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. Therefore in 1860 the global net land CO_2 balance is zero, and if the gross primary production is then overestimated, the autotrophic and heterotrophic respiration are then also overestimated. See below for more discussion on the mean and seasonality biases and their relevance for comparing our two model formulations.

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 (Fig. S13 in the revised version) 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 (against FLUXCOM, but more in line with the estimations from Welp et al. (2010), as mentioned in the manuscript) 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 (see below for the places and excerpts stating this).

The new version of the sentence that was previously in line 392 is now (located in lines 476-477):

"The differences in absolute CO_2 *and* XCO_2 *levels against the surface observations and the satellite retrievals, respectively, with modelled* CO_2 *are caused by the modelling system, but this bias does not influence the analysis performed."*

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.

In the revised version of the manuscript the Fig. 2 is now Fig. S11 (including soil maps also from data based estimates from SoilGrids and WHSD) and the previous Fig. S4 is now Fig. 4.

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 now show a plot of NPP, heterotrophic respiration (HR) and NEE, both global cycles (Fig. 2) as well as divided into latitudinal bins (Fig. 3). Additionally we have added a plot of autotrophic respiration (AR) to the supplement (Fig. S2), 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 is shown in the figures 2 and 3 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 and revised versions. The autotrophic respiration is relatively large. In the revised version of the manuscript we will show the annual cycles of NPP and NEE globally (Fig. 2) and in latitudinal bins (Fig. 3) and autotrophic respiration similarly in the supplementary material (Fig. S2). The compensation by the autotrophic and heterotrophic respirations become visible in these plots and we 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 the 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. The nutrient cycles were not modelled in our study. It is likely that including the nitrogen cycle the GPP values would be lower, as some vegetation would be nitrogen limited.

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.

It is not straightforward to come up with a solution for fixing the GPP, as that would be a study on its own and without thorough analysis and further model development (being out of the scope for this paper) it is not obvious which solutions would be required. We have now added to the lines 426-431 the following text:

"*Furthermore, the high GPP values predicted in the current run would likely be lower, if the nutrient cycles of nitrogen and phosphorus were included in the used version of JSBACH. Beside using a JSBACH version with nutrient cycles, further development work in the phenological cycle could improve the estimated GPP. The difference of the modelled GPP to the FLUXCOM product (Fig. S9) suggests that the maximum leaf area index might be overestimated in the tropics. Also, the timing of the phenological cycle north of 60 N might benefit from re-parametrization.*"

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. We now acknowledge that when it comes to ranking the models, we would need a more data-driven system 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 a 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 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.

To make this point raised by the reviewer in the manuscript, we have written in lines 423-426, when discussing the JSBACH results:

"*That GPP of JSBACH is biased high compared to observations is likely of secondary importance to our study comparing two model formulations, because GPP was the same for both formulations and the GPP bias did not lead to strong biases in the seasonal cycle predictions in different latitudinal zones were (Fig. S9). However, to assess the absolute skill of each model formulation in terms of net ecosystem exchange, GPP biases need to be reduced.*"

And to lines 472-475, when discussing the atmospheric CO_2 results:

"*The simulated GPP had a larger magnitude and some bias in its seasonal cycle, and therefore its evaluation against atmospheric* CO_2 *observations is influenced by it. Even though the atmospheric observations provide a valuable and informative comparison for the model results, their use as a benchmark metric needs careful consideration.*"

and have added to the Conclusions in lines 513-514 the following:

"*Also, the evaluation was done within a land surface model that is biased in its GPP predictions when compared to an upscaled GPP product and this hampers the use of atmospheric* CO_2 *as a numeric benchmark.*"

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 have added 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 have removed the abbreviations for SCA (with some exceptions in the figures to save space) and GAW. We have now paid attention to the length of the paragraphs and have developed short ones or merged them with other paragraphs, as suggested.

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 have now added a sentence (lines 27-29):

"*For example, while Crowther et al. (2016) concluded in data-based analysis that large carbon stocks will lose more carbon due to warming conditions, but van Gestel et al. (2018) questioned this view by an analysis based on more comprehensive dataset.* "

L56: The connection between benchmarking global soil C models (the topic of the last paragraph) 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 have now used the Basile paper in bringing this work better to the context. We have now improved in the introduction section about 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. In the new introduction we have now introduced these points and they are also mentioned in the discussion.

