Reply to reviewer #1 by Thum et al. (The original comments by the reviewer are in violet and the replies by the authors are in black.)

Thum and colleagues present an interesting study looking at how different soil models influence atmospheric CO2 fluxes, which can be compared with ground and space-based observations. The study is thorough and well written, but in my estimation,it somewhat glosses over some of the caveats that should likely be considered with trying to evaluate component fluxes (here HR), with atmospheric CO2 observations that serve as a proxy for NEE.

We thank the reviewer for these in-depth comments and views. We hope we are able to address the concerns the reviewer is bringing up in our replies and in the revised version of the manuscript in a satisfactory manner.

Major concerns: Introduction, I found the basic primer on the global carbon cycle a little pedantic at times, with phrases like "Photosynthesis takes place in the green plant parts". As well as wandering, combining information on soil respiration and soil C stocks without really developing these ideas in a focused way that supports the direction or intent of the paper. I realize this is more stylistic that substantiative comment, but would encourage the authors to revise the introduction with focused topic sentences for each paragraph that introduces background literature and develop ideas in a way that focuses attention on the work to be presented.

We thank the reviewer for this advice in improving the introduction. We will re-write the introduction to be less pedantic and also take on the other good hints forward provided here. The reason we wrote the sentence "*Photosynthesis takes place in the green plant parts*" this way was to highlight the ending of this sentence: that the remote sensing is therefore able to detect the terrestrial photosynthesis.

I'm trying to wrap my head around how biases in the magnitude (and timing) of GPP simulated by JSBACH confound results presented here

In the JSBACH simulations the biosphere has been first spun-up to steady state in 1860 and the current land sink is resulting from the simulation period between 1860 to present. In 1860 the global net land CO_2 balance is therefore zero, and if the gross primary production is then overestimated, the autotrophic and heterotrophic respiration are then also overestimated.

and I fundamentally disagree with the statement on Line 392.

The sentence in line 392 is: "*The biases between XCO_2 from satellite retrievals and the model results originating from the JSBACH simulations are relatively large and this is likely caused by the use of a posteriori ocean fluxes from the CTE2016.*"

The biases we talk about in the sentence in line 392 is connected to the absolute differences in the CO_2 concentration values, that were reported in lines 294 and 322 [we should have mentioned also the site level observations, and not only the satellite observations in this sentence].

This bias we mean here is not connected to the sink or source terms separately but to the net land sink, that we report in Table 2 to be -1.68 PgC for CBA and -1.75 for YAS, and how it differs from the net land sink from the CTE2016 framework. With Fig. S10 we aimed to demonstrate how development of the bias in the absolute CO_2 values is in line with these different net land sink estimates, as we have explained in lines 394-397.

Re-writing the sentence in line 392 requires emphasizing better which bias we are meaning. We agree with the reviewer, that within the JSBACH model the modelled GPP is overestimated and the timing doesn't match perfectly. We will bring this up more in the new version of the manuscript, adding that also the bias of GPP influences the modelling skill of the system and not only the performance of the soil carbon model itself.

Should SI4 and Fig 2 be combined into one display item in the main text, using the same y-axis units for both?

The Fig. 2 displays a soil carbon map and Fig. S4 the anomalies from the turnover rates. If we put the color code to be the same for the both models, the spatial patterns won't be anymore visible. Since the reviewer 2 requested to also show observation based soil carbon maps, we will add observations to Fig.2 and still keep Fig. S4 and Fig. 2 separate.

Could you show the annual cycle of NPP or HR globally, in addition to the latitudinal bins shows in SI2? [maybe this goes into SI].
Would similar plots of regional & global AR or NPP also be helpful (in SI)?

We will present a plot of NPP, heterotrophic respiration (HR) and NEE, both globally as well as divided into latitudinal bins. Additionally we add a plot of autotrophic respiration (AR) to the supplement, showing both the global and latitudinally divided yearly cycles.

