MS Title: Effect of land use legacy on the future carbon sink 1 for the conterminous U.S.

Author: Benjamin S.Felzer

General comment

The manuscript entitled "Effect of land use legacy on the future carbon sink for the conterminous U.S." by Benjamin S.Felzer assesses the effects of land use and land cover change (LULCC) from 1750 to 2014 using the Terrestrial Ecosystems Model (TEM) for the conterminous U.S. Comparing model outputs from four experiments, the author finds that LULCC legacy has a considerable impact on carbon pools and fluxes. While the topic is of general interest, I don't follow the experimental design of the study. My primary concern is that the author combines meterological forcing data from one climate model (MPI-ESM-P 1750-1900) with quasi-observed data (CRU, 1901-2014) and climate projections from another model (CCSM4, 2015-2099) in one continuous simulation. Each data set comes with their own sets of biases. This lack of consistency will manifest as a forcing, causing changes in carbon fluxes. So the final signal is a combination of changes in environmental conditions, such as climate change, as well as the transition between inconsistent data sets. Also, one of the experiments (CONDENSED) has been re-equilibrated to the environmental conditions of 2015-2045. It remains unclear to me why the author chooses to equilibrate the model with projected climate conditions when trying to isolate the role of LULCC. I recommend that the author either adjusts the methods, or provides additional explanations that justify his approach. I recommend that the manuscript may be considered for publication in Biogeosciences after major revision.

Thank you for these comments.

The issue of combining multiple climate datasets is an interesting one. I believe it is important to use actual climate data (i.e. CRU4.04) rather than modeled climate data for the historical period when it is available. Otherwise, modeled climates will not capture the correct interannual variability – i.e. warm/cold years, wet/dry years when they actually occur. The modeled "data" may have the correct trend, but will not be correct on a year-to-year basis. It is absolutely critical that the future climate model data be downscaled and bias-corrected, which is why I chose to use the MACA datasets. While MACA was bias corrected using GRIDMET, that dataset starts in 1979 so does not provide a long historical period of data. However the resulting time series are continuous so there are no large biases. The decision to use modeled data prior to 1901, from 1750 to coincide with the period of LULCC, was not taken lightly, as I originally used a more traditional approach of using the first 30 years of the historical climate repeatedly for the period 1750-1901. However, when I published the precursor to this paper (Felzer et al., 2018) a reviewer took me to task for that approach (earlier review comment "It is also difficult to understand why the model was run from 1700 since climate data were not available until 1901. I would think climate before 1901 has affected the carbon cycle in the US and that effect cannot be reproduced by using a 1901-1930 average climate that includes big fire years around 1910 and the beginning of the drought of the 1930s. The little ice age is not represented well by the 1901-1930 average climate"). **For that reason, I responded by finding a millennial simulation, and chose**

the model with the highest resolution, the MPI-ESM-P (I also had to stitch together the millennial and historical periods to go from 1750-1900). Importantly, I did my own downscaling and bias correction to ensure continuity of climate, which is evident in Figure S3. To thoroughly explain these issues to the reader, I added the following paragraph to the Experimental Design section (including 3 new references) – lines 230-241:

The decision to base climate prior to 1900, prior to the gridded historical data, was made to capture more realistic climate variations during the period from 1750 to 1900, such as the Little Ice Age (LIA), which lasted through the 19th century (Bradley and Jones, 1993; Mann, 2002). The temperature record from the MPI-ESM-P model does show signs of temperature climbing out of a cold peak after 1818 but remaining cool throughout the rest of the century (Figure S3), which is consistent with Northern Hemisphere proxy records (Mann et al., 2008). Since this study is for the conterminous U.S., it does not show as strong an LIA signal as would be expected from records in the North Atlantic. The decision to then use historical CRU4.04 climate rather than modeled climate from 1901-2014 is to more accurately capture the true interannual variability, which would be entirely lost by using output from a climate model. All three datasets have been downscaled and bias corrected to produce a seamless record of climate from 1750-2099.

The second point about equilibration period I agree warranted a model rerun of the CONDENSED experiment. The reviewer accurately points out that I should be using the prior 30 years to 2015, rather than the post 30 years, as the basis for equilibration. Therefore, I ran the dynamic equilibration from 1986-2015, and used that as the basis for the initial conditions of the CONDENSED experiment. This change is noted in the Experimental Design section (lines 249-250). The new results are now used in all the figures involving the CONDENSED run, and numerical values throughout the text changed, where necessary. This change did not significantly alter any of the results. Both old and new figures are present in the revised document, so the reviewer can confirm that the differences exist but are minor.

Detailed Comments

L19 It is common practice to account for LULCC that started during the pre-industrial period. For instance, the TRENDY model ensemble that informs the annual publication of the global carbon budget accounts for LULCC starting in the year 1701.