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

We also thank the reviewer for pointing to these useful references that we have used in the revised version of the manuscript. We used the paper by Keppel-Aleks et al. (2011) to highlight the influence of several factors to the column XCO_2 profiles. We referenced to Keppel-Aleks et al. (2013) as a work that has used column XCO2 profiles in model evaluation.

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 have written in the revised version already to the introduction the following sentence (line 55):

"*In this work we used a three-dimensional atmospheric chemistry transport model TM5 (Krol et al., 2005, Huijnen et al.,2010.*"

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 3: There is a positive significant correlation between the litterfall and heterotrophic respiration only in region 30 N - 60 N with both of the model versions. The increase of heterotrophic respiration is anyhow preceding the litter flux. We have added a plot of this to the supplement (Fig. S8) and discuss this in the manuscript.

An 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 have now mentioned this and the previous point in the Discussions section (lines 403-407) in the revised version of the manuscript as:

"*The observations show that the litterfall has strong influences on heterotrophic respiration (Chemidlin Prevost-Boure et al., 2010), but this process is not included in the used models, so at seasonal timescales in the different latitudinal zones no clear influence of litterfall driving* R_h *was seen. However, changes in the chemical composition of litterfall is considered to be potential reason for changes in the amplitude of atmospheric* CO_2 *(Randerson et al., 1997) and this is something we could study with the YAS model.*"

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 have clarified 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 have added 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 have modified this sentence to say that the TM5 model is run with the 3-hourly meteorological data from ERA-Interim (lines 158-159):

"The 3-hourly meteorological fields from ECMWF ERA-Interim (Dee et al., 2011) were used as forcing to run the TM5 model."

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 be a proxy for soil moisture. Precipitation has the advantage that it is much easier to observe than soil moisture and was thus the applied as driving variable in the development of the heavily observation-based YASSO. While an approximation of soil moisture effects by precipitation works relatively well 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 (new version lines xx) 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.*” [New version, lines 415-418: "*Precipitation was originally used in the YAS model as a proxy for soil moisture, since enough accurate soil moisture observations for model development were not available. Clearly, this idea needs reconsideration as our results show that at zonal spatial scales and monthly temporal scale, Rh from YAS is not at all correlated to the soil moisture.*"]

and in the Discussion and in line 423-424 (new version line 520) 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. S5, 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 R_h, 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 have now added interannual variability of the seasonal cycle amplitude to the plots. Additionally we have calculated an estimate for the uncertainty in the observations so that we calculated the difference between the fit to the observations and the actual observations and estimated the standard deviations from these values for each day with the observations. This is now represented in the figures as shaded area.

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 have now done so in the revised version.

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 adding 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 (now Fig. 5), where the turnover rates are shown as a function of air temperature and the precipitation is visible via coloring. We also removed the sentence from the conclusions that was re-stating the sentence, so that it has now less emphasis.

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 (see also above). The fire fluxes of the CTE2016 are also different to JSBACH, so that could be another source for the discrepancy. We have modified this paragraph to add this point and also clarified 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 have added the information content to a paragraph that discusses the results from GOSAT and flux observations.

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 have rephrased this sentence to be:

"The differences in absolute CO_2 and XCO_2 levels against the surface observations and the satellite retrievals, respectively, with modelled CO_2 are caused by the modelling system, but this bias does not influence the analysis performed."

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

Reply to reviewer #2 by Thum et al. (The original comments by the reviewer are in violet and the replies by the authors are in black.)

Thum et al. present an interesting approach for the evaluation of soil carbon models in a land surface model by using atmospheric CO2 observations. I like the basic idea of the study and the work is methodological sound. However, I actually got bored and disappointed when I was reading the paper. This is too a large degree caused by the presentation (text and figures) of the material (I agree with all the points by reviewer 1):

We thank the reviewer for these views on the paper.

1) It is not clear what the purpose of the study is. Only in the discussion it's written that the "aim was to use atmospheric observations to benchmark soil carbon models". If this was the aim, the evaluation of just two modules of JSBACH is insufficient (and causes my disappointment). It would be better to clearly state already in the introduction that the aim was to evaluate two soil modules of JSBACH.