It's not really clear that the magnitude seasonal cycle of HR with YAS is so large in mid latitudes (30-60N; Fig 4), and is of equal and nearly opposite magnitude to the high GPP biases in this region, resulting in a lower than observed NEE (Fig 6-7).

Magnitude of the seasonal cycle at these latitudes by YAS is 2.5 Pg month^{-1}, and the GPP bias in this region is approximately 0.05 PgC day^{-1} = 1.5 PgC month^{-1}. How the difference between the heterotrophic respiration from the YAS and CBA models influence the global and latitudinally segregated NEE values will be shown in a figure in the revised manuscript.

I'm wondering how any version of JSBACH captures appropriate seasonal peak to trough dynamics of NEE Fig 6 given biases in the magnitude of GPP fluxes? This means that the timing and or magnitude of AR or HR fluxes must be compensating to generate reasonable seasonal cycle of NEE.

The annual magnitudes of the autotrophic and heterotrophic respiration have been given in Table 2 in the first version. The autotrophic respiration is relatively large. In the revised version of the manuscript we will show the annual cycles of NPP and NEE globally and in latitudinal bins and autotrophic respiration similarly in the supplementary material. The compensation by the autotrophic and heterotrophic respirations become visible in these plots and we will also discuss these plots in the revised manuscript.

Given known biases in the seasonal cycles of GPP (Fig S2), what modifications are needed to improve the representation of plant and soil dynamics in JSBACH?

In the data assimilation study by Castro-Morales et al. (2019), where satellite observed FAPAR and atmospheric CO_2 molar fractions were used in the assimilation, the maximum LAI value in the tropics was lowered via optimization. This itself will not enhance the seasonal cycle in the tropics, but would bring its absolute value down. To improve the seasonality of the tropical GPP, re-parametrization of the phenology model for the tropical plant functional type would likely help.

The timing is not so much off in the region 30ºN - 60ºN, but for the phenology during senescence period there might be room for improvement. North of 60º the modelled GPP is peaking too late, but the start of the growing season is occurring at the same time with the FLUXCOM results.

A study done for Finland, comparing growing season onset by satellite observations to JSBACH simulation output found out that it was enough to improve the onset of the deciduous forest by re-parameterizing the phenology module, however, to improve the onset of the coniferous forest it was necessary to add seasonality to the temperature responses of the photosynthesis parameters (Böttcher et al., 2016). It is therefore not so straightforward to say, which changes would improve the seasonal cycle of GPP. Improving the phenology cycle might be a step forward, but in the tropics how the plants experience dry conditions might be also off.

For the soil dynamics, based on the results of this study, we strongly recommend moving towards using the soil moisture as a driver for the YAS model. This was already mentioned in the earlier manuscript version.

In summary,could one erroneously chose an inferior soil biogeochemical model that give 'better' NEE fluxes with atmospheric observations, but that's fundamentally just masking over /compensating biases in GPP? This is, 'getting the right answer for the wrong reasons?

The reviewer is right, that the biased GPP might influence the results. Therefore we will modify the wordings of this manuscript so that we will not talk about benchmarking, but will emphasize that the aim was to see how well we can try to assess the differences between the soil carbon formulations within this kind of system and acknowledge that when it comes to ranking the models, we'd need a more data-driven system such as testbed, where the soil carbon models can be forced with certain plant productivity inputs (Wieder et al., 2018) or a systematic assessment of several variables against observations. The aim of this study was comparison of the two soil carbon models, which both had the same GPP input, and while the GPP bias compared to observations does have implications for numerical benchmarking, it does not take away the importance and conclusions of this work.

While the GPP of the JSBACH is not a perfect match to observations, it is the same in the both model simulations and the aim was to evaluate the soil carbon modules within a global land surface model, that includes several different process descriptions, e.g. also fires and land use changes. JSBACH is also part of the Earth System Model of the Max Planck Institute and an IPCC model, and it is state-of-the-art land surface model.

