Response to B. Guenet of the paper entitled "Soil carbon response to land-use change: Evaluation of a global vegetation model using observational meta-analyses"

Ref.: bg-2016-161

Below are the reviewers suggestions (***bold italic font***) and our responses to each point (normal font). In some of our responses, we have cited text from the revised manuscript (*italic font*).

The study by Nyawira et al., it is nice attempt to evaluate a large-scale model using meta-data. This study focus of LUC effect on SOC dynamic but the methodology presented here might probably use in another context (compare long term and short term effect of atmospheric CO2 increase on NPP for instance).

Thank you for your comments. We are happy that you find our method useful for other applications.

The paper is generally well written but the methods section needs to be a bit more de- tailed to be useful to any modeller interested in applying the method. In particular, how the idealized simulations were sampled for non-equilibrium cases.

We have expanded the simulation setups section (section 2.3) and the model-data comparison approach section (section 2.4) to make the method easily understandable to readers interested in doing similar analysis.

Another missing point is how tillage is represented in the model and in particular its effect on SOC.

The model does not represent other crop management practices such as tillage. We discuss this and the implications of tillage for SOC in section 4.2.1.

The take home message I found in the paper is that using observed GPP and with harvest representation the model fits better with the data. These results are not very surprising except if different approaches has been tested but not presented. Nevertheless, the main interest of the paper to my opinion is methodological. Therefore I suggest to add the scripts used in supplementary material to facilitate the use of the method by other.
Finally I suggest accepting the paper with minor revisions.

The scripts and the data used for our analysis are archived by the institute and are available upon request, in accordance with the guidelines of good scientific practice of the Max Planck Society. We have added this information in the acknowledgement.

Minor comments: P4 l4: I don't understand this part. If you did idealized simulation using one vegetation type per grid, why give those details about grid cells with more than one vegetation type?

In this paragraph we describe why we use idealized simulations and also why realistic LUC simulations with a mix of vegetation types cannot be used for evaluating DGVMs. We have re-written the paragraph to make this point clear.
"We perform idealized LUCs in which only one vegetation type covers the entire globe and which is subsequently transformed to another type. The idealized simulations approach prevents interference of soil C changes that occur due to different types of LUCs occurring simultaneously in a grid cell or due to sequences of LUC over time. Such interferences occur in realistic LUC simulations. Here, most grid cells in the globe contain a mixture of different vegetation types and at a given year different LUCs may occur. For example, part of the forest in a grid cell may be converted to crop and at the same time part of the grass be converted to crop. Many DGVMs do not separate the soil C for the different PFTs and have one soil C pool for all the PFTs. Those that separate the soil C, e.g. JSBACH, typically add the soil C of the old PFT to the new PFT after LUC. Therefore, soil C change resulting from a specific LUC cannot be obtained using such realistic simulations. The idealized simulations approach used in this study

ensures that starting with equilibrium soil C from one land use then changing to another land use, the resulting soil C change can be associated with the specific LUC."

P4 l16: If this is the case here the word "usually" is not necessary.

We have removed the sentence.

P4 l29: When LUC is performed it is not very clear how the new vegetation type is split into the different PFTs?

We have expanded the description of the initial PFT distribution (paragraph 2 of section 2.3) as well as the changes in PFT distribution with LUC (paragraph 4 of section 2.3).
" To perform the LUCs in Table 1, starting from the obtained equilibrium state for each land cover, we use JSBACH land use transition matrices as described in Reick et al. (2013). We modify the transition matrix to perform the respective LUC transition in all the grid cells in the entire globe at the first simulation year with no other LUC transitions during the rest of the simulation time. The distribution of PFTs for the target land cover map is taken from the idealized land cover maps described before, with the exception that the LUC transition to pasture assumes an equal distribution of C3 and C4 pastures (following the default JSBACH assumptions). These simulations represent the standard model version results."

P5 l6: The product of degradation of this new pool goes back to litter (to simulate composting for instance) or is this OM totally exported?

The harvest pool decays solely into the atmosphere within one year. Additional organic matter that may be transported back to the field in form of manure is captured implicitly by the biomass left in the field after harvest. We have added this in the text.

P6 l4: It is quite a big assumption to fix this β value since it is likely controlled by several factors (Matthieu et al., 2015). A sensitivity analysis to this parameter might be useful in the supplementary materials.

We have removed the section on scaling soil carbon in the manuscript.

Fig. 2: In the title: it is not "equilibrium" anymore right?

We have changed the figure caption.

Tab. 4: It seems that to force the model with observed GPP and to better reproduce harvest improved the model-data agreement, what about doing both?

This is a good suggestion. However, we did not perform this simulation due to technical issues related to how we trigger the harvest events based on phenology in the model. Harvest in the regions with a well defined growing season (e.g., temperate regions) is done at the end of the growing season. In the observation driven simulation, the prescribed LAI seasonality had drops during the growing season that would lead to constant harvesting during the growing season, which would introduce an artificial bias between the model-driven and the observation-driven simulations. However, since the observation-driven and jsbach-driven simulations results for the different LUCs were similar this simulation would not change the conclusions in the paper.

Tab. 5: Comparison with data might be useful in particular to see the error associated to autotrophic respiration in the model.

The NPP values shown in Table 5 are model inputs and not outputs. The obs_drvn simulation values represent the NPP obtained from GPP using ratios. We have added a discussion in the supplementary to discuss uncertainties associated with scaling the GPP to NPP.

Response to D. Schepaschenko of the paper entitled "Soil carbon response to land-use change: Evaluation of a global vegetation model using meta-analyses"
Ref.: bg-2016-161

It seems our point by point responses at the quick report stage of the unfortunately were not passed to the reviewer. The requested changes were accounted for before the paper was published on discussions. Below are the reviewers suggestions (***bold italic font***) and our responses to each point (normal font) based on our changes prior to publication as discussion paper and additional changes in response to the other reviewers' comments.

This study demonstrates an approach for evaluating perfomance of DGVMs to account soil carbon changes following land-use change. It is important to estimate how far DGVM simulations are from the reality and which model setup is closer to observation. The article has rich discussion section, where most of the questions are covered. The paper is well written. Finally I suggest accepting the paper with minor revisions.

We are happy that the reviewer finds the evaluation of soil carbon changes in DGVMs useful.

Application of universal function for scaling soil carbon pool to 100 cm (page 3, line 15) could introduce substantial bias. In many cases carbon pool changes in top layer only, but you propagate observed value down to 1m and increase therefor the magnitude. That might be the reason of having higher amplitude in meta-data. I would either suggest use soil map and soil specific equations or make the analysis for the top layer only if no observation for deep soil layers available.

The reviewer is correct that the scaling of the meta-data with depth is quite uncertain. We didn't find reliable land-use specific functions for scaling soil carbon densities. Therefore, we have removed the scaling of the meta-analysis and discussed the depth issue as a major challenge in the model-data comparison.

Selection of area/climate for simulation is important. By including extra area (where LUC not going to happen) or excluding potential LCC area, you might bias the overall estimation. Authors suggest three different extends and each is not ideal:

1. ***Entire vegetated area of the land surface (too big. Low chances e.g. for forest on permafrost to be converted to cropland)***
2. ***Area where LUC has taken place historically (too narrow, LUC might came to new places, e.g. tropical deforestation)***
3. ***Where meta-data were available (even more narrow)***

I would suggest to overlay PFT map with climate one and define climatic patterns where one or another PFT can appear. This would cover all current and potential LUC.

This comment from the reviewer indicates that our descripition on how we selected model regions for comparison with the meta-data as described in section 2.4 was not clear. We actually did not suggest that the meta-data can be used to evaluate the entire vegetated land surface as pointed out by the reviewers' approach 1. We have revised this section in the manuscript and added a more detailed explanation on why we used the other two approaches for selecting the regions for comparison (Approach 2 and 3). The reviewer is correct that these two approaches may not be representative of regions where LUC has not taken place historically. However, the meta-data on LUC exist only in regions where LUC has taken place historically. The only approach not introducing biases is thus to assess regions of LUC, since future LUC may move to regions where the meta-data are not representative. Our study presents a method to identify suitable models that can also be used to project soil carbon changes due to future LUC. However, such projections are beyond the scope of our study, which focusses on model evaluation.

Minor comments

Page 7, line 8. In fact, forest might have lower NPP compare to cropland, but most of the dead matter come to soil surface where decomposition is slow

We have re-written the sentence to show that we are explaining the reasons for the simulated increase in the model. We have further clarified that the change in soil carbon is driven by on average higher productivity.

Page7, line 10. Here could be different explanation. Soil respiration is higher in cropland compare to grassland because of the tillage. That is why having similar NPP grassland accumulate more carbon.

We agree with the reviewer that tillage leads to more soil carbon losses in croplands compared to grasslands. However, in this sentence in the manuscript we the explain response of our standard model simulation, which does not include tillage. We demonstrate the effects of crop management practices with the simulation accounting for crop harvesting and further discuss the implications in section 5.1.1.

Page 9, line 10 "Without accounting for crop harvesting" makes sense only to demonstrate how DGVMs are far from reality while not taking into account such evident things like management and disturbances.

We have rephrased this sentence to indicate that the results are for the sensitivity simulations neglecting burning in our standard model simulation, which does not account for crop harvesting.

Page 9, line 10 "switching off disturbances in grass... leads to the right direction of soil carbon change" I hope we aim is to describe the reality with the model, but not just have a similar estimation. Disturbances exist. If result is better without disturbances, then the model makes mistake in its different part.

We agree with the reviewer that the phrasing of this sentence suggests that we switch off burning to get to the right direction of change. We have rephrased the sentence to clarify that this result represents a sensitivity simulation. In addition, we have included an additional sentence to explain that we aim to show the choice of the vegetation types affected by disturbances in DGVMs has an influence on the soil carbon response to LUC.

Page 10, Line 1. Grassland and cropland NPP generally larger compare to forest in temperate region also because they allocated on best locations (soil, slope, etc). However if you try to convert existing forest to grassland or cropland you might not get increase of NPP.

The meta-data include local-scale measurements that are mainly done using paired plots designs; hence such sub-grid scale heterogeinities are accounted for in the meta-data. Therefore, an assessment of the soil carbon response to LUC associated with such heterogeinities is beyond the scope of our study.

Response to E. Marin-Spiotta of the paper entitled "Soil carbon response to land-use change: Evaluation of a global vegetation model using observational meta-analyses"
Ref.: bg-2016-161

Below are the reviewers suggestions (***bold italic font***) and our responses to each point (normal font). In some of our responses, we have cited text from the revised manuscript (*italic font*).

Thank you for your comments, which have helped us greatly in improving our manuscript. We would like to clarify that there were some changes requested by one of the reviewers prior to the publication of the manuscript on discussion. Therefore, some of the line numbers in this review refer to the older version of the manuscript at the quick report stage. To be consistent, we have added the new line numbers that match the current manuscript on discussion in parenthesis.

This study uses published results from temperate and tropical meta-analyses that calculate mean responses of soil carbon change to different land-use transitions from field measurements of paired plots to better constrain estimates of belowground response to land-use change at a global scale simulated by a dynamic global vegetation model. To my knowledge, this is the first time global syntheses of data from published meta-analyses have been used to compare to results from models. The research takes advantage of a large effort to synthesize global temperate and tropical data on soil C to estimate the response of soil C stocks to major land use transitions.

Overall, the writing could be improved to more clearly describe the modeling approach and to distinguish it from past efforts. Some sections have minimal text (for example, description of the observational data and the meta-analyses approach), and while the attempt to be concise is appreciated, more information would make it easier for readers to understand and attempt to replicate the approach.

We have expanded section 2.3 and 2.4 to make it easier for readers interested in replicating the results (see response to B. Guenet). We have also expanded the section describing the meta-analyses (section 2.1) as follows:
"In this study, we use results from the meta-analyses by Poeplau et al. (2011) in the temperate regions and Don et al. (2011) for the tropical regions including 95 and 385 published studies, respectively. The published studies include sites from different countries in the tropics and temperate regions. The site studies were conducted using two main experimental designs: paired plots comparing soil C between two adjacent sites with different land use types, and time series where the soil C of a particular site was monitored overtime after LUC. The paired plot approach is used to construct chronosequences comprising of plots with different ages after LUC that use one of the plots, with the prior land use, as the reference site. The paired plot based approach goes a long with a higher methodological uncertainity in the data due to differences in the inherent soil properties such as texture between the plots, which affect the response of soil C to LUC. In contrast, the time series observational data are without such uncertainties, but very few time series are available to investigate the response of soil C to LUC. In calculating the soil C changes across the different sites, the reference site was always assumed to be in equilibrium.
The meta-analyses defined the following criteria for including the site studies: (1) climate conditions, age of the current land use, and the relevant site characteristics such as soil type, texture and land-use history had to be provided, (2) studies on organic and wetland soils were not included and (3) for paired plots the sites had to be adjacent to each other to reduce uncertainties due to the spatial variability of soil properties unrelated to the LUC (Don et al., 2011; Poeplau et al., 2011). Any studies that did not match any of the criteria were excluded in the compilation. The soil bulk densities were used to calculate the soil organic carbon in Mg/ha. Mass correction was applied to account for changes in density with depth (Ellert and Bettany, 1995). In addition, Poeplau et al. (2011) used different variables, such as climate, time after LUC and the clay content, to derive carbon response functions

(CRFs) describing the temporal response of soil C to LUC for the temperate regions. The response functions include general CRFs that account for only time after the LUC and specific CRFs that account for other site properties. Table 1 shows the LUCs represented in the two meta-analyses that are included in our study."

The discussion dives directly into details of the model but would benefit from an overall summary highlighting the main findings of the paper and being organized around the take home messages of the research. An effort to frame the discussion in a bigger context will help identify novel insights to a broader audience who may not be familiar with the modeling approach but is still very interested in the findings. For example, consider starting the discussion with section 4.1.4 (which has a great discussion of scale between the models and the field observations) then going into the details of the crop harvest and fire, then discussion of the challenges (current section 4.2.1).

We have re-organized the discussion section to focuss on three key aspects that are important for the broader audience. (1) The general approach for evaluating DGVMs against the meta-analyses, (2) the causes of model deviation from the meta-analyses that we identified for the DGVM JSBACH, and (3) the challenges that are involved in model-data comparison.