We have removed the word "benchmark" from the manuscript, as the aim of this paper is not to do a purely numerical ranking of two different models. We change the title to include 'JSBACH', therefore the scope of this paper should be then visible already in the title.

However, if this is the case,the manuscript would do better as a model evaluation paper in Geoscientific Model Development. Generally, the text is written as a model evaluation study and I don't find important results for the general Biogeosciences. My feeling is that the paper should go beyond a model evaluation and include some more substantial scientific questions, hypotheses and findings in order to fit into Biogeosciences.

The aim of this study was to provide a proof of concept that atmospheric CO_2 concentrations are a useful variable to evaluate soil carbon models and to apply this to two different model formulations of JSBACH to reveal what we can learn about the functionality of the soil carbon models. Our aim is not a rigorous numerical benchmarking, but to see what causes differences we see in the model results versus the observations.

Furthermore, we do address scientific questions in this study, such as whether the soil carbon content or environmental responses of the decomposition are more responsible for the differences seen in the results. We therefore see our manuscript well suited for Biogeosciences. In writing the revised version of the manuscript we will include the following scientific questions this study addresses to the introduction:

**How do can we use a land surface model together with a transport model to evaluate soil carbon model and what problems do we face when doing that?*
**What is the role of soil carbon stocks, the variables driving their decomposition and the functional dependencies of those variables on modelled heterotrophic respiration at global scale and how this leads to differences in the atmospheric CO_2 signal?*

In the revised version of the manuscript we now better highlight how the changes in the heterotrophic respiration influence the modelled terrestrial NEE, making it more visible how the process of changes in heterotrophic respiration transfers to the atmospheric signal. This therefore better illustrates the process chain and connects better the terrestrial modelled carbon cycle to the atmospheric CO_2 signal. Since the representation of the land carbon cycle and its gross fluxes are similar across most state-of-the-art DGVMs, such analysis provides more universal insight than just a JSBACH evaluation. Also we show environmental dependencies of the modelled turnover rates, which brings them better into context of ecological understanding. Additionally we show a latitudinal comparison of the modelled soil carbon stocks to the observation-based estimates. We believe that all these new contributions (suggested by the reviewers) tie this manuscript better in line with the Biogeosciences.

2) If the “aim was to use atmospheric observations to benchmark soil carbon models” (as stated), I would expect a more detailed description of the assumptions and a detailed analysis on how to use atmospheric CO2 observations for the benchmarking, including how to disentangle the contributions of GPP and Reco on CO2, the role of uncertainties in observations and atmospheric transport, and how different regions contribute to the CO2 seasonality. Especially the later points would help to potentially benchmark soil carbon simulations if different parts of the world.

We have paid more attention to the points raised by the reviewer in the revised version of the manuscript. It is not possible in our current set-up to disentangle the contributions of GPP and Reco, but the GPP is same for both of the model simulations and we re-wrote the text to remove using the word benchmarking, so that our main aim to see the contribution of two different soil carbon modules in the atmospheric CO_2 molar fraction becomes more clear.

We have added text on the observational uncertainty concerning the surface CO_2 observations (in Section 2.4), spaceborn XCO_2 observations and atmospheric transport uncertainties (in Introduction)**.** The atmospheric transport model was the same in the both model simulations, so we don’t expect this to be causing differences we see in the two different model runs. For the modelled atmospheric signal we also have contributions from the ocean fluxes and anthropogenic emissions. Errors in these estimates will also influence the atmospheric CO_2, but they were the same for both model simulations. The role of different regions contributing to the seasonality of the CO_2 molar fraction can be studied by setting the land fluxes to zero at certain regions, as done earlier by Dalmonech and Zaehle (2013). We did not for now consider this to be necessary for this analysis.

3) As already stated by reviewer 1, the text needs substantial rewrite. The text has no clear structure, topic sentences are missing, some chapters are too long (especially 1 and 3.1). For example, the first section of the results (3.1) report mainly minor results (including references to the supplement) but does not report the most important results. In addition, I recommend to split this section in further sub-chapters to improve the structure of the text.