(see https://blogs.exeter.ac.uk/trendy/protocol/)

Thanks for the comment. I would say the industrial period begins with the start of the Industrial Revolution in the 1860s. However, in my earlier study (Felzer and Jiang 2018) I did start the runs in 1700. I actually made a conscious decision to change the starting period of this study to 1750 to be consistent with the IPCC AR6 report, which generally uses 1750 as their baseline. I first realized this when reviewing a paper in which the authors started in 1750 and referred to it as some standard IPCC baseline, which is why I researched the issue further and decided to start in 1750 rather than 1701.

L41 To my understanding you don't compare model output against observation-based reference data. I would therefore not write that "carbon stocks are overestimated". Instead, I would either describe how carbon stocks differ among experiments or expand the analysis by comparing results against observations.

Good point – fixed to read "The carbon stocks are larger than using all the cohorts if condensed cohorts …". (line 41). I tried to correct this in other places as well.

L55 Replace "address" with "addresses".

Done

L80 Would a carbon sink related to regrowth not be larger if disturbance rates reduce, rather than "continue"?

I went back to the Pugh et al. (2019) paper, and their meaning is if disturbance rates continue at 1981-2010 levels, so I added "continue at historical levels" (line 81).

L80 Spell out the FIA acronym.

Done

L149 This section describes the different experiments (historical, restart, condensed, and temrestart) with respect to their initial values and whether they are based on the full or condensed version of the cohort. Please add how you treat LULCC, atmospheric CO_2 concentrations, nitrogen deposition and nitrogen fertilization when describing each experiment, and include this information in the table.

I replaced the table to include these new items, as well as ozone. Note that more detail is in the references or in the references of the associated text.

L156 Please motivate why you condense the full cohorts to 1 cohort/PFT.

Added the following sentence: The motivation for these two condensed-PFT runs is to reduce computational time by eliminating the need to run potentially thousands of land-use legacy cohorts for each grid when starting from present-day conditions (lines 175-178).

L166 You write that "the difference between the RESTART and CONDENSED runs shows the effect of including land legacy on future carbon dynamics". The difference between both runs is that the CONDENSED run re-equilibrates using climate data from 2016-2045 (L210). If the CONDENSED run has been re-equilibrated, then it is also in equilibrium with respect to the meteorological forcing, CO2, and N deposition + fertilization that correspond to 20162045. How do you separate the impact of land legacy from these other factors then? Please clarify or adjust your method.

This is an interesting point. However, the idea is to start a model in present-day conditions, you either need to provide it initial conditions, as in the TEMRESTART run, or reinitialize somehow. Two approaches to reinitialization are to use the 30-year average climate from, say, 1986-2014, or use those years for a dynamic equilibration, as I have done. Other conditions, such as CO_2, N deposition, N fertilization, and ozone, will be the values pertinent to those 30 years, as will, in fact, the climate. So, I think that is the appropriate and only way forward. CO_2, N deposition, N fertilization, and ozone changes all really ramped up in the latter part of the 20th century, and were nonfactors for most of the prior period. I have added the sentences "Note that the RESTART run will also incorporate effects of changing climate, CO_2, ozone, N deposition and fertilization, which cannot be captured in the CONDENSED run. (lines 179-181)" to acknowledge this point.

To separate out the effects of each would require factorial experiments running the model with only one factor, which is essentially what I did in the Felzer and Jiang (2018) study. However, to respond to this point further, I have done a new run (HISTCONST) in which land cover is held constant at 2014 values, so the difference between HISTORICAL and HISTCONST illustrates the effects of the other factors (note that since N fertilization is part of management, I do not list it along with N deposition when describing the environmental factors of change). I added this run into the Methods (lines 155-164) and Results (lines 275-285) sections, as well as included a new figure (Fig. 2) – that includes this run and another new run, HISTCOND, discussed below in response to the other reviewer.

L183 You Combine climate model data from one model, (MPI-ESM-P 1750-1900) with quasiobserved data (CRU, 1901-2014) and climate projections from another model (CCSM4, 2015-2099) in one continuous simulation. The more conventional approach is to conduct simulations that are either based on quasi-observed data or on data from one climate model. As for the quasi-observed climate data you could have used an early chunk of the historical data (e.g. 1901-1920) and spun up the model by iterating this climate data for whatever period it takes to equilibrate your model. This would have also freed you from the need of bias correction and downscaling MPI-ESM-P. The problem with your approach is that you combine data that come with their own set of biases and thereby mix the forcing from environmental factors with differences between these data sets. Please provide an explanation that justifies your experimental setup or adjust your method.

Please see explanation provided above.