The use of the term "meta-data" in the title and throughout the paper to represent results from a meta-analyses (a specific statistical test that calculates differences (effect size or response ratio) between data points) is confusing and inaccurate (and distracting) as this term has a different formal definition ("data that describes other data"). To avoid unnecessary confusion, please use an alternate term, such as "field data", "observations", "observational data" or "results from meta-analyses." The term meta-data in the paper is used to refer to meta-analyses (published studies using the specific statistical approach), to the results of these analyses and to general syntheses of published data, further adding to confusion, as these are not the same.

We have removed the word meta-data and adopted meta-analyses and observational data in our manuscript.

In a few places, it is unclear how data used in the model simulations is then related to the land cover types and information associated with the observations of soil carbon change. For example, how are the different plant functional types, especially the different forest PFT, related to the 4 idealized land use classes? Given that the observational data used in this paper is heavily biased towards tropical sites (from Don et al. 2011 vs Poeplau and Don 2015), it is expected that the land cover description of the sites in the published literature do not match the PFTs at the global scale in the DGVM.

This is a good point. The land cover type description is indeed an important factor when comparing soil carbon changes with meta-analyses. In JSBACH there are four forest PFT distinguished in terms of their phenology (broadleaf and decidous) and location (tropical and extratropical). Our grid cell selection criteria ensures that the selected regions include only the PFTs existing in the particular regions. Therefore, the comparison of the simulated soil C changes with the data by Don et al 2011 includes regions with only tropical PFTs, while the comparison with Poeplau et al 2011 includes extratropical PFTs. We added the paragraph below in Section 2.3 to clarify how we derive the distribution of the different PFTs contained in each land cover map.

"We create idealized land cover maps for four vegetation types; forest, crop, grass and pasture. In these cover maps the entire globe is covered by each of the four vegetation types. The regions where one of these vegetation types does not exist are masked out in our comparison of simulated results to the meta-analyses (see section 2.4). Each land cover map consists of several PFTs: Forest land cover contains evergreen and broadleaf PFTs in the tropical and extratropical regions, while crop, grass and pasture land cover contains both C3 and C4 PFTs. To create the idealized land cover maps we start with a present day JSBACH land cover map obtained by remapping observed vegetation distribution into PFTs (see Friedl et al. (2010) and supplementary material section S1). In the grid cells where two PFTs belonging to the same vegetation type already exist, e.g., in a grid cell with both tropical deciduous and tropical

evergreen from observed vegetation distribution, we scale the cover fraction to the entire grid cells based on their relative distribution."

In addition, where do the plant productivity measurements used in the model come from and how do these relate to the types of vegetation and their growth rates from the observational soil carbon studies?

Plant productivity is either simulated directly by JSBACH (standard set of simulations) or prescribed from observations. Section 2.3 describes where these measurements come from (flux net measurements extended globally using machine learning algorithms). See previous response on how the PFT distribution within a land cover type is derived.

How do the model simulations address uncertainties in field soil C measurements? For example, the values given in section 3.1 are averages with some associated error. What is the size of this error, how does this variability affect the carbon response functions, and how then do these influence modeled results?

You are correct that field measurements can be quite uncertain. However, in our comparison we do not force the model with the observed soil C from the meta-analyses, but just compare the two. Because we do not use observed soil C as forcing in our model, there are no propagated errors from the meta-analyses to our simulated results. The carbon response functions used are derived from the meta-analyses. The standard deviation provided in our comparison provides a measure of the spread in the considered regions and sites. Therefore, assessing the error associated with the meta-analyses is outside the scope of our study.
We have added a sentence to discuss uncertainties associated with the methodological designs used to obtain the observational-data in the meta-data analyses (section 2.1).

Page 1, line 23: Soil C changes with LUC are not only influenced by differences in inputs, but also outputs, and alteration to processes that store C in soils.

We have re-written the sentence as follows;
"Soil C changes due to LUC are caused by changes in soil C inputs and outputs when one vegetation type is replaced by another. Changes in soil C inputs stem from differences in litter quality and quantity, while the changes in outputs stem from alteration of soil decomposition processes that govern stabilisation of carbon in soils."

The last paragraph of section 4.1.1 in the discussion briefly starts to address other factors that can influence soil C decomposition that are not included in the model. Some further discussion on how the focus on plant litter chemistry and climatic variables as controls on C cycling that is the basis for the biogeochemical component of the model and the absence of other mechanistic controls on soil C turnover and how they may influence differences between simulation results and observational data would enhance the paper.

We have extended the discussion in section 4.1.1.
"The carbon model used in this study simulates soil C based on the plant chemistry and climate. Recent studies have shown that the inclusion of microbial dynamics and priming processes in biogeochemical models can improve model agreement with observations (e.g., Wieder et al., 2013). As these processes are different across land-use types, the inclusion of such processes in future generation of DGVMs may lead to improved simulated soil C response to LUC."

Page 4, line 13-20 (line 2-10): This paragraph discusses grid cells with only one vegetation type and also proportions of grid cells undergoing different land use transitions from different vegetation types. Please clarify which approach was taken in the paper and distinguish between old approaches (for example, additive soil C pools with LUC) and the new one proposed in this study.

We have re-written the paragraph to clarify that we use idealized and not realistic LUC simulations. The goal of this paragraph is also to discuss why we need to do idealized simulations and not use realistic LUC simulations in evaluating DGVMs.

"We perform idealized LUCs in which only one vegetation type covers the entire globe and which is subsequently to another type. The idealized simulations approach prevents interference of soil C changes that occur due to different types of LUCs occuring simultaneously in a grid cell or due to sequences of LUC over time. Such interferences occur in realistic LUC simulations. Here, most grid cells in the globe contain a mixture of different vegetation types and at a given year different LUCs may occur. For example, part of the forest in a grid cell may be converted to crop and at the same time part of the grass be converted to crop. Many DGVMs do not separate the soil C for the different PFTs and have one soil C pool for all the PFTs. Those that separate the soil C, e.g. JSBACH, typically add the soil C of the old PFT to the new PFT after LUC. Therefore, soil C change resulting from a specific LUC cannot be obtained using such realistic simulations. The idealized simulations approach used in this study ensures that starting with equilibrium soil C from one land use then changing to another land use, the resulting soil C change can be associated with the specific LUC."

Page 5, line 3 (Page 4, line 26): What are the four idealized land use cases? These could be identified here or earlier in the description of the observational data and the meta-analyses.

We have clarified that the four idealized land use cases refer to crop, forest, grass and pasture.

Minor comments

Page 1, line 5-11: Consider the following sentence reorganization: "Our simulated results show model agreement with the observational data on the direction of changes in soil carbon for some land-use changes, although the models generally estimated smaller magnitudes of change. The conversion of crop to forest resulted in a simulated soil carbon gain of 10% compared to a gain of 42% in the data, whereas the forest to crop change resulted in a simulated loss of -15% compared to -40%. The model and field data disagreed for the conversion of crop to grassland. The simulations estimated a small soil carbon loss (-4%), while field data indicate a 38% gain in soil C with the same land-use transition. These model deviations from the observations are substantially reduced by explicitly accounting for crop harvesting and removing burning in grasslands from the model."

We have adopted the suggested changes and re-written the section as follows;

"Our simulated results show model agreement with the observational data on the direction of changes in soil carbon for some land-use changes, although the model simulated generally smaller magnitude of changes. The conversion of crop to forest resulted in soil carbon gain of 10% compared to a gain of 42% in the data, whereas the forest to crop resulted in a simulated loss of -15% compared to -40%. The model and the observational data disagreed for the conversion of crop to grasslands. The model estimated a small soil carbon loss (-4%), while observational data indicate a 38% gain in soil C for the same land-use change. These model deviations from the observations are substantially reduced by explicitly accounting for crop harvesting and neglecting burning in grasslands in the model."

Page 1, line 17: suggest deleting: "(hereafter meta-data)" as this is an incorrect use of this term.

We have deleted this and adopted meta-analyses throughout the manuscript.

Page 1, line 18: add references to the meta-analyses here

We have added an example reference for the meta-analyses here. The full list of references is provided later in paragraph 3 in the introduction where we discuss the meta-analyses.

Page 2, line 7: Rewrite: "Despite the dependence of the soil carbon response to local conditions of soils, climate and management practices, regional and global syntheses of published data can be useful to aggregate local-scale measurements on soil carbon changes and estimate mean responses to different

LUCs using a meta-analyses approach."

We have re-writen the sentence as suggested.

Page 2, line 8 and 10: Here meta-data should be replaced by meta-analyses.

See our response to the terminology concern.

Page 2, line 10 and Page 3, line 7: what is meant by "harmonize"a temperate ? Can you use another term?

We have re-written the sentences and removed the term "harmonize".

Page 2, line 15: Marin-Spiotta and Sharma (2013)'s work did not use a meta-analyses approach

We have removed the reference from the paragraph describing meta-analyses.

Page 3, line 5 (line 3): replace "the meta-data" with "results from the meta-analyses"

We have replaced meta-data with results from the meta-analyses.

Page 3, line 6 (line 4): The "quality criteria" sentence structure is awkward. Consider: "These meta-analyses were conducted on paired plots of similar soil type and texture, to reduce uncertainties from hetereogeneous soil properties unrelated to the land-use transition."

See the response to the meta-analyses section (section 2.1).

Page 3, line 30 (line 21): Replace windbreak (check definition) with windstorm.

We have replaced windbreak with windthrow.

Page 4, line 3-5 (Page 3, line 22-25): Consider rewording: "The decomposition rate of litter is controlled by its chemical composition, as determined by its solubility (acid, water, ethanol and non-soluble hydrolysable pools) and the presence of a slow decomposing humus pool." It is unclear from the text whether the humus pool is part of the plant litter or a soil organic matter pool? Does the model include above and belowground litter pools?

We have clarified that all the litter pools and the humus pools in YASSO are treated as part of soil organic matter. We have added two sentences to explain the difference between above and belowground litter.

"Additionally, litter is split into aboveground and belowground where the aboveground litter burns while belowground litter does not. All the litter pools–aboveground and belowground–and the humus pool are summed up in obtaining the total soil carbon."

Page 4, line 26 (line 15): replace "ran" with "run"

We have replaced ran with run in the sentence.

Page 10, line 6: above refers to what?

We have removed the term above in the restructuring of the discussion section.

Figures 1 and 2 are hard to read. Consider that the green, orange and brown colors will be difficult to distinguish for readers with color-blindness, which affects almost 10% of males in many European and English-speaking countries. The grey background also reduces the contrast between the lines, and the lines are too small and hard to see.
Figure 3. See earlier comment about choice of colors.

We have changed the colours in the figures to colours that aid in color-blindness. Additionally we have increased the thickness of the different lines.

Response to T. Pugh of the paper entitled "Soil carbon response to land-use change: Evaluation of a global vegetation model using observational meta-analyses"
Ref.: bg-2016-161

Below are the reviewers suggestions (***bold italic font***) and our responses to each point (normal font). In some of our responses, we have cited text from the revised manuscript (*italic font*).

Nyawira et al. develop a framework to evaluate the response of soil carbon stocks in the DGVM JSBACH to land-use change, using meta-analyses of observations of soil carbon stock change. They find that the baseline model is unable to reproduce the observations for most transition types tested, but that inclusion of crop harvest and exclusion of fires on pasture land notably improves the fit of the model response.
The analysis has been carefully executed and the manuscript is well written. It provides a useful framework for evaluation of soil carbon response to land-use change, which is generally poorly evaluated in DGVMs used to provide land-use emission estimates, despite constituting a substantial part of the overall emission. I am happy to recommend publication, subject to addressing the following minor comments.

Thank you for your comments. We are happy that you find our results worth of publication in Biogeosciences.

Section 2.2 -In order to understand the differences between the various land-use types considered, more information about the different PFT types is required. In particular, how do C3/C4 grasses, C3/C4 pastures and C3/C4 crops differ from one another? I suggest to add a table listing the differences in any important PFT parameters, or other parameters which may be important to the land-use type (e.g. different soil decomposition rates?).

We have added two sentences in this section. In paragraph one we explain the difference in terms of photosynthesis: *"The PFTs differ with respect to their phenology, albedo and photosynthetic parameters; photosynthesis is based on Farquhar et al. (1980) for C3 plants and Collatz et al. (1992) for C4 plants."*
In paragraph two we added a sentence to explain that there is no distinction of the decomposition rates among the different PFTs, but only between woody versus green litter. *"Non-woody litter has the same decomposition rates for all the PFTs, while the decomposition of woody litters depends on the woody diameter".*

Pg. 4, line 21 - I wasn't quite sure here if the forest PFTs had been extended to cover areas not presently covered by forest, or not. Can the authors clarify?

In the idealized simulations we extended the vegetation type, e.g., forest, to cover the entire globe in our simulations. However, our criteria for the grid cells selections based on climate or historical LUC removes regions where the forest do not exist. We have clarified this in section 2.3.

Pg. 5, line 12 - 50% seems a high proportion of crop NPP to be allocated belowground. This will have a substantial influence on the size of the flux lost to harvest and should therefore be discussed in relation to published literature in the discussion section.

We are sorry this was a mis-communication from our part. There is no 50% aboveground and 50% belowground allocation of NPP in the model. The choice of the 50% harvest is based on root to shoot ratio and not only accounts for root biomass, but also for other unharvestable plant parts. We have also added a discussion on the uncertainties associated with this parameter choice. We have the sentence describing the aboveground and belowground allocation and added the following sentence;
The choice to transfer 50% to the litter is approximated from the average root to shoot ratio of several crop types (Extended data Fig. 2 in Gray et al., 2014). The 50% accounts for root biomass, unharvestable parts of the stem

biomass being left in the field and a potential return of carbon to soil in the form of manure.

Pg. 5, line 27 - I think these are grid-cells where just the relevant transition type has taken place (based on Fig. S2), and not where any LUC has taken place at all? This isn't clear to me from the text. Also, on first reading I thought climate and LUC criterion were being applied simultaneously, and it only later became clear that they were being applied separately.

We have re-written the sentence to make it clearer that it is where the relevant transition has taken place. In addition, we have made it clear that the two criteria are independent.

Pg. 6, line 4 - Should beta have units of length? Also, please define the units of d_0 (presumably cm).

We have removed the scaling section from the manuscript.