Thank you for the advice in improving the re-writing of the manuscript. Concerning the order of the results section: This work was done in a step-wise manner, when first the JSBACH simulations were performed and they were then fed to TM5. Starting with the atmospheric CO_2 results (which we would consider being main results of the study) potentially makes it challenging for the reader to follow the storyline. Having said that, we bore this view in mind during re-writing and tried to highlight important results before minor results.

In order to achieve this, we put the flux results before the stock comparisons in section 3.1. Further, we have divided 3.1. into further sub-chapters (3.1.1. Flux comparisons, 3.1.2. Stock comparisons, 3.1.3. Box model) as suggested. The reviewer also noted that the introduction has been too long. We have tried to shorten it in the revised version. Unfortunately the section 3 that was also too long according to the reviewer is now longer, due to the suggestions from the reviewers, but we hope that using the subsections for this section makes it easier to follow. Effort has been made to add missing topic sentences.

4) Figures: I’m sorry, but reading the figures in the main text and in the supplement wasa nightmare! The figures are too small and the grey colours make it almost impossible to distinguish

the different model runs and observations. Please improve all figures.

We thank the reviewer for this insight and improve the figures in the revised version of the manuscript. We apologize the inconvenience -- the size of the figures was caused by following the template provided by the publisher and not noticing that it should have been wider for two column figures. We have now made the figures bigger.

5) The discussion of GPP is over-simplistic. JSBACH overestimates GPP and has in some regions shifts in the seasonality. Hence it remains unclear which soil carbon model is the better one because the comparison of CO2 seasonality is also affected by wrong simulations of GPP. Could it be an option to force more realistic GPP estimates into JSBACH or mix GPP from data-driven estimates with Rh from JSBACH in TM5?

As this study was also a methodological test to see if using different soil models would indeed result in differences in the atmospheric CO_2 cycle, we wanted to only use the 'off-shelf' version of JSBACH. This is also in line with our goal of a proof of concept study for evaluation of other land surface models, which would likewise likely not be optimized in their performance. It is obvious that using a full land surface model brings up constraints. What was not successful in writing of the earlier version, was to bear this in mind when interpreting the results. Overall, in this model evaluation exercise, a performance competition of the two different models was not the main goal, but to find out what characteristics of the soil carbon models need improvement and how their different structures influence the results and how the differences transfer all the way to the atmospheric CO_2 molar fractions. For the YAS model the finding that the precipitation should not be used as a proxy for soil moisture at this temporal and spatial resolution, is relevant and independent of JSBACHs GPP overestimation.

The reviewer has a suggestion of forcing more realistic GPP into JSBACH. This aim can be achieved by data assimilation, as done by Nyawira et al. (2016) and Castro-Morales et al. (2019) where they used other, earlier versions of JSBACH. However, this approach would be a study by itself to be done first.

As for using more data-driven GPP estimate, this unfortunately doesn't work in our simulation set-up, as this would break mass-balance of the terrestrial carbon cycle and the litterfall and GPP are strongly coupled in such a model. There would be several biases resulting from this. There certainly is a lot of value in more data-based approaches and that kind of work would be best done by a test-bed set-up similar to works by Wieder et al. (2018).

We have now added a point to the potential biases in the atmospheric CO_2 signal caused by the modelled GPP to two places in the discussion and once in the conclusions (see the responses to reviewer 1).

6) Please describe if permafrost was simulated in the JSBACH runs and how the simulation or non-simulation of permafrost contributed to soil carbon simulations.

Permafrost was not included in the simulations and we will mention this in the new version of the manuscript clearly. Recent model development adding permafrost to JSBACH includes only the YAS model (Castro-Morales et al., 2018). Including permafrost would increase the carbon stocks in high latitude regions. The exact influence on the atmospheric signal is speculation without the actual model runs, but it would be expected that the seasonal cycle of heterotrophic respiration in high latitude regions would be dampened, since the active layer producing heterotrophic respiration would be thinner.

