bg}} \quad . \end{aligned} \tag{2}$$

Where one can also simplify $CA \cdot Y$ to a production value, as we did in the last version of the manuscript; keeping area and yield explicit in the new version however makes the dynamic harvest index more explicit.

We also added more detail to Sect. 2.3.2 on "Crop and crop residues production" and units for the three crop-group cg specific ratios: * above-ground residues to harvested biomass $r_{cg}^{\text{ag,prod}}$ in $(tDM\,ha^{-1})(tDM\,ha^{-1})^{-1}$ * above-ground residues to harvested area $r_{cg}^{\text{ag,area}}$ in $tDM\,ha^{-1}$ * below-ground residues to above-ground biomass r_{cg}^{bg} in $tDM\,tDM^{-1}$.

9. For the Tier 1 method, IPCC divides the reference carbon stocks by climate and soil types. Did the authors also divide the grid cells by climate and soil because only climate is mentioned in the text? And, I found a diagram in Figure 5.1 in the IPCC report that divides low, medium and high input categories. Did the authors use this diagram to classify the input? It is not clear if the authors use the diagram or developed their own. If they developed their own, is it consistent with the IPCC factors?

Answer to 9.: We do not calculate SOC stocks and SOC changes based on the Tier 1 method. Instead we calculate stylized stock change factors based on the Tier 1 method, and compare them to our Tier 2 approach. Soil types therefore do not influence these factors, as the management is multiplied on top of the stocks. Our approach therefore stays fully consistent with the IPCC factors.

10. What is 'resp' is 'area reduction resp' on line 110? This sentence should be revised to improve readability. I also found other sentences that were difficult to read or missing words in some cases, but did not make a list during my review. Suggest a careful review before final publication.

Answer to 10.: We changed in LN 137-138:

"with A_{lu} being the land-use type specific areas, AR_{lu} and AE_{lu} the area reduction resp. area expansion of the two land-use types"

to

"with A_{lu} being the land-use type specific areas, AR_{lu} the area reduction and AE_{lu} the area expansion of the two land-use types."

11. I found the Tier 2 method in Chapter 5 of Volume 4 of the 2019 IPCC report, and would suggest that the authors cite this chapter rather than the entire 2019 IPCC report,which has 5 volumes. This would make it easier for others interested in the study to find the method in the IPCC report.

Answer to 11.: We took up the reviewers comment and changed from citing the whole IPCC guidelines Vol. 4 (Eggleston et al., 2006) and their refinements (Calvo Buendia et al., 2019) to referencing the Cropland chapters (Lasco et al., 2006; Ogle et al., 2019). Note that not all references changed, since part of the methodology is stated in other chapter (e.g. on "Generic Methodologies Applicable to Multiple Land-use Categories" - Chapter 2).