Pg. 6, line 13 - How do you sample simulated soil C changes over the ages? Do you take a simple mean over the age range in the observations, or do you weight the mean by the number of observations in each age range? I would argue the second is much better, if you have the data to do it.

We have added two sentences to clarify that the second approach is indeed taken.
"For this we use the age represented by each site in the meta-analyses to select the transient years in the simulations to include in averaging the soil C response. We average the soil carbon response over these years and spatially for the selected regions. This average represents the simulated soil C response over the different ages represent in the meta-analyses."

Section 3.2, para. 1 - I think you can be a bit more assertive here in saying that the reason for the results from the crop to grass simulations is fire. That seems to be very clearly demonstrated at the end of the paragraph, and I'm not sure why the section beginning "we suspect" (line 11) is included.

We have re-written the sentence.

Pg. 8, line 14 - Should Fig. S2 be cited here?

We have included the figure citation.

Pg. 9, line 24 - What is meant specifically by "forest floor"? Surface litter?

Correct. We have changed forest floor to surface litter.

Pg. 9, line 27 - Is the larger NPP for forests than pastures in accordance with the literature? Would be good to discuss this briefly with some references.

We added a discussion on this in section 4.2.3.
"For most of the considered regions in the tropics, the larger simulated NPP for forests compared to pastures is consisted with other observations (Smith et al 2012). Murty et al. (2002) associated the observed increase in soil C following conversion of forest to pasture with low initial content of soil C, application of fertiliser and careful management that avoided overgrazing. Table 3 shows low previous land use soil C for forest to pasture compared to forest to crop in the meta-analyses. However, this is not the case for the simulated soil C in the considered regions."

End of section 4.1.4 - Absolutely agree with this sentiment, but shouldn't we then be aiming for a more stringent test than getting within the very large standard deviation that results from this small-scale heterogeneity?

We agree with the reviewers sentiment. However, in our model-data comparison we do not use the standard deviation as a measure of agreement between the simulated results and the meta-analyses. In this part we discuss that the model may not capture spatial variability in soil C changes due to other missing processes, which the meta-analyses may capture.

Pg. 11, line 23 - What exactly is meant by "top soil"?

We have clarified that the top soil refers to the upper 30cm.

Section 4.3 - I agree with the general statement regarding absolute estimates, but the way this section is written seems to imply that JSBACH was successful in capturing the observations in this evaluation, which I feel would be stretching it a bit for several of the transition types, especially grass to crop (based on Fig. 2)

The reviewer is right that our phrasing was misleading. We have re-written the section as follows:
"Even though DGVMs provide land-use-related absolute soil C changes, our comparison focused on relative changes. This is the preferred variable in the meta-analyses because spatial heterogeneity partly cancels in relative terms when two sites in close proximity are compared to each other, as done in chronosequences. Only relative changes allow for deriving robust carbon response functions (Poeplau et al 2011). In the jsb_drvn_harv simulation, the equilibrium changes indicate a decrease in soil C of about 11 kgC m^{-2} and 3 kgC m^{-2} for forest to crop and grass to crop, respectively, in the temperate region. The decrease for forest to crop in the tropics is about 9 kgC m^{-2} (Fig. 1b). The reverse LUCs result in soil C increase of about the same magnitude. Because DGVMs are unaffected by small-scale spatial heterogeneity, their estimates of absolute changes are expected to be more robust than those of meta-analyses and therefore better representative for global carbon responses. After successful evaluation against relative changes, DGVMs can therefore be used to assess large-scale soil C changes in the absolute terms that are relevant for carbon budget estimates."

Table 4 - I'm not clear on the logic of having this table in addition to Table 3. It would seem more helpful to add the obs_drvn and jsbach_drvn_harv data to Table 3 (appropriately adjusted for 30 cm depth), to allow them to also easily be assessed against the observations.

We have removed the model scaling section following concerns raised by another reviewer and included the meta-analyses soil C densities in Table 4 (now Table 3).

Soil carbon response to land-use change: Evaluation of a global vegetation model using ~~meta-data~~observational meta-analyses

Sylvia S. Nyawira[1,2], Julia E. M. S. Nabel[1], Axel Don[3], Victor Brovkin[1], and Julia Pongratz[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany
[2]International Max Planck Research School on Earth System Modelling, Hamburg, Germany
[3]Thünen Institute of Climate-Smart Agriculture, Braunschweig, Germany

Correspondence to: Sylvia Sarah Nyawira (sylvia.nyawira@mpimet.mpg.de)

Abstract. Global model estimates of soil carbon changes from past land-use changes remain uncertain. We develop an approach for evaluating dynamic global vegetation models (DGVMs) against existing observational ~~meta-data~~ meta-analyses on soil carbon changes following land-use change. Using the DGVM JSBACH, we perform idealized simulations where the entire globe is covered by one vegetation type, which then undergoes a land-use change to another vegetation type. We select the grid cells that represent the climatic conditions of the ~~meta-data~~ meta-analyses and compare the mean simulated soil carbon changes to the ~~meta-data~~meta-analyses. Our simulated results show model agreement with the ~~meta-data~~ observational data on the direction of changes in soil carbon for some ~~, but not all~~ land-use changes, ~~while the magnitude of simulated changesis smaller than in the meta-data~~although the model simulated generally smaller magnitude of changes. The conversion of crop to forest ~~results~~ resulted in soil carbon gain of 10% ~~and that of~~ compared to a gain of 42% in the data, whereas the forest to crop ~~to a~~ resulted in a simulated loss of -15% compared to ~~a gain of 42and loss of~~ -40%~~, respectively, in the meta-data. However,~~ . The model and the observational data disagreed for the conversion of crop to ~~grass results in~~ grasslands. The model estimated a small soil carbon loss (-4%), while ~~the meta-data indicate a~~ observational data indicate a 38% gain in soil ~~carbon of 38~~C for the same land-use change. These model deviations from the ~~meta-data~~ observations are substantially reduced by explicitly accounting for crop harvesting and ~~switching off~~ neglecting burning in grasslands in the model. We conclude that our idealized simulation approach provides an appropriate framework for evaluating DGVMs against ~~meta-data~~ meta-analyses and that this evaluation helps to identify the causes of deviation of simulated soil carbon changes from the ~~meta-data~~meta-analyses.

1 Introduction

Global model estimates of land-use-related soil carbon (soil C) changes rely on dynamic global vegetation models (DGVMs). To judge the reliability of DGVMs in simulating past and future changes of soil C, models have to be evaluated against observations. A range of ~~meta-data analysis (hereafter meta-data)~~ meta-analyses on soil C changes following land-use change (LUC) has been published recently, aggregating local-scale measurements to spatial scales potentially applicable to DGVMs (e.g., Guo and Gifford, 2002). Here, we develop an approach for evaluating DGVMs against the ~~meta-data~~ observational data and apply the approach to evaluate the DGVM JSBACH.

A major driver of soil C changes in recent centuries has been LUC. For example, the replacement of natural vegetation with croplands usually leads to soil C loss while the reverse leads to a gain (Guo and Gifford, 2002). Unlike for vegetation, soil dynamics include slower processes ranging from decadal to centennial timescales; hence the carbon response to LUC lags the changes in vegetation carbon. Soil ~~carbon changes following LUC result from the modified vegetation productivity that influences the quantity and quality of litter~~C changes due to LUC are caused by changes in soil C inputs and outputs when one vegetation type is replaced by another. Changes in soil C inputs stem from differences in litter quality and quantity, while the changes in outputs stem from alteration of soil decomposition processes that govern stabilisation of carbon in soils. The response of soil C to LUC depends on the local conditions, such as soil type, mineralogy and texture (Lugo et al., 1986) and on climate influences, such as temperature and soil moisture or precipitation (Marín-Spiotta and Sharma, 2013). Also, management practices can influence the soil C response; for example, Poeplau and Don (2015) showed that planting cover crops during winter and tilling them into the soil as additional carbon input can significantly enhance soil C on croplands. Due to the slow response of soils to LUC, soil C changes from past LUCs continue to have a long-term effect on the global carbon budget (Pongratz et al., 2009).

Despite the dependence of the soil C response to local conditions of soils, climate and management practices, ~~meta-data~~ regional and global syntheses of published data can be useful to aggregate local-scale measurements on soil C changes ~~due~~ and estimate mean responses to different LUCs ~~to mean responses~~using a meta-analyses approach. Over the recent past, several of these ~~meta-data~~ meta-analyses have been published (Post and Kwon, 2000; Guo and Gifford, 2002; Paul and Polglase, 2002; Murty et al., 2002; Laganiére et al., 2010; Poeplau et al., 2011; Don et al., 2011). An advantage of the ~~meta-data~~ meta-analyses is that they apply several quality checks ~~and harmonize~~ to combine and aggregate the local-scale measurements~~to common variables. The meta-data~~. The meta-analyses provide estimates of the average magnitudes of relative changes and additionally the temporal response of soil C to LUC (Poeplau et al., 2011; Poeplau and Don, 2015). ~~Meta-data~~ These analyses have also been used to understand the factors influencing the spatial and temporal variability of soil C changes following LUC. This has been done by correlating variables such as temperature, precipitation and clay content with the soil C changes (Don et al., 2011; Wei et al., 2014)~~and using regression tree analyses~~. However, the applicability of this observational data for global modeling has not been tested so far.

DGVMs are used to study the effects of LUC on soil C globally. They combine information on the past vegetation distribution, climate and LUC data and incorporate various processes to quantify global changes in terrestrial carbon stocks resulting from past LUCs (e.g. Pongratz et al., 2009; Stocker et al., 2011). In addition, by simulating climate and LUC scenarios following the representative concentration pathways, DGVMs are used to make future projections in terrestrial carbon stocks (e.g. Brovkin et al., 2013). However, global estimates of LUC carbon fluxes by different DGVMs show a large spread (Ciais et al., 2013). This spread has been attributed to several factors: the different climate used in driving the DGVMs (Anav et al., 2013), different modeling approaches of LUC (Houghton et al., 2012; Wilkenskjeld et al., 2014), inconsistent definition of land-use fluxes (Pongratz et al., 2014), parameterizations related to fluxes of land-use and land cover change (Brovkin et al., 2013; Goll et al., 2015) and land-management processes (Houghton et al., 2012). In a recent study, Tian et al. (2015) used the same model setup, with the same climate and LUC input data, to quantify global changes in soil C resulting from past LUCs across different DGVMs. They found that these changes differ widely across the models with some models showing almost no change and others showing a large decrease in soil C. Until now, soil C changes resulting from different LUCs in DGVMs have not been compared to observational data compiled by the ~~meta-data~~different meta-analyses. This is because an approach for comparing these changes to the ~~meta-data~~ meta-analyses is still lacking and many of the ~~meta-data~~ meta-analyses have only become available relatively recently.

Our study aims at developing an approach that can be applied to any DGVM for evaluating the soil C changes following LUC against the ~~meta-data~~meta-analyses. We test the applicability of the approach using the DGVM JSBACH and identify what the comparison reveals in terms of model processes. Further, we highlight the challenges involved in comparing simulated results to the ~~meta-data~~ meta-analyses and suggest what can be done to overcome these challenges. This is to our knowledge the first time that simulated soil C response to different LUCs in a DGVM are compared systematically to ~~meta-data~~meta-analyses.

2 Methods

2.1 ~~Meta-data~~Meta-analyses

In this study, we use ~~the meta-data~~ results from the meta-analyses by Poeplau et al. (2011) in the temperate regions and Don et al. (2011) for the tropical regions including 95 and 385 published studies, respectively. ~~These meta-data are based on quality criteria such as soil type and texture of the paired plots. They harmonize the local-scale measurements by applying mass corrections to account for different soil mass due to sampling the soil to different depth increments at the paired sites. Additionally, by accounting for different variables influencing the soil C response to LUC~~The published studies include sites from different countries in the tropics and temperate regions. The site studies were conducted using two main experimental designs: paired plots comparing soil C between two adjacent sites with different land use types, and time series where the soil C of a particular site was monitored overtime after LUC. The paired plot approach is used to construct chronosequences comprising of plots with different ages after LUC that use one of the plots as the reference site. The paired plot based approach goes along with a higher methodological uncertainty in the data due to differences in the inherent soil properties such as texture between the plots, which affect the response of soil C to LUC. In contrast, the time series observational data are without such uncertainties, but very few time series are available to investigate the response of soil C to LUC. In calculating the soil C changes across the different sites, the reference site was always assumed to be in equilibrium.

The meta-analyses defined the following criteria for including the site studies: (1) climate conditions, age of the current land use, and the relevant site characteristics such as soil type, texture and land-use history had to be provided, (2) studies on organic and wetland soils were not included and (3) for paired plots the sites had to be adjacent to each other to reduce uncertainties due to the spatial variability of soil properties unrelated to the LUC (Don et al., 2011; Poeplau et al., 2011). Any studies that did not match any of the criteria were excluded in the compilation. The soil bulk densities were used to calculate the soil organic carbon in Mg/ha. Mass correction was applied to account for changes in density with depth (Ellert and Bettany, 1995). In addition, Poeplau et al. (2011) ~~derived generalized and specific~~ used different variables, such as climate, time after LUC and the clay content, to derive carbon response functions (CRFs) describing the temporal ~~evolution~~ response of soil C ~~following LUCs~~

to LUC for the temperate regions. The response functions include general CRFs that account for only time after the LUC and specific CRFs that account for other site properties. Table 1 shows the LUCs represented in the two ~~meta-data~~ meta-analyses that are included in our study.

2.2 Carbon cycle model in JSBACH

We use the DGVM JSBACH (Raddatz et al., 2007; Reick et al., 2013), the land surface model of the Max Planck Insitute Earth System Model (Giorgetta et al., 2013). Vegetation distribution in JSBACH is represented with 12 plant functional types (PFTs), of which 8 are natural types (4 forest types, 2 shrub types, 2 grass types (C3 and C4)), and 4 are anthropogenic types (C3 and C4 pastures and crops). The PFTs differ with respect to their phenology, albedo and photosynthetic parameters; photosynthesis is based on Farquhar et al. (1980) for C3 plants and Collatz et al. (1992) for C4 plants. The carbon cycle model in JSBACH describes the carbon allocation, the storage in the vegetation and soils, and losses through respiration and natural disturbances. For each PFT, the net primary production (NPP) is allocated to three vegetation carbon pools: the “green pool“ containing living tissues, the ”reserve pool“ containing sugar and starches and the “wood pool“ containing woody material. Each of these pools has different turnover rates, influenced by a background natural mortality and foliage losses due to seasonal and climatic influences. The carbon lost from the vegetation pools via turnover goes into the soils in form of litter where it is decomposed. Following LUC, a fraction of the vegetation carbon goes into litter and the other is released directly to the atmosphere. Additionally, carbon can be lost from the vegetation and soil through disturbances in the form of fire and ~~windbreak~~windthrow.