We have now added to lines 447-452 the following text:

"This JSBACH version also didn't have permafrost described. If permafrost would be modelled, the seasonal cycle of heterotrophic respiration at high latitudes would likely be dampened, as the depth of the active layer determines the amount of soil capable of respiring. The YAS model has been used in a JSBACH version containing permafrost in a study concentrating on the Russian Far East (Castro-Morales et al., 2018). Both, CBA and YAS, were originally developed for mineral soils and for applications with organic soil, so model development and testing at smaller than global scale could be useful."

Figure 1: Explain the numbers in the caption.

Thank you for noticing this was missing. We have now changed the caption to be:

“Locations of GAW stations, denoted as black dots, and different TransCom regions (different numbers denote the different TransCom regions in this study) as different colors.”

Figure 2: Even if the soil carbon stocks have been already evaluated, it would be still helpful to add 1 or 2 maps from an observation-based product for comparison.

We have added maps from SoilGrids and HWSD (Fig. S11) products for the comparison with the soil carbon stocks, as suggested. Additionally, we added a figure showing a latitudinal gradient of the carbon stock, to better illustrate the differences between the modelled and observation-based estimates (Fig. S12).

Table 2: There seems to be a mistake in the results for TER, as those can’t be the same numbers.

Yes, thank you, the YAS number should be 155 PgC.

References

Castro-Morales, K., Kleinen, T., Kaiser, S., Zaehle, S., Kittler, F., Kwon, M. J., Beer, C., and Göckede, M.: Year-round simulated methane emissions from a permafrost ecosystem in Northeast Siberia, Biogeosciences, 15, 2691–2722, https://doi.org/10.5194/bg-15-2691-2018, 2018.

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.

Nyawira, S. S., Nabel, J. E. M. S., Don, A., Brovkin, V., and Pongratz, J.: Soil carbon response to land-use change: evaluation of a global vegetation model using observational meta-analyses, Biogeosciences, 13, 5661–5675, https://doi.org/10.5194/bg-13-5661-2016, 2016.

Evaluating two soil carbon models within ~~a~~ the global land surface model JSBACH using surface and spaceborne observations of atmospheric CO_2~~mole fractions~~

Tea Thum[1], Julia E. M. S. Nabel[2], Aki Tsuruta[3], Tuula Aalto[3], Edward J. Dlugokencky[4], Jari Liski[3], Ingrid T. Luijkx[5], Tiina Markkanen[3], Julia Pongratz[2, 6], Yukio Yoshida[7], and Sönke Zaehle[1]

[1]Max Planck Institute for Biogeochemistry, Jena, Germany
[2]Max Planck Institute for Meteorology, Hamburg, Germany
[3]The Finnish Meteorological Institute, Helsinki, Finland
[4]NOAA Global Monitoring Laboratory, Boulder CO, United States
[5]Meteorology and Air Quality Group, Wageningen University and Research, Wageningen, The Netherlands
[6]Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany
[7]Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan

Correspondence: Tea Thum (tthum@bgc-jena.mpg.de)

Abstract. The trajectories of soil carbon ~~(C) in the~~ in our changing climate are of utmost importance, as soil ~~carbon~~ is a substantial carbon ~~storage~~ reservoir with a large potential to impact the atmospheric carbon dioxide (CO_2) burden. Atmospheric CO_2 observations integrate all processes affecting ~~C~~ carbon exchange between the surface and the atmosphere ~~. Therefore they provide a benchmark~~ and therefore are suitable for carbon cycle ~~models. We evaluated~~ model evaluation. In this study, we present a framework for how to use atmospheric CO_2 observations to evaluate two distinct soil carbon models (CBALANCE and YASSO) that ~~were implemented to a~~ are implemented in the global land surface model (JSBACH)~~against atmospheric observations~~. We transported the biospheric carbon fluxes obtained by JSBACH using the atmospheric transport model TM5 to obtain atmospheric CO_2. We then compared these results with surface observations from Global Atmosphere Watch ~~(GAW)~~ stations as well as with column XCO_2 retrievals from the GOSAT satellite. The seasonal cycles of atmospheric CO_2 estimated by the two different soil models differed. The estimates from the CBALANCE soil model were more in line with the surface observations at low latitudes (0°N-45°N) with only 1% bias in the seasonal cycle amplitude~~(SCA)~~, whereas YASSO ~~was underestimating the SCA~~ underestimated the seasonal cycle amplitude in this region by 32%. YASSO, on the other hand, gave more realistic seasonal cycle amplitudes of CO_2 at northern boreal sites (north of 45°N) with underestimation of 15% compared to 30% overestimation by CBALANCE. Generally, the estimates from CBALANCE were more successful in capturing the seasonal patterns and seasonal cycle amplitudes of atmospheric CO_2 even though it overestimated soil carbon stocks by 225% (compared to underestimation of 36% by YASSO) and its predictions of the global distribution of soil carbon stocks was unrealistic. The reasons for these differences in the results are related to the different environmental drivers and their functional dependencies of these two soil carbon models. In the ~~tropical region~~ tropics, heterotrophic respiration in the YASSO model ~~showed earlier increase in season of the heterotophic respiration~~ increased earlier in the season since it is driven by precipitation instead of soil moisture~~as~~, as in CBALANCE. In ~~the~~ temperate and boreal ~~region~~ regions, the role of temperature