Throughout on display items, the shades of grey make it kind of hard to distinguish models and observations. This is especially true for Figs 5-7). Is there any harm in using colors?

We will add colors to all of the figures that were black and white in the earlier version.

Minor and technical concerns: There are enough abbreviations in the text that it some what distracts from the readability of the manuscript. I'd encourage removing some of these less standard ones if possible (e.g. SCA, GAW).There are also several very short paragraphs (even just one sentence long), that should either be further developed or merged into related paragraphs.

We will use less abbreviations in the second version of the manuscript and pay attention to the length of the paragraphs.

L19: the IAV of fluxes are never communicated here (although they could be easily brought into text and display items.

We have added the annual values to the plots that show seasonal cycles of NPP, TER and NEE.

L26: Maybe remove ‘natural’

Removed.

L39: I might include van Gestel et al’s 2018 critique of the Crowther paper to make this point.

Thank you for bringing up this point, we will do this.

L56: The connection between benchmarking global soil C models (the topic of the lastparagraph) and global CO2 flux observations is a little rocky and unclear. Reading between the lines, I think it’s a very good idea- but the connections about how / why it may be considered a useful way to evaluate soil biogeochemical models should be clarified. That said, others have recently used a similar approach (Basile et al 2020), which could be useful in contextualizing the introduction of the present work.

Thanks, we will use this paper in bringing this work better to the context and discuss better the link between the connection of atmospheric CO_2 molar fractions and soil model evaluation.

I’m also assuming the authors will discuss some of the assumptions being made in evaluating HR fluxes with atmospheric CO2 observations that may include biases in the timing and magnitude of GPP and AR fluxes from (JSBACH), potential errors imparted by the atmospheric transport model (TM5), or challenges in interpreting total column CO2 observations- especially from space. Maybe it’s worth foreshadowing some of these in the introduction?

We thank the reviewer for this insight, and find the idea of bringing them up in the introduction a good idea. We will do this.

See also refs from Keppel-Aleks et al (below).

We also thank the reviewer for pointing to these useful references that we'll make use of in the revised version of the manuscript.

Line 67: this sentence seems awkwardly phrased, maybe drop ‘in’ and ‘far’.

We'll rephrase this sentence.

Line 84: Is there reference for TM5, or example of where / how it’s been used?

In the earlier version we had added more information about TM5 only in section 2.3. We will in the new version add some references already here in the introduction.

Line 97: Randerson et al 1997 found that the timing of litterfall was important for controlling the timing of HR fluxes. Is the same true in JSBACH? How well does the model simulate this phenology?

As shown in Table 3, the decomposition is for a large part regulated by the environmental conditions. For completeness, we now checked the correlations between litterfall and heterotrophic respiration (HR) similarly as we did for the environmental drivers in Table 2: There is a positive significant correlation between the litterfall and heterotrophic respiration in region 30 N - 60 N with both of the model versions. The increase of heterotrophic respiration is anyhow preceding the litter flux. We will add a plot of this to the supplement and discuss this in the manuscript.

Interesting point in the study by Randerson et al is that changing litter quality would be making changes to the decomposition they are seeing in their study. This is something that we could actually test with the YAS model and we will also mention this in the new version of the manuscript.

Line 140, I realize it's likely in your previous publications, but is it worth noting how litter chemistry from JSBACH PFTs is translated onto the YASSO litter quality definitions? The way this is presented in kind of confusing & disconnected

We will clarify this in the revision of the manuscript. The division of the incoming litter from the JSBACH model per PFT is based on observations from different ecosystems (Trofymow et al., 1998; Berg et al., 1991a, 1991b; Gholz et al., 2000). We will add this information to the new version of the manuscript.

Line 143 define PFT

Thank you, we provide a definition for it here.

Line 159 constraints for what?

This comment refers to the sentence: '*The 3-hourly meteorological fields from ECMWF ERA-Interim (Dee et al., 2011) were used as constraints.*' We will modify this sentence to say that the TM5 model is run with the 3-hourly meteorological data from ERA-Interim.