Decomposition of litter in JSBACH is simulated by the YASSO model. YASSO is calibrated globally based on results from litter bag experiments (Tuomi et al., 2008, 2009, 2011) and has been evaluated on site to regional scale (Karhu et al., 2011; Thum et al., 2011; Lu et al., 2013). Decomposition of litter is distinguished in terms of the solubility of litter in four different compounds (acid, water, ethanol and non-soluble hydrolysable pools) and ~~a~~ an additional slow decomposing humus pool. Each of these pools has ~~different decomposition rates~~ a different decomposition rate derived from the litter bag experiments. The heterotrophic respiration depends on temperature based on a Gaussian model (Tuomi et al., 2008) and on precipitation based on an exponential function (Tuomi et al., 2009). ~~Woody and~~ For all PFTs non-woody litter ~~is treated separately;~~ has the same decomposition rates, while the decomposition of woody litter depends on the ~~diameter of the woody litter~~ woody diameter. Additionally, litter is split into aboveground and belowground where the aboveground litter burns while belowground litter does not. All the litter pools–aboveground are belowground–and the humus pool are summed up in obtaining the total soil carbon. YASSO shows a better correlation of present-day carbon stocks with the Harmonized World Soil Data Base compared to JSBACHs’ previous soil model CBALANCE (Goll et al., 2015). YASSO has been shown to have a lower sensitivity to some uncertain model parameterizations such as the fraction of carbon lost to the atmosphere following LUC (Goll et al., 2015). A detailed description of the implementation of YASSO can be found in Thum et al. (2011) and in Goll et al. (2015).

2.3 Simulation setups

We perform idealized LUCs ~~assuming grid cells with~~ in which only one vegetation type ~~. Historical JSBACH simulations as performed, e.g., for the fifth phase of the Coupled Model Intercomparison Project and the global carbon budget prescribe LUC each year in form of annual transitions. In every grid cell that contains~~ covers the entire globe and which is subsequently transformed to another type. The idealized simulations approach prevents interference of soil C changes that occur due to different types of LUCs occurring simultaneously in a grid cell or due to sequences of LUC over time. Such interferences occur in realistic LUC simulations. Here, most grid cells in the globe contain a mixture of ~~vegetation types,~~ different vegetation types and at a given year different LUCs may occur~~each year~~. For example, part of the forest in a grid cell may be converted to crop and at the same time part of the grass be converted to crop. ~~In JSBACH upon LUC, the soil C of the old PFT is added to the soil C of the new PFT. Moreover, many~~ Many DGVMs do not separate the soil C for the different PFTs ~~but instead~~ and have one soil C pool for all the PFTs. Those that ~~do distinguish soil Cin the different PFTs (as JSBACH) still have a legacy history resulting from different LUCs. Our~~ separate the soil C, e.g. JSBACH, typically add the soil C of the old PFT to the new PFT after LUC. Therefore, soil C change resulting from a specific LUC cannot be obtained using such realistic simulations. The idealized simulations approach ~~prevents the interference of soil C changes due to different LUCs applied at different times~~ used in this study ensures that starting with equilibrium soil C from one land use then changing to another land use, the resulting soil C change can be associated with the specific LUC.

We create idealized land cover maps for four vegetation types; forest, crop, grass and pasture. In these cover maps the entire globe is covered by each of the four vegetation types. The regions where one of these vegetation types does not exist are masked out in our comparison of simulated results to the meta-analyses (see section 2.4). Each land cover map consists of several PFTs: Forest land cover contains evergreen and broadleaf PFTs in the tropical and extratropical regions, while crop, grass and pasture land cover contains both C3 and C4 PFTs. To create the idealized land cover maps we start with a present day JSBACH land cover map obtained by remapping observed vegetation distribution into PFTs (see Friedl et al. (2010) and supplementary material section S1). In the grid cells where two PFTs belonging to the same vegetation type already exist, e.g., in a grid cell with both tropical deciduous and tropical evergreen from observed vegetation distribution, we scale the cover fraction to the entire grid cells based on their relative distribution.

The carbon cycle model in JSBACH can be executed as part of the entire vegetation model or as a stand-alone model isolating the actual carbon cycle simulation from the simulation of other processes, such as photosynthesis and hydrological processes. In the stand-alone mode, the model is driven by net primary production (NPP), leaf area index (LAI), precipitation and 2 m air temperature together with the vegetation distribution. This setup has the advantage that the model can be run for centennial to millennial timescales at low computational costs. ~~Therefore, the idealized land-use simulations can be ran for many years until equilibrium is reached. The input variables used for driving the stand-alone carbon model are usually derived from a simulation with JSBACH.~~

To obtain the inputs for the stand-alone carbon model, we first perform idealized land-use simulations with JSBACH ~~where the entire vegetated area of the land surface is covered by only one vegetation type, being either crop, forest, natural grass ,~~

~~or pasture. The distribution of the different PFTs belonging to the vegetation type (e.g., the four forest PFTs in the all forest simulation) is kept at the relative distribution as prescribed by the global vegetation map from observations remapped to JSBACH PFTs (see and supplementary material section S1). In each of the simulations, JSBACH is driven by~~ with each of the four created land cover maps (forest, crop, grass and pasture). We use observed climate from the climate research unit (CRU) for the years 2001 to 2010 as forcing for JSBACH in these simulations (Harris et al., 2014).

In a second step, we force the stand-alone carbon cycle model using the ~~obtained~~ NPP and LAI obtained from the JSBACH idealized land-use simulations, precipitation and temperature from CRU, and the idealized vegetation distribution used in the JSBACH simulations. ~~The model is run for each of the four idealized land use cases~~ We run the model until the soil C pools are in equilibrium for each of the four land covers. We consider the total soil C in YASSO to be in equilibrium when the relative change in soil C from one year to the next in the grid cell is less than 1%. ~~From this equilibrium , we perform LUC in the first year for each of the LUCs represented in Table 1 with the NPP and LAI replaced with that of the new land use. We run the model again for the new land use until equilibrium is reached~~To perform the LUCs in Table 1, starting from the obtained equilibrium state for each land cover, we use JSBACH land use transition matrices as described in Reick et al. (2013). We modify the transition matrix to perform the respective LUC transition in all the grid cells in the entire globe at the first simulation year with no other LUC transitions during the rest of the simulation time. The distribution of PFTs for the target land cover map is taken from the idealized land cover maps described before, with the exception that the LUC transition to pasture assumes an equal distribution of C3 and C4 pastures (following the default JSBACH assumptions). These simulations represent the standard model version results.

Vegetation productivity as simulated by JSBACH has been shown to be higher as compared to observations (Anav et al., 2013; Todd-Brown et al., 2013). ~~To assess if this bias in~~ We perform additional simulations where we prescribe observed NPP and LAI instead of using the NPP and LAI simulated by JSBACH. This set of simulations serves two purposes: to assess if the model bias in vegetation productivity has an effect on the soil C response to LUC ~~, we perform additional simulations where we replace the JSBACH NPP and LAI with observation-based plant productivity. For this purpose,~~ and to obtain soil C response that is more representative of the observational data in the meta-analyses. In this simulation we use gross primary production (GPP) obtained by extending flux net tower measurements using machine learning algorithms and LAI obtained from MODIS satellite (Tramontana et al., 2016)~~for the different vegetation types are remapped~~. The global vegetation classification used for the GPP and LAI data is not the same as the PFTs classification used in DGVMs. We remap the GPP and LAI into JSBACH PFTs; subsequently ~~NPP is derived~~ we derive NPP from GPP (details in supplementary material section S1). We replace the model NPP and LAI with the remapped ones and run the model to equilibrium for the different land cover maps and LUCs.

From the results shown below, we find that one reason for the deviation of simulated soil C response to LUC from the ~~meta-data~~ meta-analyses could be the lack of explicitly accounting for crop harvest in the model. To account for the influence of crop harvesting in the model, we introduce a crop harvesting similar to what has been previously done in other DGVMs (Shevliakova et al., 2009; Bondeau et al., 2007; Stocker et al., 2011; Lindeskog et al., 2013). We introduce a harvest pool for the crops that decays into the atmosphere on a timescale of one year. This is in contrast to the earlier model version, where all material harvested from crops was transferred to the litter. In the grid cells with an explicit growing season, harvesting is thereby done at the end of the growing season. In the grid cells without an explicit growing season, as occurs in the humid tropics, harvesting is done constantly throughout the year, imitating that each grid cell contains many individual fields that are harvested at different points in time. 50% of what is harvested is kept in the harvest pool while the other 50% goes to litter. ~~This choice is based on the carbon allocation in JSBACH where~~ The choice to transfer 50% ~~of the crop NPP is allocated above ground and~~ to the litter is approximated from the average root to shoot ratio of several crop types (Extended data Fig. 2 in Gray et al., 2014). The 50% ~~below ground and is in accordance with other studies where the harvested material is taken to be the entire above-ground carbon material~~ accounts for root biomass, unharvestable parts of the stem biomass being left in the field and a potential return of carbon to soil in the form of manure.

We perform additional simulations to test the sensitivity of simulated soil C changes towards the representation of natural disturbances, in particular fire. As discussed in section 4, in the standard setup of JSBACH fire affects natural grasslands but not pastures and croplands. Our sensitivity simulations therefore exclude fire on natural grasslands as well. Table 2 summarizes the simulations performed in this study and the names used to represent the respective simulations.

2.4 Model-data comparison approach

The idealized LUC simulations represent soil C changes for the entire vegetated areas including regions where LUC does not take place. Therefore, we need a criterion for selecting the model regions to consider in the comparison to the ~~meta-data~~meta-analyses. We select regions in the model based on two different criteria~~;~~ : climate and LUC applied independently. For the climate-criterion, we select the grid cells that ~~fulfil~~ fulfill the precipitation and temperature range represented by the ~~meta-data~~ meta-analyses in Table 1. Previous studies found that the soil C response to LUC varies spatially due to many factors, among them precipitation and temperature (Don et al., 2011; Wei et al., 2014; Marín-Spiotta and Sharma, 2013). Therefore, the climate-criterion excludes grid cells with different climatic conditions from the ~~meta-data~~meta-analyses, which have potentially different response to LUC. The climate-criterion-based regions are shown in Fig. S1. To assess if the regions obtained using the climate-criterion are representative of regions where ~~LUC has taken place, we select grid cells based on where LUC has taken place historically (LUC-criterion)~~the specific LUC has occured historically, we obtain other regions using the differences between present-day and historical land cover in JSBACH. We select the grid cells where more than 10% of the ~~vegetated area~~ specific vegetation type within the grid cell has undergone LUC~~using a present-day (2005) and a historical (1850)vegetation distribution map~~. These regions represent the LUC-criterion-based regions (regions in Fig. S2). The results shown in section 3 are averages over the climate-criterion-based regions. We also include a comparison of the simulated changes for these two criteria.

The ~~local-scale measurements contained in the meta-data are done at different sampling depths, with most of the measurements at the top layers, while the model simulates soil C at the depth of 100 cm. To compare the mean soil C densities in the meta-data with the simulated soil C densities, we scale the simulated soil C densities for the standard model simulation (jsbachSUBSCRIPTNBdrvn in Table 2) to the upper 30 cm. We apply the depth scaling approach represented in Eq. (**??**)~~

~~as used in and based on .~~

$$C_{d_0} = \frac{1-\beta^{d_0}}{1-\beta^{100}} C_{100},$$

~~where C_{100} is the soil C density at 100 cm depth, β is the relative decrease in the soil C with soil depth set to 0.9786 as in , d_0 is the original measuring depth at which the measurements were done and C_{d_0} is the soil C density at the required depth (30 cm).~~

~~The~~ comparison of soil C changes includes two variables; the relative and absolute soil C changes. We calculate the absolute soil C change by subtracting the soil C of the previous land use from the soil C of the current land use. The relative changes are then calculated with respect to the previous land use. Additionally, we use the generalized CRFs derived from the ~~meta-data~~ meta-analyses in Poeplau et al. (2011) to compare the simulated transient response with the ~~meta-data~~meta-analyses. In this case, only the CRFs with high model efficiency for the crop to grass and crop to forest LUCs are used.

The measurements for the individual observations contained in the ~~meta-data~~ meta-analyses are done at different ages following LUC. Therefore, the observations may not be in equilibrium for the current land use. To account for this, we sample the simulated soil C changes over the ages represented in the ~~meta-data~~meta-analyses, which makes a direct comparison of the simulated and the observed soil C changes more appropriate. For this we use the age represented by each site in the meta-analyses to select the transient years in the simulations to include in averaging the soil C response. We average the soil carbon response over these years and spatially for the selected regions. This average represents the simulated soil C response over the different ages in the meta-analyses. In section 3 we show both the simulated equilibrium relative and absolute changes and the changes obtained by sampling over the ages represented by the ~~meta-data~~meta-analyses.