is more dominant. There~~the~~ , heterotophic respiration from the YASSO model had ~~larger annual variability~~a larger seasonal amplitude, driven by air temperature, compared to ~~the CBALANCE~~CBALANCE, which is driven by soil temperature. The results underline the importance of using sub-yearly data in the development of soil carbon models when they are used in shorter than annual time scales.

1 Introduction

The ~~global carbon (C) balance is mainly influenced by terrestrial and oceanic carbon fluxes. Since the early 1960s, when accurate measurements of atmospheric began, the land and ocean have absorbed approximately half of annual anthropogenic carbon dioxide () emissions (**?**). The terrestrial natural~~ terrestrial carbon cycle consists of uptake ~~by vegetation~~ of CO_2 by vegetation for photosynthesis and release of carbon by plants' autotrophic respiration, soil decomposition by heterotrophic organisms and natural disturbances (**?**). ~~At global scale, recent advances in remote sensing have allowed global spatially distributed estimates for gross primary production (GPP) generated together with machine learning methods (**?**). Photosynthesis takes place in the green plant parts, which can be detected from space as well as the sun-induced chlorophyll fluorescence that is related to photosynthesis (**?**). Release of carbon from the terrestrial biosphere remains more elusive, since below ground processes cannot be directly detected globally and the understanding of the below ground processes is still developing (**?**).~~

~~For soil respiration, including both autotrophic respiration by roots and heterotrophic respiration, some global datasets have been developed (**??**). Recently efforts to estimate heterotrophic respiration have also been made using machine learning techniques, **?** and **?** being the first attempts. The separation of heterotrophic respiration from the autotrophic respiration remains still a challenge and in these two studies it has been done at yearly scales. The soil stores~~ Soils store twice as much carbon as the atmosphere (**?**) and its fate in changing climate remains uncertain (**?**). For example, while **?** concluded from a data-based analysis that large carbon stocks will lose more carbon due to warming conditions, **?** questioned this view with an analysis based on a more comprehensive dataset. To have reliable predictions of future carbon stocks, process-based understanding of the below ground carbon cycle ~~needs to be achieved.~~

~~Heterotrophic respiration, R_h, at global scale is a very important part of the carbon cycle. It is influenced by moisture and temperature conditions (**?**). In many global modelling approaches, moisture and temperature dependencies are influencing a first-order decay process of soil carbon pools (**?**). The magnitude of modelled R_h is therefore dependent on these parameters, the pool sizes and their turnover rates as well as the incoming litter input from vegetation. The role of the turnover rate is crucial, as the pre-industrial turnover rates are the key uncertainty in simulation of soil organic matter stocks for the 21st century (**?**). Soil carbon modelling needs improvements so that future changes in soil carbon stocks can be better predicted (**?**). Current global carbon cycle models give a wide range of estimates for changes in soil carbon by the end of this century (**?**)~~

~~. There is currently on-going development of soil carbon models, and these next generation models include detailed microbial dynamics with nutrient cycles (**??**).~~ is needed (**?**).