Line 175 is 'atmospheric' redundant here?

Thanks, removed.

Line 205: Should this be S2.

Thanks, corrected.

Line 250: Why run statics on YAS fluxes and alpha, when the model uses precipitation to moderate decomposition rates (eq5)?

As explained in the manuscript in line 378, the precipitation is used as a driver in YAS, since it has been considered to estimate the soil moisture. While this might be true at the annual timescales, in which this model has been originally developed, this comparison here in Table 3 shows that it is not justified at monthly scale. We have commented this in the earlier manuscript version in line lines 384-386 with: "*Precipitation has been originally used in the YAS model as a proxy for soil moisture, since enough accurate soil moisture observations for model development haven't been available. Clearly, this idea needs reconsideration as our results show that at zonal spatial scales and monthly temporal scale the Rh from YAS is not at all correlated to soil moisture variable α.*" in the Discussion and in line 423-424 in the Conclusions: "*This suggests that use of precipitation as a proxy for soil moisture might not be sensible in sub-annual time scales.*"

Are the correlations between environmental drivers and HR fluxes really that surprising or interesting (Table 3, Line 241-254 & 366)? The models have these assumptions a priori (Methods, Fig. S1). In places with large seasonal variation in soil or air temperature (arctic), temperature is important control over seasonal HR fluxes. In places with little annual variation in temperature (tropics) moisture is a more important control. I'm not really sure what readers are supposed to learn from this analysis?

While we do have these clear response functions that are driving the soil carbon models, it is not necessarily clear at which part of the response we're in these different ecosystems. E.g. in the tropical zone, area 10ºS - 0º in Fig. S7, the YAS model is actually reaching the saturation level in respect to moisture.

And as the reviewer later mentions, the litter flux could play a role, but this is not a role in our model and these high correlation values to the environmental drivers already suggest that. These correlations also function as comparison for the soil moisture vs. YAS Rh, as mentioned.

Fig 6: Dotter is not a word.

Thanks, it was a typo. In the revised version we will have colors in this figure and not have a dotted line there.

Fig 6 & 7, can uncertainty estimate (or interannual variability) be shown for observations?

Thanks for the suggestion, we will add interannual variability of the seasonal cycle amplitude to the plots.

Table 3, why not just report r values, so negative correlations can be more clearly illustrated. Can't statistically significant correlations be highlighted (not just results withhigh r2)?

Thanks for this suggestion, we will do so.

L345, I'm not sure better agreement with CMIP5 models is necessarily a good thing,based on Kathe's work. Moreover, the calculation of global turnover times seems to mask important regional patterns. Instead, see Koven et al. 2017.

We considered to add this metric here, since it has been used, but this was more to complement the Figure S4 that was showing a map of the turnover rate anomalies by the two models. We agree with the reviewer that a global value does not include many important features and add here also a plot similar to Koven et al. 2017, where the turnover rates are shown as a function of air temperature and the precipitation is visible via coloring.

L395: where are results supporting these claims shown?

The sentence in this line is: '*The global land sink of JSBACH is approximately -1.7 PgCyr −1 (Table 2) and therefore the used ocean fluxes cause a bias to the simulated atmospheric CO_2 molar fraction.*'

We used posteriori ocean fluxes from the CTE2016 in this study. These ocean fluxes have been optimized using the terrestrial carbon cycle of the CTE2016, the SiBCASA-GFED4 model (van der Velde et al., 2014), and the fire emission fluxes that have been estimated from satellite observed burned area (Giglio et al., 2013). The fossil fuel emissions have not been optimized.

Since the optimization has been done with this other set of terrestrial carbon fluxes, this would be the likely cause for the bias. The fire fluxes of the CTE2016 are also different to JSBACH, so that could be another source for the discrepancy. We will modify this paragraph to add this point and also to clarify why we consider the ocean fluxes to be causing this bias.