3 Results

3.1 Soil carbon densities for previous and current land use

Before comparing the simulated changes in soil C against the ~~meta-data~~ meta-analyses in the next section, we present an assessment of the soil C densities prior to LUC. For this we compare the mean soil C densities in the ~~meta-data~~ meta-analyses to the soil C densities for different ecosystems used in bookkeeping models and compiled by the Intergovernmental Panel on Climate Change. For the temperate regions, the previous land use mean soil C of 14.7 kgC m^{-2} for the forests in the ~~meta-data~~ meta-analyses (Table ~~??~~3) is slightly higher than the 13.4 kgC m^{-2} for the undisturbed forest in Houghton et al. (1983), but much higher than the 9.62 kgC m^{-2} in Watson et al. (2000). However, most carbon densities are lower than earlier estimates, such as for tropical forests: 11.7 kgC m^{-2} (Houghton et al., 1983) and 12.27 kgC m^{-2} (Watson et al., 2000); temperate grassland: 18.9 kgC m^{-2} (Houghton et al., 1983) and 23.6 kgC m^{-2} (Watson et al., 2000); and cropland 6-9 kgC m^{-2} (Houghton et al., 1983) and 8 kgC m^{-2} (Watson et al., 2000). A key reason for the lower carbon densities is the limited sampling of only the top-soil in the sites of the ~~meta-data~~ meta-analyses (Table 1), while the soil C densities for the different ecosystems in Houghton et al. (1983) and Watson et al. (2000) are up to a depth of 1 m. ~~The simulated soil C densities for the standard model simulation scaled to the upper 30 cm, leads to lower carbon densities compared to the meta-data particularly for forests and grasses (Table ??).~~

The soil C densities in Table 3 obtained at the model simulation depth are much higher compared to the ~~meta-data~~meta-analyses. The lower carbon densities in the ~~meta-data~~ meta-analyses are again due to sampling only the top soils. Moreover, the model is in equilibrium for each of the considered land use while the local-scale measurements are done at different times. The average soil C densities for the previous land use in the ~~jsbach~~jsb_drvn simulation are higher than in the ~~obs~~t16_drvn simulation for all the LUCs (Table 3). The higher soil C densities result from the generally higher NPP in the ~~jsbach~~jsb_drvn simulation compared to the ~~obs~~t16_drvn simulation (Table 4), which in turn leads to higher litter fluxes (Table 5). Accounting for crop harvesting in the ~~jsbach~~jsb_drvn_harv simulations decreases the litter fluxes (Table 5), which significantly decreases the equilibrium soil C densities. By explicitly accounting for crop harvest in the model the soil C densities for croplands decrease by about 16-24% for the considered regions.

3.2 Simulated changes in soil C for the different land-use changes

Figures 1 and 2 show an increase and decrease in soil C following conversion of crop to forest and forest to crop, respectively, for both the ~~jsbach~~jsb_drvn and the ~~obs~~t16_drvn simulations, consistent ~~with the meta-data~~results from the meta-analyses. In the model this change stems from the higher average productivity in forests compared to croplands for both simulations (Table 4), which leads to higher litter fluxes (Table 5). In addition, woody material in forests decomposes slowly compared to leaf material from croplands. The conversion of crop to grass results in ~~a decrease in soil C~~ soil C decrease, while the reverse leads to a gain in both of these simulations, which is inconsistent with the ~~meta-data~~ meta-analyses (Figs. 1 and 2). ~~This is despite the grass NPP being similar or slightly higher than the crop NPP and their litter fluxes not differing much (Table 4 and 5). We suspect the~~ The reason for this deviation ~~of simulated soil C changes from the meta-data to be related to the~~ is related to litter fluxes or ~~processes~~ process other than soil decomposition leading to soil C losses, because of observational constraints on the other parts of the carbon cycle ~~: soil~~ model: soil carbon decomposition rates in YASSO are calibrated against a wealth of measurements, and the simulations driven by ~~observation-based plant productivity result qualitatively~~ observation based plant productivity (t16 SUBSCRIPTNBdrvn) result in the same deviation as the JSBACH-driven ones (jsb SUBSCRIPTNBdrvn). The deviation may stem from an overestimate of ~~crop~~ cropland relative to grassland litter fluxes, or from an overestimate in the model of non-respiratory processes for grass. Although crop and grass have the same decomposition rates in YASSO, burning in grasslands leads to the loss of more litter carbon to the atmosphere ~~. This results in~~ and shorter turnover time ~~for grasslands~~ (Table 6). ~~In the jsbach~~This explains the simulated soil C decrease when croplands are replaced with grasslands. In the jsb_drvn_nofire simulation, switching off disturbances in grasslands leads to model agreement with the ~~meta-data~~ meta-analyses on the direction of soil C change (Figs. 1 and 2). The inclusion of crop harvesting in the model reduces the litter fluxes for crops (Table 5) and significantly increases the simulated soil C changes for the different LUCs (Figs. 1 and 2).

Although the simulated equilibrium relative and absolute changes for the conversion of temperate crop to forest and vice versa are larger than in the ~~meta-data~~ meta-analyses (Fig. 1), the current land use ~~in the meta-data~~ at the different sites in the meta-analyses may not be in equilibrium. Sampling over the ages represented by the ~~meta-data~~ meta-analyses results in relative changes of about 10% for the ~~jsbach~~jsb_drvn simulation and 25% for the ~~obs~~t16_drvn simulation for the crop to forest conversion (Fig. 2a). These values are lower compared to the 40% relative changes in the considered ~~meta-data~~ meta-analyses and the 53% in Guo and Gifford (2002). For the forest to crop, the relative changes are about -15% for the ~~jsbach~~jsb_drvn and ~~obs~~t16_drvn simulations compared to the -42% in the ~~meta-data~~ meta-analyses (Fig 2a). In both of these simulations, the relative changes following the conversion of crop to grass and vice versa are relatively small (Fig. 2a). Despite ~~meta-data~~ meta-analyses showing an increase of about 8% for a tropical forest to pasture conversion (Guo and Gifford, 2002; Don et al., 2011), our model results indicate a decrease of about -15%. In addition, the absolute changes are smaller compared to the ~~meta-data~~ meta-analyses for all LUCs (Fig. 2b).

Accounting for crop harvesting leads to larger relative and absolute changes in the model. The crop to forest LUC results in an increase of 42%, while the forest to crop results in a decrease of -22%. In line with the ~~meta-data~~meta-analyses, the crop to grass LUC results in an increase of 13%, while the grass to crop results in a decrease of -6% (Fig. 2a). Although these changes are still often smaller than the ~~meta-data~~meta-analyses, they are within the standard deviation represented in the ~~meta-data~~ meta-analyses for most of the LUCs (Fig. 2). Comparing the transient response with the generalized CRFs from Poeplau et al. (2011) and the individual observation points for the crop to grass and crop to forest LUCs, we find that accounting for crop harvesting leads to a stronger soil C response to afforestation in the model and a gain in soil C for the crop to grass conversion, in accordance with the ~~meta-data~~ meta-analyses (Fig. 3).

The climate-criterion (temperature and precipitation) used in the selection of the model grid cells for comparison with the ~~meta-data~~ meta-analyses leads to small selected regions for the temperate regions (supplementary material Fig. S1). Selecting larger regions based on where the specific LUC has taken place historically, helps in judging if the soil C changes for the climate-criterion are representative of soil C changes in LUC regions (supplementary material Fig. S2). We find that averaging the soil C changes over regions where LUC took place historically results in the same direction of soil C changes as the climate criterion (supplementary material Figs. S3 and S4) with slight differences in the magnitudes of the relative and absolute changes (Fig. 4).

4 Discussion

~~Improved process understanding by~~ Our results show that the use of meta-analyses provides an opportunity for evaluating simulated soil C response to LUCs. In this section we discuss general issues related to the applicability of meta-analyses for DGVM evaluation, such as scale-related issues, explore the causes of model deviation from the observational data and identify the challenges involved in model-data ~~comparison~~comparison.

4.1 Application of meta-analyses for DGVM evaluation

DGVMs simulate soil C processes at large spatial scales and are widely used to provide soil C estimates relevant for the global carbon budget (Le Quéré et al., 2015). Reliability on these estimates depends on the ability of DGVMs to correctly represent present-day soil C changes from past LUCs. Site-level simulations are often used to evaluate DGVMs for CO_2 fluxes, such as net ecosystem exchange and terrestrial ecosystem respiration (e.g., Thum et al., 2011). While vegetation processes representing such variables are well represented in the models, soil processes that are important at the local scale, such as soil chemistry, are not represented in DGVMs. Although it may be impossible for a DGVM to capture the soil C response at an individual site, in particular if the site is not representative of a larger region, the model should be able to match average responses across observations covering a wide region. It is therefore possible to evaluate DGVMs at the scales they are meant for.

In our comparison we choose the grid cells, over which we average the response, based on two independent criteria: the climate space covered by the meta-analyses and the regions where LUC has taken place historically. This selection helps in judging how robust the simulated results are and testing if the meta-analyses are indeed representative of regions where LUC has taken place. Our results for the climate-criterion are qualitatively the same as those of the LUC-criterion (Figs. 1, 2, S3 and S4). Small differences occur for forest to crop and crop to forest in the temperate regions, where the LUC regions have smaller changes compared to the regions captured by the climate-criterion (Fig. 4). This suggests that the regions captured by the meta-analyses by Don et al. (2011) and Poeplau et al. (2011) are generally representative of regions where LUC has taken place historically, although the latter may not be representative of whole-ecosystem averages (see Pongratz et al., 2011). Although the site studies in the meta-analyses may have biases towards regions of similar soil and climatic conditions (Powers et al., 2011), the mata-analyses still show a large variability compared to our simulated results as indicated by the usually substantially larger standard deviation in the observational data (Fig. 2). This can be explained by the lack of DGVMs in representing the spatial heterogeneity of local soil and climate conditions and land-management practices.

Even though DGVMs provide land-use-related absolute soil C changes, our comparison focused on relative changes. This is the preferred variable in the meta-analyses because spatial heterogeneity partly cancels in relative terms when two sites in close proximity are compared to each other, as done in paired-plots setups. Only relative changes allow for deriving robust carbon response functions (Poeplau et al., 2011). In the jsb

SUBSCRIPTNBdrvn

SUBSCRIPTNBharv simulation, the equilibrium changes indicate a decrease in soil C of about 11 kgC m^{-2} and 3 kgC m^{-2} for forest to crop and grass to crop, respectively, in the temperate region. The decrease for forest to crop in the tropics is about 9 kgC m^{-2} (Fig. 1b). The reverse LUCs result in soil C increase of about the same magnitude. Because DGVMs are unaffected by small-scale spatial heterogeneity, their estimates of absolute changes are expected to be more robust than those of meta-analyses and therefore better representative for global C responses. After successful evaluation against relative changes, DGVMs can therefore be used to assess large-scale soil C changes in the absolute terms that are relevant for carbon budget estimates.

4.2 Causes of model deviation from meta-analyses

4.2.1 Accounting for crop harvesting

The importance of accounting for crop management practices, such as crop harvesting, irrigation and tillage, in DGVMs has been highlighted by recent studies (Levis et al., 2014; Pugh et al., 2015). In particular, Pugh et al. (2015) showed that the inclusion of tillage, grazing and crop harvesting in the LPJ-GUESS model increases the historical land-use carbon emissions. The increased emissions result from the reduced carbon inputs to the soil by removal of harvested material off-field and increased turnover rates via tillage. Our results show that lack of explicitly accounting for crop harvesting does not only lead to underestimation of soil C changes following the conversion of crop to forest and vice versa, but it also contributes to the wrong direction of change for the crop to grass LUC (Figs. 1 and 2). Figure 3 shows that accounting for crop harvesting in JSBACH improves the temporal response of soil C to the conversion of crop to grass and crop to forest. The removal of 50% crop biomass to the harvest pool–based on root to shoot ratios–is uncertain as it differs across crop types (Table 1.2 in Fageria, 2012); hence this value may not may not be representative of all the sites in the meta-analyses. Despite the uncertainty associated with the harvested crop biomass, our results show that accounting for crop harvesting significantly reduces the soil C for croplands (Table 3).

We note that our model does not represent other crop management practices. For example, tillage in croplands leads to the exposure of mineral surfaces that are often inaccessible to decomposition causing more soil C loss (Post and Kwon, 2000). However, Pugh et al. (2015) showed that accounting for crop harvesting had larger effects on the historical carbon emissions compared to the inclusion of tillage. Moreover, fertilization can affect cropland soil C stocks by enhancing productivity and hence increasing soil C inputs, and compensating effects by enhancing decomposition by activating microbes (Russell et al., 2009). The carbon model used in this study simulates soil C based on the plant chemistry and climate. Recent studies have shown that the inclusion of microbial dynamics and priming processes in biogeochemical models can improve model agreement with observations (e.g., Wieder et al., 2013). As these processes are different across land-use types, the inclusion of such processes in future generation of DGVMs may lead to improved simulated soil C response to LUC.

4.2.2 Accounting for fire

DGVMs include process representation of vegetation fires to account for the annual emissions of carbon resulting from fires and to allow dynamical shifts in vegetation distribution. However, the choice of which vegetation type burns varies across different DGVMs. Earlier representations of fire in DGVMs accounted for burning only for natural vegetation types (e.g., Kloster et al., 2010; Reick et al., 2013), while recent studies included burning in pastures (e.g., Lasslop et al., 2014) and croplands (e.g., Li et al., 2013). Remote sensing data show that the burned area for different vegetation types varies across different regions. For example, Giglio et al. (2013) showed that while crops contribute to more than 50% of the burned area in Europe and Middle East, grasslands contribute to more than 50% of the burned area in Central Asia. Our model accounts for burning only in natural vegetation types. We perform sensitivity simulations where grasslands are treated the same as croplands by neglecting burning in grasslands in the standard model simulation (~~jsbach~~jsb_drvn, which does not account for crop harvesting). The sensitivity simulations show a direction of change that is in accordance with the ~~meta-data~~ observational data for crop to grass and grass to crop. (Figs. 1 and 2). In the simulations accounting for crop harvesting (~~jsbach~~jsb_drvn_harv), neglecting burning in grasslands would lead to even larger relative and absolute changes for the crop to grass and grass to crop LUCs. This shows that DGVMs assumptions on which vegetation types burn plays a major role on the soil C response to LUC. However, it remains unclear if the ~~meta-data~~ site studies included in the meta-analyses represent regularly burned regions or not. Establishing observational evidence for the sensitivity of soil C changes for a given land use towards frequency and intensity of fire events, similar to how ~~meta-data~~ meta-analyses show the sensitivity of responses to factors like precipitation, temperature or soil texture, would allow to evaluate the relevance of this process as currently represented in DGVMs.