L197 It would be more convincing if you had used radiation rather than cloud data for the historical period as well. The change in your method from one period to the next may create an unnecessary artificial forcing, which then mixes with the impact of climate change. Please justify your approach or adjust your method. T

This is a good point. However, net irradiance is not available in the CRU4.04 dataset, only clouds, which is why clouds were used. Added the following sentence: "The CRU4.04 data does not include irradiance, which is why it was necessary to use clouds for the historical

period, but since net irradiance is more directly used by the model, that was chosen for the future period" (lines 221-223).

L210 You write that the CONDENSED run is first equilibrated based on repeated use of the 2016-2045 climate. It is not clear to me why you use projected future climate conditions to equilibrate your model. Please explain or adjust your method.

True – as explained above, I did adjust the methods to use 1986-2014, and reran the CONDENSED run as a result.

L211 Please define NCE before using the acronym.

Sorry, now defined, with added sentence, "NCE is the NEP plus carbon lost through land-use conversion or by decomposition of agricultural or timber harvest products." (lines 252-254)

L218 Please mention in the text what experiments you are referring to.

I added (from HISTORICAL).

L238 You write that "reinitializing each grid is based on the assumption of NEP as close to zero". Should it not be Net Biome Productivity (NBP) rather than NEP that should be close to zero, as NBP also includes fluxes associated with disturbances, such as wildfires?

The TEM approach is to equilibrate NEP. In TEM (from McGuire et al. 2001), the NCE is the flux term that includes fluxes associated with disturbances. The equilibration procedure does not include disturbances, so that NEP = NCE during equilibration. This is an interesting point that I have long wrestled with – in fact, how to equilibrate a model with disturbances because you can't just run the model for a set historical period with known disturbances, but have to run for hundreds of years to reach equilibration. The assumption itself that ecosystems can be in a state of equilibration is not true, but it is a necessary assumption to establish some baseline of initial conditions at some known starting year.

L275 Please explain why you expect that more mature forested in CONDENSED would have lower NEP. Also, please mention what period you are referring to. Finally, replace "mature forested" with "mature forests", if this is what you mean.

I have added two new references (Besnard et al., 2018; He et al., 2012), but the basic idea is that NEP reaches its peak in mid succession and eventually goes back toward 0 or slightly positive for mature trees. That is a good point about the time period, so I added the sentence "By the end of the century regrowing forests in the RESTART run will still be younger than those in CONDENSED run, and 85 years is not enough time to reach full equilibration in the model." (lines 346-349). Corrected typo.

L278 The CONDENSED run has been re-equilibrated to the environmental conditions of 20152045. Is that not the main reason why vegetation carbon is 16% larger compared to

the RESTART simulation? Also, I don't recommend using the term bias here if you are not comparing against observation-based reference data (here and elsewhere).

Addressed this problem with the rerun of the CONDENSED experiment. Changed to "The larger values …" (line 351).

L280 Explain exactly what you mean by "fixes most of the problem".

Changed to "lowers the vegetation carbon so that it is close to that of using the full cohorts" (lines 353-355).

L364 A difference of 1.0% or 1.8% does not seem very large and may not even be statistically significant.

Good point. I looked at the last 30 years, and did a t-test at the 0.05 confidence level – turns out the CONDENSED rerun is not different but the TEMRESTART is, so changed sentence to: "The soil moisture of the CONDENSED run over the last 30 years is not statistically different from RESTART, while the TEMRESTART is 1.8% higher during that time period (Fig. 11a)" (line 444-447).

L391 I believe it is Chapin "III" et al. Also, this statement seems a little vague to me. Please explain what you mean.

Fixed, and added "such that more mature trees have higher amounts of vegetation carbon" (lines 475-476).

L411 Please mention what simulation you are referring to.

Added, "from the HISTORICAL run" (line 496).

L433 One reason why the modelled NEE IAV is smaller than observed may be related to the fact that your model does not represent mortality. That may be worth mentioning here as well.

I was actually trying to explain that the model does get within the correct range of measured IAV, and instead trying to point out that the entire time series shifts. Of course, if the values are still within the IAV, there are not significant changes. But I thought it still interesting to point out that the IAV did not change, only the means. If you feel this sentence should be removed, I can still do that.

L481 I assume that "models of the future" refer to models that that project future changes in vegetation dynamics? Please rephrase.

Changed to "models simulating the future" (lines 580-581).

L519 To decide whether a simulation is more realistic, you would need to evaluate your model output against some kind of observation-based reference data.

Changed to "achieve a result more consistent with a detailed representation of land-use cohorts. (lines 618-619)"

Tables

Table 1: Add information on how CO2, climate, and LULCC are treated in each experiment.

Done

Figures

Figure01 The time axis covers the period 1750-2014. Why does the caption say that the curves corresponds to the HISTORICAL and RESTART run, if the RESTART run starts in 2015?

Sorry, my mistake – it's just the HISTORICAL run.

S1b Please make larger to enhance readability.

Done