L405: This doesn't seem like a standalone paragraph, nor is it really clear how it relates to the results being discussed regarding carbon tracker and JSBACH land C uptake.

Thanks for noting this, we will tie this information to the other paragraphs.

L408: what biases are not important here?

We admit that this line is unclear. We were referring here to the biases in the absolute CO_2 molar fraction values, and that these are not important, since the differences in the absolute values are not playing a role, since we concentrated in this analysis on the seasonal cycles and only mentioned the bias in the absolute CO_2 for completeness. We will rephrase this sentence.

L414: while I agree with this statement, however it's not done in the work presented. Multiple benchmarks, however, could be presented, again see Koven et al. 2017, or Todd-Brown's work that's already cited). Indeed it seems for a flux based analysis like this some more rigorous evaluation of upstream fluxes (e.g. GPP) is pretty important.

The sentence on the line is: "*We demonstrated how atmospheric CO_2 observations can be used to benchmark soil carbon models and that it is important to benchmark models across several different variables.*" We will remove the word benchmarking from the revised version of the manuscript, since we did not perform rigorous numerical benchmarking in this study, but were evaluating two different soil carbon models within one global land surface model within the constraints given by that system.

References:

Berg, B., et al. (1991a), Data on needle litter decomposition and soil climate as well as site characteristics for some coniferous forest sites. Part I. Site characteristics, Rep. 41, Dep. of Ecol. and Environ. Res., Swed. Univ. of Agric. Sci., Uppsala, Sweden.

Berg, B., et al. (1991b), Data on needle litter decomposition and soil climate as well as site characteristics for some coniferous forest sites. Part II. Decomposition data, Rep. 42, Dep. of Ecol. and Environ. Res., Swed. Univ. of Agric. Sci., Uppsala, Sweden.

Böttcher, K., Markkanen, T., Thum, T., Aalto, T., Aurela, M., Reick, C.H., Kolari, P., Arslan, A.N. and J. Pulliainen 2016. Evaluating Biosphere Model Estimates of the Start of the Vegetation Active Season in Boreal Forests by Satellite Observations. Remote Sensing, 8, 580.

Castro-Morales, K., Schürmann, G., Köstler, C., Rödenbeck, C., Heimann, M., and Zaehle, S.: Three decades of simulated global terrestrial carbon fluxes from a data assimilation system confronted with different periods of observations, Biogeosciences, 16, 3009–3032, https://doi.org/10.5194/bg-16-3009-2019, 2019.

Dalmonech, D. and Zaehle, S.: Towards a more objective evaluation of modelled land-carbon trends using atmospheric CO_2 and satellite-based vegetation activity observations, Biogeosciences, 10, 4189–4210, https://doi.org/10.5194/bg-10-4189-2013, 2013.

Gholz, H. L., D. A. Wedin, S. M. Smitherman, M. E. Harmon, and W. J. Parton (2000), Long term dynamics of pine and hardwood litter in contrasting environments: toward a global model of decomposition, Global Change Biol., 6, 751–765, doi:10.1046/j.1365-2486.2000.00349.x.

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), J. Geophys. Res.-Biogeo., 118, 317–328, https://doi.org/10.1002/jgrg.20042, 2013.

Trofymow, J. A., et al. (1998), The Canadian Intersite Decomposition Experiment (CIDET): Project and site establishment report, Inf. Rep. BC-X-378, Pac. For. Cent., Victoria, B. C., Canada.

van der Velde, I. R., Miller, J. B., Schaefer, K., van der Werf, G. R., Krol, M. C., and Peters, W.: Terrestrial cycling of 13 CO2 by photosynthesis, respiration, and biomass burning in SiBCASA, Biogeosciences, 11, 6553–6571, https://doi.org/10.5194/bg-11-6553-2014, 2014.

Wieder, WR, Hartman, MD, Sulman, BN, Wang, Y P, Koven, CD, Bonan, GB. Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Glob Change Biol*. 2018; 24: 1563– 1579. https://doi.org/10.1111/gcb.13979