4.2.3 Conversion of forests to managed grasslands

~~Meta-data~~ Results from the meta-analyses have shown that the conversion of forest to pasture in the tropics leads to negligible changes in the soil C and in some cases an increase ~~. For the temperate regions, the conversion of grassland to forest increased soil C when the forest floor was included, while without forest floor a decrease in soil C was observed~~ (Guo and Gifford, 2002; Murty et al., 2002; Don et al., 2011). We find that in the model the conversion of forest to pasture for the tropics leads to a decline in soil C comparable to that of converting forest to crop (Fig. 1). This is associated with larger NPP for forests compared to pastures, which leads to larger litter fluxes (Table 4 and 5). For most of the considered regions in the tropics, the larger simulated NPP for forests compared to pastures is consisted with other observations (Smith et al., 2012). Murty et al. (2002) associated the observed increase in soil C following conversion of forest to pasture with low initial content of soil C, application of fertiliser and avoided grazing. Table 3 shows low previous land use soil C for forest to pasture compared to forest to crop in the meta-analyses. However, the model does not simulate low previous land use soil C for the forest to pasture transition in the considered regions (Table 3).

For the temperate regions, the conversion of grassland to forest increased soil C when the surface litter was included, while without surface litter a decrease in soil C was observed (Poeplau et al., 2011). Our comparison does not include conversions between forest and grass in the temperate regions, the smaller change for grass to crop as compared to forest to crop suggests, however, also here a simulated loss of carbon for the forest to grassland LUC.

Schulze et al. (2010) in their review of the European carbon balance found that grasslands store more carbon compared to forests. They attribute this to the higher below-ground allocation for grasslands compared to forests, annual root turnover and possibly nitrogen fixation. Our model does not explicitly represent the potentially deep rooting of grasses, which likely contributes to the disagreement in sign of change for the tropical forest to pasture transition and the weaker simulated response for the temperate grass to crop transition. The latter may further be explained by our simulations not capturing the differences in productivity of grasslands compared to forests and cropland found ~~accross~~ across various eddy covariance sites in Europe (Schulze et al., 2010). Schulze et al. (2010) found generally larger NPP for grasslands and croplands, while the simulated results shows on average higher productivity for forests for the considered temperate regions (Table 4).

4.2.4 ~~Sensitivity of results to grid cell selections~~

~~By selecting larger regions for comparison with the meta-data based on where LUC has taken place historically, instead of selecting regions based on the climate-criterion, we find that our results are qualitatively the same as the ones represented above, based on the climate-criterion (Fig. 4). Small differences occur for forest to crop and crop to forest in the temperate regions, where the LUC regions have smaller changes compared to the regions captured by the climate-criterion. This suggests that the regions captured by the meta-data by and are generally representative of regions where LUC has taken place historically, although the latter may not be representative of whole-ecosystem averages (see).~~

~~Even though meta-data may have biases towards regions of similar soil and climate conditions , the meta-data still show a larger variability compared to our simulations results as indicated by the usually substantially larger standard deviation in the observational data (Fig. 2). This can be explained by the lack of DGVMs in representing the spatial heterogeneity of local soil and climate conditions and land-management practices. Site-level simulations are often used to evaluate DGVMs for~~ ~~CO_2 fluxes, such as net ecosystem exchange and terrestrial ecosystem respiration . While vegetation processes representing such variables are well represented in the models, soil processes that are important at the local scale, such as soil chemistry, are not represented in DGVMs. Although it may be impossible for a DGVM to capture the soil C response at an individual site, in particular if the site is not representative of a larger region, the model should be able to match average responses across observations covering a wide region. It is therefore possible to evaluate DGVMs at the scales they are meant for, e.g., providing estimates relevant for the global carbon budget .~~

4.3 Challenges in model-data comparison

4.3.1 Sampling at different times following land-use change

The local-scale measurements constituting the ~~meta-data~~ meta-analyses are taken at different times after LUC; hence the current land use is often not in equilibrium. Yet often sites at different stages of disequilibrium are included in average responses, which have been subsequently interpreted in modeling studies as indication for the observation-based evidence of effects of historical land-use change on equilibrium soil C stock changes (Pongratz et al., 2009; Reick et al., 2010; Stocker et al., 2011). Idealized simulations such as presented here can account for this transience in soil C response by sampling over the same ages as represented by the ~~meta-data~~meta-analyses. Due to the larger availability of sites that have recently undergone LUC, averaging over all available sites of different ages in the ~~meta-data~~ meta-analyses has a strong bias towards smaller soil C changes than would be expected in equilibrium. In our model, this bias becomes apparent in the smaller relative and absolute changes compared to the equilibrium changes (Figs. 1 and 2). The bias can be quantified in our simulations and amounts to about 20-40% of the equilibrium response that is captured by an average across the simulations accounting for crop harvest (supplementary material Table S4). Therefore, parametrization and evaluation of DGVMs using ~~meta-data~~ meta-analyses needs to account for the transient state of the mean soil C changes for the different LUCs represented in the ~~meta-data~~meta-analyses.

4.3.2 Different soil sampling depths

Soil carbon models used in DGVMs typically simulate soil processes up to a depth of 1 m and are meant to capture the complete soil C stock changes after LUC. By contrast, some of the observations, in particular in the tropics, covered only a shallow sampling depth (Table 1). Analysis of the depth-dependence of observed soil C changes revealed that most of the change occurs in the top ~~soil (0-30 cm depth)~~ 30 cm (Poeplau and Don, 2013), in line with the fact that in most ecosystems the majority of soil C is stored in the upper layers with around 1520 Pg C, which is more than 56% of the total soil C globally, in the upper 1 m (Jobbágy and Jackson, 2000). ~~To compare simulated soil C densities and changes with the meta-data, a~~ For comparison of the relative and absolute of soil C changes at consistent depth, scaling of the ~~simulated soil C to the top layer or the local-scale measurements to the entire 1 m~~ site studies to the model depth can be ~~performed. In our assessment of carbon densities we~~ applied (Yang et al., 2011; Deng et al., 2014). However, these scaling approaches used an equation calibrated across a wide range of ecosystems ~~to scale the simulated soil C in our standard model simulation to the top layer. Being independent of vegetation type, such an approach does not alter the results of the relative changes, which excludes additional uncertainty introduced by scaling on the comparison of~~ and are thus independent of the land use types. Hence, such as a scaling would only affect the comparison of the absolute changes, but not the relative changes.

~~However, previous~~ Previous studies have shown that the amount of soil C varies with depth differently in different ecosystems. For example, Jobbágy and Jackson (2000) found that 42% of the total soil C in grasslands is stored in the upper 20 cm while for forests 50% of the carbon is in the upper 20 cm. Guo and Gifford (2002) argued that while forests have high above-ground inputs in the top layers, tree roots are less important sources of organic matter because much of the tree root systems lives for many years. On the other hand, the annual root turnover in grasslands contributes to larger soil C storage in deeper depths. Therefore also the changes of soil C vary with depth differently for different LUCs. Poeplau and Don (2013), using several local scale measurements, found that 91% of the total soil C change occurs in the ~~top soil~~ upper 30 cm following afforestation, while 65% of the change occurs in the top soil following the conversion of crop to grass. In line with this, DGVMs may need to consider including vertically resolved soil profiles to represent the distribution of soil C with depth across different ecosystems, to represent that different types of LUCs act differently depending on the sampled depth and to be better comparable with ~~meta-data~~meta-analyses. Conversely, to capture the full impacts of LUC on soil C as relevant for carbon budgeting and to allow a direct comparison to DGVMs, local-scale measurements need to consider a deeper sampling of the soil profile.

4.4 ~~Relative versus absolute changes comparison~~

~~In our comparison, we have foremost focused on relative changes in soil C rather than absolute ones. This is the preferred variable in the meta-data because spatial heterogeneity partly cancels in relative terms when two sites in close proximity are compared to each other, as done in chronosequences. Only relative changes allow for deriving robust carbon response functions . However, because DGVMs are unaffected by small-scale spatial heterogeneity, their estimates of absolute changes are expected to be more robust than those of meta-data and therefore better representative for global carbon responses. After~~

~~successful evaluation against relative changes, DGVMs can therefore be used to assess large-scale soil C changes in the absolute terms that are relevant for carbon budget estimates. In the jsbachSUBSCRIPTNBdrvnSUBSCRIPTNBharv simulation, the equilibrium changes indicate a decrease in soil C of about 11 kgC m^{-2} and 3 kgC m^{-2} for forest to crop and grass to crop, respectively, in the temperate region. The decrease for forest to crop in the tropics is about 9 kgC m^{-2} (Fig. 1b). The reverse LUCs result in soil C increase of about the same magnitude.~~

5 Conclusions

Our comparison used a typical DGVM, JSBACH, which has been applied in a range of model intercomparison projects (e.g. Brovkin et al., 2013; Le Quéré et al., 2015). It revealed successful representation of some, but not of all LUCs. The comparison supports previous studies that found that inclusion of crop harvesting is a crucial component in DGVMs to accurately represent soil carbon losses with agricultural expansion and historical land-use emissions (Stocker et al., 2011; Pugh et al., 2015). Additionally, we find that natural disturbances by fire, which are not well documented in the ~~meta-data~~meta-analyses, may substantially influence the soil carbon response to LUC simulated by models.

Challenges for this comparison remain. First, ~~meta-data~~ meta-analyses cover many observations where the current land use may not be in equilibrium; hence the mean relative changes in the ~~meta-data~~ meta-analyses represent a transient response. Idealized LUC simulations can account for this by sampling over the ages represented by the ~~meta-data~~meta-analyses. Second, the ~~meta-data~~ meta-analyses include local-scale observations that are done at different sampling depths. Ultimately this challenge can be overcome only by deeper sampling in observational data or by DGVMs considering in the future including a vertically resolved soil profile.

Despite such challenges, our study shows that the use of the ~~meta-data~~ meta-analyses on soil carbon changes following LUC offers the opportunity for evaluation and improvement of DGVMs. We developed a systematic approach that is applicable to any DGVM for comparing simulated soil carbon changes due to different LUCs using the ~~meta-data~~meta-analyses. Extending this comparison to other DGVMs or to model intercomparison projects would not only provide an observational reference for validation, but also help investigate across a larger range of processes the key influences on models' sensitivity to LUC.

Acknowledgements. We thank Jari Liski for his helpful discussions during the course of this study and Daniel Goll for his contribution in the implementation of the YASSO model into JSBACH. We thank Martin Jung and Ulrich Weber for providing us with the GPP and LAI data used in some of the simulations. We thank Goran Georgievski for his comments, which helped to improve the manuscript. This work was supported by the German Research Foundation's Emmy Noether Program (PO 1751/1-1).

References

Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R., and Zhu, Z.: Evaluating the Land and Ocean Components of the Global Carbon Cycle in the CMIP5 Earth System Models, Journal of Climate, 26, 6801–6843, doi:10.1175/JCLI-D-12-00417.1, 2013.

Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., and Smith, B.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, Global Change Biology, 13, 679–706, doi:10.1111/j.1365-2486.2006.01305.x, 2007.

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, doi:10.1175/JCLI-D-12-00623.1, 2013.

Ciais, P., Sabine, C., Bala, G., Brovkin, V., Canadell, A., Chhabra, A., DeFries, R., Galloway, J., Heinemann, M., Jones, C., Le Quéré, C., Myneni, R. B., Piao, S., and Thornton, P.: Carbon and Other Biogeochemical Cycles. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group 1 to the Fifth Assessment Report on Climate Change, chap. 6, pp. 486–494, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Collatz, G., Ribas-Carbo, M., and Berry, J.: Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C4 Plants, Australian Journal of Plant Physiology, 19, 519, doi:10.1071/PP9920519, 1992.

Deng, L., Liu, G.-B., and Shangguan, Z.-P.: Land-use conversion and changing soil carbon stocks in China's 'Grain-for-Green' Program: a synthesis., Global change biology, 20, 3544–56, doi:10.1111/gcb.12508, 2014.

Don, A., Schumacher, J., and Freibauer, A.: Impact of tropical land-use change on soil organic carbon stocks - a meta-analysis, Global Change Biology, 17, 1658–1670, doi:10.1111/j.1365-2486.2010.02336.x, 2011.

Ellert, B. and Bettany, J.: Calculation of organic matter and nutrients stored in soils under contrasting management regimes, Canadian Journal of Soil Science, 75, 529–538, doi:10.4141/cjss95-075, 1995.

Fageria, N. K.: The Role of Plant Roots in Crop Production, chap. 1, pp. 1–44, CRC Press, Florida, USA, doi:10.1201/b12365-8, 2012.

Farquhar, G. D., Caemmerer, S. V., and Berry, J. A.: A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species, Planta, 90, 78–90, doi:10.1007/BF00386231, 1980.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sensing of Environment, 114, 168–182, doi:10.1016/j.rse.2009.08.016, 2010.

Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), Journal of Geophysical Research: Biogeosciences, 118, 317–328, doi:10.1002/jgrg.20042, 2013.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, Journal of Advances in Modeling Earth Systems, 5, 572–597, doi:10.1002/jame.20038, 2013.

Goll, D. S., Brovkin, V., Liski, J., Raddatz, T., Thum, T., and Todd-Brown, K. E. O.: Strong dependence of CO2 emissions from anthropogenic land cover change on initial land cover and soil carbon parametrization, Global Biogeochemical Cycles, 29, doi:10.1002/2014GB004988, 2015.

Gray, J. M., Frolking, S., Kort, E. a., Ray, D. K., Kucharik, C. J., Ramankutty, N., and Friedl, M. a.: Direct human influence on atmospheric CO2 seasonality from increased cropland productivity., Nature, 515, 398–401, doi:10.1038/nature13957, 2014.

Guo, L. B. and Gifford, R. M.: Soil carbon stocks and land use change: a meta analysis, Global change biology, 8, 345–360, doi:10.1046/j.1354-1013.2002.00486.x, 2002.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, International Journal of Climatology, 34, 623–642, doi:10.1002/joc.3711, 2014.

Houghton, R. A., Hobbie, J. E., Melillo, J. M., Moore, B., Peterson, B. J., Shaver, G. R., and Woodwell, G. M.: Changes in the Carbon Content of Terrestrial Biota and Soils between 1860 and 1980: A Net Release of CO" 2 to the Atmosphere, Ecological Monographs, 53, 235–262, doi:10.2307/1942531, 1983.

Houghton, R. A., House, J. I., Pongratz, J., van der Werf, G. R., DeFries, R. S., Hansen, M. C., Le Quéré, C., and Ramankutty, N.: Carbon emissions from land use and land-cover change, Biogeosciences, 9, 5125–5142, doi:10.5194/bg-9-5125-2012, 2012.

Jobbágy, E. G. and Jackson, R. B.: The vertical distribution of soil organic carbon and its relation to climate and vegetation, Ecological Applications, 10, 423–436, doi:10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2, 2000.

Karhu, K., Wall, A., Vanhala, P., Liski, J., Esala, M., and Regina, K.: Effects of afforestation and deforestation on boreal soil carbon stocks—Comparison of measured C stocks with Yasso07 model results, Geoderma, 164, 33–45, doi:10.1016/j.geoderma.2011.05.008, 2011.

Kloster, S., Mahowald, N. M., Randerson, J. T., Thornton, P. E., Hoffman, F. M., Levis, S., Lawrence, P. J., Feddema, J. J., Oleson, K. W., and Lawrence, D. M.: Fire dynamics during the 20th century simulated by the Community Land Model, Biogeosciences, 7, 1877–1902, doi:10.5194/bg-7-1877-2010, 2010.

Laganiére, J., Angers, A., and Paré, D.: Carbon accumulation in agricultural soils after afforestation: a meta-analysis, Global Change Biology, 16, 439–453, doi:10.1111/j.1365-2486.2009.01930.x, 2010.

Lasslop, G., Thonicke, K., and Kloster, S.: SPITFIRE within the MPI Earth system model: Model development and evaluation, Journal of Advances in Modeling Earth Systems, 6, 740–755, doi:10.1002/2013MS000284, 2014.

Le Quéré, C., Moriarty, R., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Friedlingstein, P., Peters, G. P., Andres, R. J., Boden, T. A., Houghton, R. A., House, J. I., Keeling, R. F., Tans, P., Arneth, A., Bakker, D. C. E., Barbero, L., Bopp, L., Chang, J., Chevallier, F., Chini, L. P., Ciais, P., Fader, M., Feely, R. A., Gkritzalis, T., Harris, I., Hauck, J., Ilyina, T., Jain, A. K., Kato, E., Kitidis, V., Klein Goldewijk, K., Koven, C., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lima, I. D., Metzl, N., Millero, F., Munro, D. R., Murata, A., Nabel, J. E. M. S., Nakaoka, S., Nojiri, Y., O'Brien, K., Olsen, A., Ono, T., Pérez, F. F., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Rödenbeck, C., Saito, S., Schuster, U., Schwinger, J., Séférian, R., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Vandemark, D., Viovy, N., Wiltshire, A., Zaehle, S., and Zeng, N.: Global Carbon Budget 2015, Earth System Science Data, 7, 349–396, doi:10.5194/essd-7-349-2015, 2015.

Levis, S., Hartman, M. D., and Bonan, G. B.: The Community Land Model underestimates land-use CO2 emissions by neglecting soil disturbance from cultivation, Geoscientific Model Development, 7, 613–620, doi:10.5194/gmd-7-613-2014, 2014.

Li, F., Levis, S., and Ward, D. S.: Quantifying the role of fire in the Earth system – Part 1: Improved global fire modeling in the Community Earth System Model (CESM1), Biogeosciences, 10, 2293–2314, doi:10.5194/bg-10-2293-2013, 2013.

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, doi:10.5194/esd-4-385-2013, 2013.

Lu, N., Liski, J., Chang, R. Y., Akujärvi, A., Wu, X., Jin, T. T., Wang, Y. F., and Fu, B. J.: Soil organic carbon dynamics of black locust plantations in the middle Loess Plateau area of China, Biogeosciences, 10, 7053–7063, doi:10.5194/bg-10-7053-2013, 2013.

Lugo, E. A., Sanchez, J. M., and Brown, S.: Land use and organic carbon content of some subtropical soils, Plant and Soil, 96, 185–196, doi:10.1007/BF02374763, 1986.

Marín-Spiotta, E. and Sharma, S.: Carbon storage in successional and plantation forest soils: a tropical analysis, Global Ecology and Biogeography, 22, 105–117, doi:10.1111/j.1466-8238.2012.00788.x, 2013.

Murty, D., Kirschbaum, M. U. F., Mcmurtrie, E. R., and Mcgilvray, H.: Does conversion of forest to agricultural land change soil carbon and nitrogen? A review of the literature, Global Change Biology, pp. 105–123, doi:10.1046/j.1354-1013.2001.00459.x, 2002.

Paul, K. and Polglase, P.: Change in soil carbon following afforestation, Forest Ecology and Management, 168, 241–257, doi:10.1016/S0378-1127(01)00740-X, 2002.

Poeplau, C. and Don, A.: Sensitivity of soil organic carbon stocks and fractions to different land-use changes across Europe, Geoderma, 192, 189–201, doi:10.1016/j.geoderma.2012.08.003, 2013.

Poeplau, C. and Don, A.: Carbon sequestration in agricultural soils via cultivation of cover crops – A meta-analysis, Agriculture, Ecosystems & Environment, 200, 33–41, doi:10.1016/j.agee.2014.10.024, 2015.

Poeplau, C., Don, A., Vesterdal, L., Leifeld, J., Van Wesemael, B., Schumacher, J., and Gensior, A.: Temporal dynamics of soil organic carbon after land-use change in the temperate zone - carbon response functions as a model approach, Global Change Biology, 17, 2415–2427, doi:10.1111/j.1365-2486.2011.02408.x, 2011.

Pongratz, J., Reick, C. H., Raddatz, T., and Claussen, M.: Effects of anthropogenic land cover change on the carbon cycle of the last millennium, Global Biogeochemical Cycles, 23, doi:10.1029/2009GB003488, 2009.

Pongratz, J., Reick, C. H., Raddatz, T., Caldeira, K., and Claussen, M.: Past land use decisions have increased mitigation potential of reforestation, Geophysical Research Letters, 38, doi:10.1029/2011GL047848, 2011.

Pongratz, J., Reick, C. H., Houghton, R. A., and House, J. I.: Terminology as a key uncertainty in net land use and land cover change carbon flux estimates, Earth System Dynamics, 5, 177–195, doi:10.5194/esd-5-177-2014, 2014.

Post, W. M. and Kwon, K. C.: Soil carbon sequestration and land-use change: processes and potential, Global Change Biology, 6, 317–327, doi:10.1046/j.1365-2486.2000.00308.x, 2000.

Powers, J. S., Corre, M. D., Twine, T. E., and Veldkamp, E.: Geographic bias of field observations of soil carbon stocks with tropical land-use changes precludes spatial extrapolation, Proceedings of the National Academy of Sciences, 108, 6318–6322, doi:10.1073/pnas.1016774108, 2011.

Pugh, T. A. M., Arneth, A., Olin, S., Ahlström, A., Bayer, A. D., Klein Goldewijk, K., Lindeskog, M., and Schurgers, G.: Simulated carbon emissions from land-use change are substantially enhanced by accounting for agricultural management, Environmental Research Letters, 10, 124 008, doi:10.1088/1748-9326/10/12/124008, 2015.

Raddatz, T. J., Reick, C. H., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K.-G., Wetzel, P., and Jungclaus, J.: Will the tropical land biosphere dominate the climate–carbon cycle feedback during the twenty-first century?, Climate Dynamics, 29, 565–574, doi:10.1007/s00382-007-0247-8, 2007.

Reick, C. H., Raddatz, T., Pongratz, J., and Claussen, M.: Contribution of anthropogenic land cover change emissions to pre-industrial atmospheric CO2, Tellus B, 62, 329–336, doi:10.1111/j.1600-0889.2010.00479.x, 2010.

Reick, C. H., Raddatz, T., Brovkin, V., and Gayler, V.: Representation of natural and anthropogenic land cover change in MPI-ESM, Journal of Advances in Modeling Earth Systems, 5, 459–482, doi:10.1002/jame.20022, 2013.

Russell, A. E., Cambardella, C. A., Laird, D. A., Jaynes, B. D., and Meek, W. D.: Nitrogen fertilizer effects on soil carbon balances in Midwestern US agricultural systems, Ecological Applications, 19, 1102–1113, doi:10.1890/07-1919.1, 2009.

Schulze, E. D., Ciais, P., Luyssaert, S., Schrumpf, M., Janssens, I. A., Thiruchittampalam, B., Theloke, J., Saurat, M., Bringezu, S., Lelieveld, J., Lohila, A., Rebmann, C., Jung, M., Bastviken, D., Abril, G., Grassi, G., Leip, A., Freibauer, A., Kutsch, W., Don, A., Nieschulze, J., Börner, A., Gash, J. H., and Dolman, A. J.: The European carbon balance. Part 4: integration of carbon and other trace-gas fluxes, Global Change Biology, 16, 1451–1469, doi:10.1111/j.1365-2486.2010.02215.x, 2010.

Shevliakova, E., Pacala, S. W., Malyshev, S., Hurtt, G. C., Milly, P. C. D., Caspersen, J. P., Sentman, L. T., Fisk, J. P., Wirth, C., and Crevoisier, C.: Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink, Global Biogeochemical Cycles, 23, doi:10.1029/2007GB003176, 2009.

Smith, W. K., Zhao, M., and Running, S. W.: Global Bioenergy Capacity as Constrained by Observed Biospheric Productivity Rates, BioScience, 62, 911–922, doi:10.1525/bio.2012.62.10.11, 2012.

Stocker, B. D., Strassmann, K., and Joos, F.: Sensitivity of Holocene atmospheric CO2 and the modern carbon budget to early human land use: analyses with a process-based model, Biogeosciences, 8, 69–88, doi:10.5194/bg-8-69-2011, 2011.

Thum, T., Räisänen, P., Sevanto, S., Tuomi, M., Reick, C., Vesala, T., Raddatz, T., Aalto, T., Järvinen, H., Altimir, N., Pilegaard, K., Nagy, Z., Rambal, S., and Liski, J.: Soil carbon model alternatives for ECHAM5/JSBACH climate model: Evaluation and impacts on global carbon cycle estimates, Journal of Geophysical Research, 116, G02 028, doi:10.1029/2010JG001612, 2011.

Tian, H., Lu, C., Yang, J., Banger, K., Huntzinger, N. D., Schwalm, C. R., Michalak, A. M., Cook, R., Cias, P., Hayes, D., Huang, M., Ito, A., Jain, A. K., Lei, H., Mao, J., Pan, S., Post, W., Peng, S., Poulter, B., Ren, W., Ricciuto, D., Schaefer, K., Shi, X., Tao, B., Wang, W., Wei, Y., Yang, Q., Zhang, B., and Zeng, N.: Global Patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: current status and future directions, Global Biogeochemical Cycles, 29, 775–792, doi:10.1002/2014GB005021, 2015.

Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A. G., and Allison, S. D.: Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, Biogeosciences, 10, 1717–1736, doi:10.5194/bg-10-1717-2013, 2013.

Tramontana, G., Jung, M., Camps-Valls, G., Ichii, K., Raduly, B., Reichstein, M., Schwalm, C. R., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences Discussions, pp. 1–33, doi:10.5194/bg-2015-661, 2016.

Tuomi, M., Vanhala, P., Karhu, K., Fritze, H., and Liski, J.: Heterotrophic soil respiration—Comparison of different models describing its temperature dependence, Ecological Modelling, 211, 182–190, doi:10.1016/j.ecolmodel.2007.09.003, 2008.

Tuomi, M., Thum, T., Järvinen, H., Fronzek, S., Berg, B., Harmon, M., Trofymow, J., Sevanto, S., and Liski, J.: Leaf litter decomposition—Estimates of global variability based on Yasso07 model, Ecological Modelling, 220, 3362–3371, doi:10.1016/j.ecolmodel.2009.05.016, 2009.

Tuomi, M., Laiho, R., Repo, A., and Liski, J.: Wood decomposition model for boreal forests, Ecological Modelling, 222, 709–718, doi:10.1016/j.ecolmodel.2010.10.025, 2011.

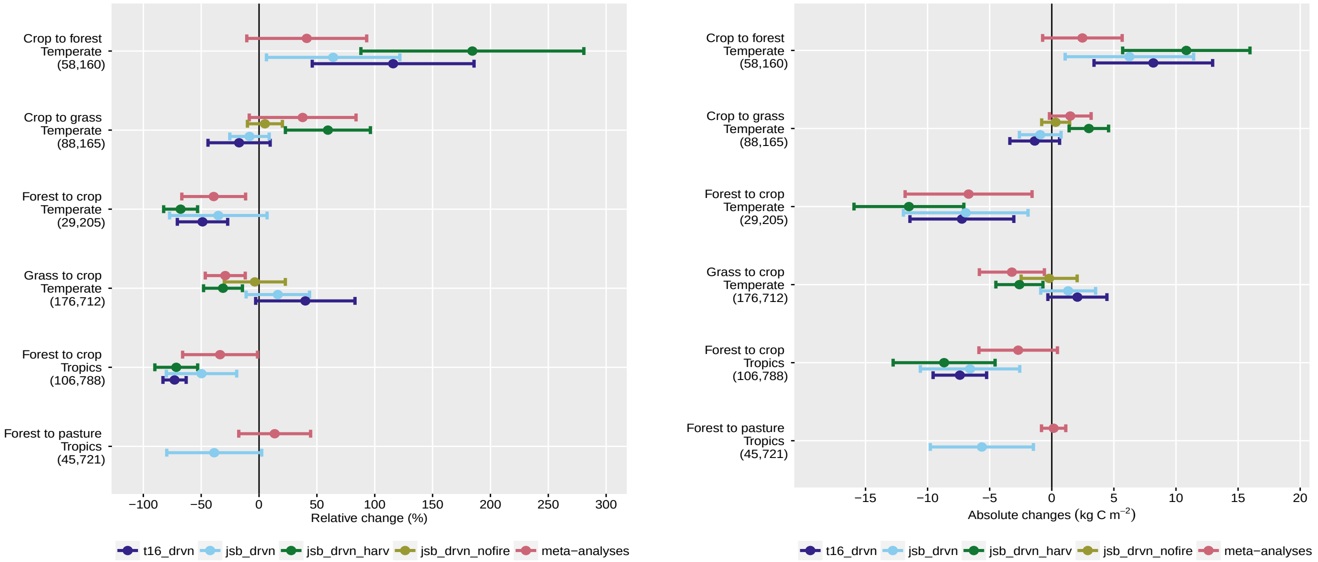
Watson, T. R., Noble, I., Bolin, B., Ravindranath, N. H., Verardo, J. D., and Dokken, J.: Land Use, Land-Use Change, and Forestry, Cambridge university Press, 2000.

Wei, X., Shao, M., Gale, W., and Li, L.: Global pattern of soil carbon losses due to the conversion of forests to agricultural land., Scientific reports, 4, 4062, doi:10.1038/srep04062, 2014.

Wieder, W. R., Bonan, G. B., and Allison, S. D.: Global soil carbon projections are improved by modelling microbial processes, Nature Climate Change, 3, 909–912, doi:10.1038/nclimate1951, 2013.

Wilkenskjeld, S., Kloster, S., Pongratz, J., Raddatz, T., and Reick, C. H.: Comparing the influence of net and gross anthropogenic land-use and land-cover changes on the carbon cycle in the MPI-ESM, Biogeosciences, 11, 4817–4828, doi:10.5194/bg-11-4817-2014, 2014.

Yang, Y., Luo, Y., and Finzi, A. C.: Carbon and nitrogen dynamics during forest stand development: a global synthesis, New Phytologist, 190, 977–989, doi:10.1111/j.1469-8137.2011.03645.x, 2011.



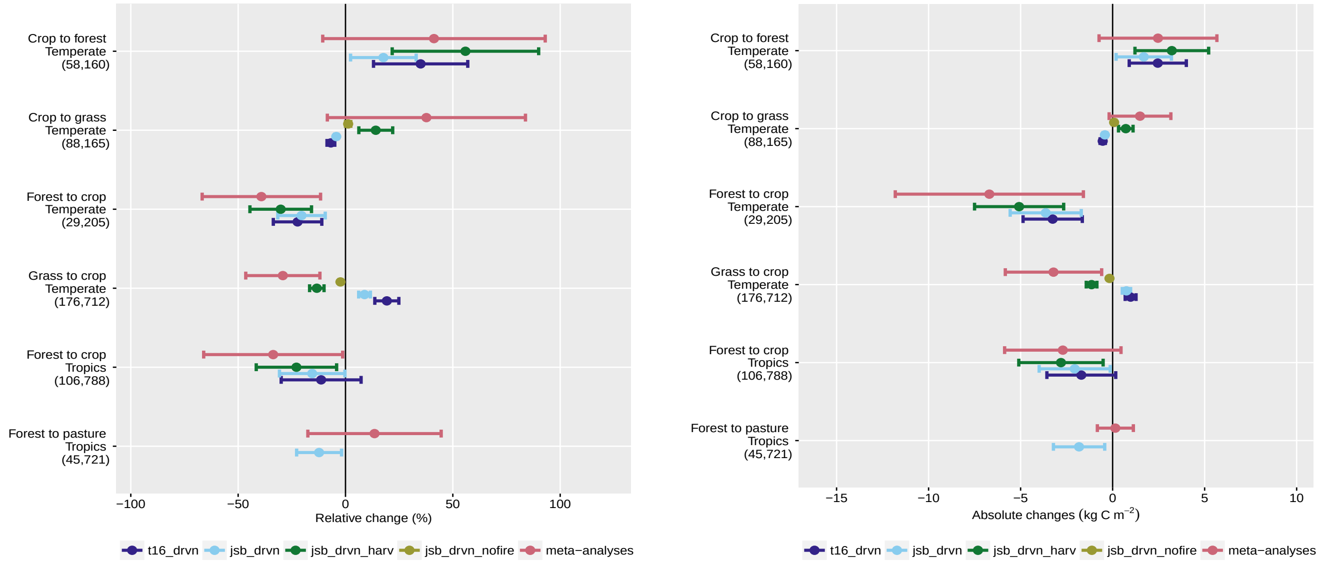
(a) Mean equilibrium relative changes

(b) Mean equilibrium absolute changes

Figure 1. Mean simulated equilibrium relative (a) and absolute changes in soil carbon (b) compared to results from the ~~mean changes for the meta-data~~meta-analyses. The first number in the parenthesis represents the number of studies in the ~~meta-data~~ meta-analyses and the second is the number of grid cells from the global simulation that ~~fulfil~~ fulfill the climate-criterion in the ~~meta-data~~ meta-analyses (regions in supplementary material Fig. S1). The dots represent the mean changes and the bars represent the standard deviation.

Table 1. Mean annual temperature (MAT) range, mean annual precipitation (MAP) range, mean sampling depths (± std) and the mean current land-use age for the local-scale observations in the ~~meta-data~~meta-analyses. We note that the different equilibrium results presented below, e.g., for crop in the crop to forest LUC and the forest to crop LUC, are due to the different climate-criterion (precipitation and temperature) for the different LUCs.

Land-use change	MAT (°C)	MAP (mm)	Sampling depth (cm)	Age (years)
Crop to forest (temperate)	5.9–10.7	540–1020	39.53±24.8	40.28
Crop to grass (temperate)	6.7–11.2	440–1030	23.44±10.5	21.7
Forest to crop (temperate)	3.4–16.4	690–1320	28.48±13.5	50.21
Grass to crop (temperate)	1–12.7	150–960	27.11±11.1	39.69
Forest to crop (tropics)	15–27.5	570–3400	17.5±12.81	22.5
Forest to pasture (tropics)	18–28	570–4000	15.79±11.55	20.67



(a) Mean sampled ages relative changes

(b) Mean sampled ages absolute changes

Figure 2. Mean simulated relative (a) and absolute changes in soil carbon (b) over the sampled ages represented by the ~~meta-data~~ meta-analyses compared to results from the ~~mean changes for the meta-data~~meta-analyses. The first number in the parenthesis represents the number of studies in the ~~meta-data~~ meta-analyses and the second is the number of grid cells fulfilling the climate-criterion in the ~~meta-data~~ meta-analyses (regions in supplementary material Fig. S1). The dots represent the mean changes and the bars represent the standard deviation.

Table 2. Summary of the simulations done in this study.

Simulation name	NPP& LAI	Land-use change	Disturbances	Crop harvest
~~jsbach~~jsb_drvn	Simulated by JSBACH	crop to forest, forest to crop, crop to grass, grass to crop, forest to pasture	on	none
~~obs~~t16_drvn	~~Observations~~ Prescribed from observations	crop to forest, forest to crop, crop to grass, grass to crop	on	none
~~jsbach~~jsb_drvn_harv	Simulated by JSBACH	crop to forest, forest to crop, crop to grass, grass to crop	on	included
~~jsbach~~jsb_drvn_nofire	Simulated by JSBACH	crop to grass, grass to crop	off	none

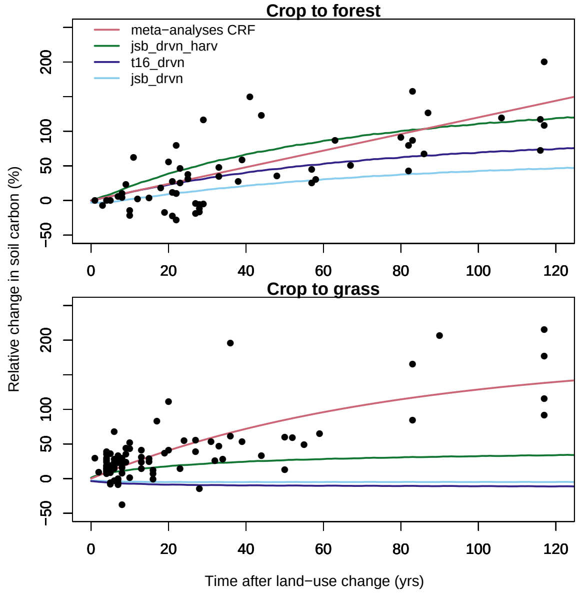
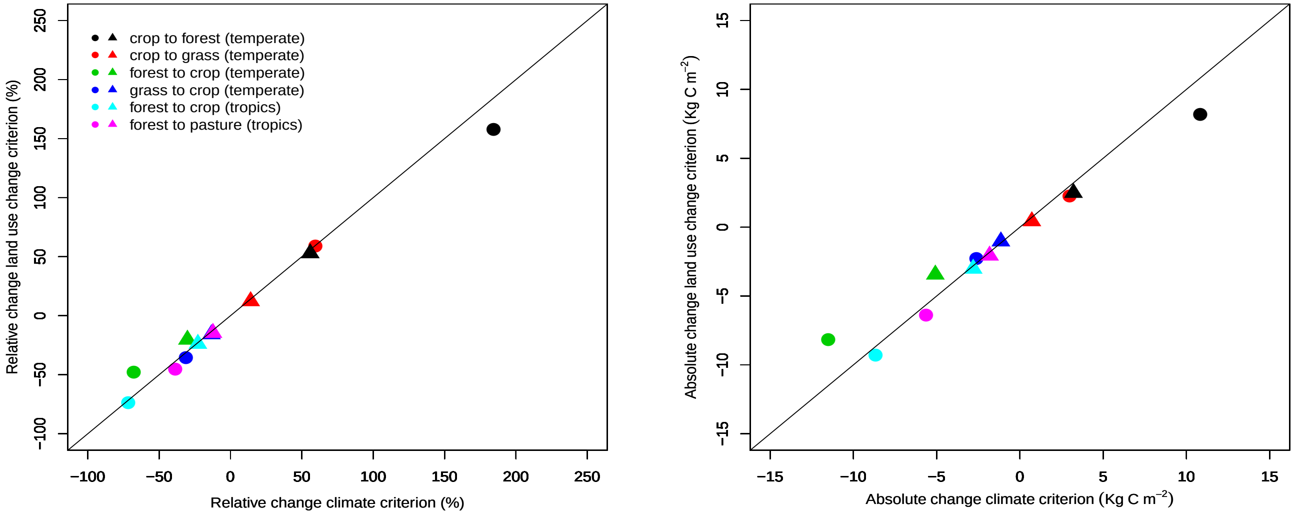


Figure 3. Mean simulated transient relative changes in soil carbon compared to the individual observations in the meta-analyses (black dots) and generalized carbon response functions (CRF) as in Poeplau et al. (2011) for the crop to grass and crop to forest LUCs.



(a) Mean relative changes

(b) Mean absolute changes

Figure 4. Mean relative (a) and absolute (b) changes for the different land-use transitions with regions based on the climate (temperature and precipitation) criterion and based on where land-use change has taken place historically for the ~~jsbach~~jsb_drvn_harv simulation. The triangles represent the mean changes over the sampled ages while the circles represent the mean equilibrium changes.

Table 3. Mean simulated equilibrium soil C densities at the model depth (100 cm) and the mean soil C in the meta-analyses in kgC m^{-2} for previous and current land use in the different LUCs and simulations (±std). The meta-analyses soil C densities represent the mean over sites with different measuring depths.

Land-use change	meta-analyses		t16_drvn		jsb_drvn		jsb_drvn_harv	
	Previous	Current	Previous	Current	Previous	Current	Previous	Current
Crop to forest (temperate)	6.8±3.1	9.3±5.1	7.2±1.7	15.4±5.5	10.1±2.9	16.9±6.4	6.0±2.9	16.9±6.4
Crop to grass (temperate)	4.6±2.1	6.1±2.6	7.4±1.7	6.0±2.1	9.5±2.6	8.6±2.8	5.6±2.6	8.6±2.8
Forest to crop (temperate)	14.7±5.3	8.0±2.7	13.4±5.1	6.1±1.7	16.5±5.4	8.4±2.7	16.5±5.4	5.0±1.8
Grass to crop (temperate)	11.5±6.7	8.3±5.6	6.2±2.1	8.3±2.9	8.4±3.9	9.8±4.7	8.4±3.9	5.7±2.8
Forest to crop (tropics)	6.4±3.9	3.7±2.3	10.1±2.4	2.7±1.1	11.4±4.3	4.8±1.7	11.4±4.3	2.7±1.1
Forest to pasture (tropics)	3.7±2.8	3.9±2.6	-	-	11.4±4.1	5.8±1.8	11.4±4.1	5.8±1.8

Table 4. Mean annual NPP for previous and current land use in kgC m^{-2} for the different LUCs and simulations (± std).

Land-use change	t16_drvn		jsb_drvn	
	Previous~~LU~~	Current~~LU~~	Previous~~LU~~	Current~~LU~~
Crop to forest (temperate)	0.42±0.10	0.73±0.24	0.58±0.15	0.90±0.34
Crop to grass (temperate)	0.43±0.09	0.41±0.14	0.57±0.15	0.63±0.17
Forest to crop (temperate)	0.77±0.26	0.44±0.12	1.04±0.34	0.58±0.14
Grass to crop (temperate)	0.32±0.11	0.34±0.12	0.48±0.24	0.44±0.23
Forest to crop (tropics)	1.21±0.28	0.35±0.10	1.42±0.60	0.69±0.22
Forest to pasture (tropics)	-	-	1.46±0.59	0.87±0.20

Table 5. Mean annual equilibrium litter fluxes in kgC m^{-2} for previous and current land use in the different LUCs and simulations (± std).

Land-use change	t16_drvn		jsb_drvn		jsb_drvn_harv	
	Previous~~LU~~	Current~~LU~~	Previous~~LU~~	Current~~LU~~	Previous~~LU~~	Current~~LU~~
Crop to forest (temperate)	0.41±0.10	0.66±0.21	0.57±0.14	0.79±0.28	0.35±0.14	0.79±0.28
Crop to grass (temperate)	0.41±0.09	0.39±0.13	0.55±0.14	0.58±0.15	0.34±0.14	0.58±0.15
Forest to crop (temperate)	0.74±0.25	0.44±0.11	0.95±0.32	0.58±0.13	0.95±0.32	0.34±0.09
Grass to crop (temperate)	0.30±0.10	0.33±0.10	0.44±0.21	0.43±0.23	0.44±0.21	0.26±0.14
Forest to crop (tropics)	1.21±0.28	0.35±0.10	1.28±0.54	0.63±0.19	1.28±0.54	0.37±0.12
Forest to pasture (tropics)	-	-	1.31±0.53	0.78±0.16	1.31±0.53	0.78±0.16

Table 6. Mean soil carbon turnover time (years) for the previous and current land use for the ~~jsbach~~jsb_drvn simulation with and without disturbances.

Land-use change	Previous ~~LU~~	Current ~~LU~~
Crop to grass, with disturbances	17.1±4.5	15±2.6
Crop to grass, no disturbances	17.1±4.5	17.2±4.3
Grass to crop, with disturbances	21.9±8.3	28.7±14.9
Grass to crop, no disturbances	28.6±14.5	28.5±14.5