One way to ~~benchmark~~ evaluate soil carbon models has been ~~by using~~ to use observations of soil carbon stocks (**?**). At ~~smaller scales,gas exchange observations with~~ small scales,rates of gas exchange measured in chambers have also been used (**?**), but separation of heterotrophic and autotrophic respiration is laborious (**?**). It is anyhow challenging to find reasons for differences in heterotrophic respiration between large scale models, as the litter input to the soil influences heterotophic respiration and this litter input varies between the models. ~~Alternative~~ One way forward is to use a ~~test-bed~~ testbed for these models, as done by **?**.

An alternative, regionally integrated approach is using observations of ~~the~~ atmospheric CO_2, which integrate all processes involved in global surface-atmosphere ~~C~~ carbon exchange. The surface observation network of atmospheric CO_2 has been used in benchmarking global carbon cycle models (**???**). Recent advances of satellite technology have enabled retrievals of space-born dry-air total column-averaged CO_2 mole fraction (XCO_2), quantifying CO_2 in the entire atmospheric column between the land surface and the top of the atmosphere. These observations reveal a more spatially integrated CO_2 signal compared to surface site observations and together they provide a complementary dataset. These two data sources have been used together to study the carbon cycle with "top-down“ inversion modelling (**?**). This kind of modelling framework uses atmospheric CO_2 observations to constrain a priori biospheric and ocean fluxes, based on the Bayesian inversion technique, which results in optimized estimates (a posteriori) of the fluxes (**????**). ~~The estimates~~ Estimates for fossil emissions are often assumed as known, i.e., not optimized in the inversion.

In this study we ~~investigate in how far~~ present a framework how to use atmospheric CO_2 observations ~~, both from surface network and space-born observations, can be used to benchmark these two~~ to evaluate soil carbon models implemented in a land surface model. We apply this to two state-of-the-art soil carbon models ~~included in~~ as a "proof-of-concept" for a more universal application. **?** did similar work within a biogeochemical testbed and concluded that heterotrophic respiration can be a valuable benchmark in carbon cycle studies. They emphasized that the seasonal phasing of heterotrophic respiration relative to the net primary production influences the net ecosystem exchange and therefore potentially introduces bias to atmospheric CO_2 that hampers its use as a benchmark.

To obtain the atmospheric CO_2 profiles from our simulations with the land surface model ~~JSBACH,~~ we applied an atmospheric transport model. In this work we used a three-dimensional atmospheric chemistry transport model TM5 (**??**). Generally, transport models, such as TM5, contain errors caused by, for example, poorly resolved advection and heavily parameterized transport schemes (**?**). With TM5 we calculated the column averaged CO_2 that can be used to evaluate model results versus the satellite observations. Also satellite observations can include errors. The uncertainty for GOSAT observations has been estimated to be around 1 to 2 ppm (**??**). Contributors to uncertainties in the retrieval algorithms originate, for example, from the solar radiation database and handling of aerosol scattering (**?**). Last, also space-borne observations have uncertainties, and also the column XCO_2 profiles have influences from e.g. advection and global scale gradients driven by weather systems (**?**). A model evaluation performed with the column XCO_2 observations enabled a more thorough study of fluxes and atmospheric physics of the modelling system (**?**).

We use in this work JSBACH, the land surface model of the ICOsahedral Non-hydrostatic Earth system model (ICON-ESM), one of the models participating in CMIP6. ~~Previous studies evaluating the new YASSO soil carbon model performance when implemented in JSBACH, have shown better performance in relation to observations of soil carbon stocks when compared to the older soil carbon model of JSBACH, CBALANCE (??)~~We are interested in seeing if the two soil carbon models lead to markedly different CO_2 signals and to explore which conclusions on model performance and process representation can be drawn that could help to improve this land surface model (and potentially other similar models) and our understanding of the land carbon cycle. Since the only difference between the two model versions is the description of the underlying soil processes and include no major feedbacks between soil and vegetation in the model set-up (excluding ~~the~~ a small effect of litter accumulation on fire ~~fluxes~~emissions), the difference in the release of carbon to the atmosphere originates only from the soil carbon models. The two soil carbon models ~~used in this work have different environmental drivers. CBALANCE has~~ are both first-order decay models. However, they have different pool structures and environmental drivers and have differing response functions to these variables. CBALANCE uses soil moisture and soil temperature as driving variables and the YASSO model precipitation and air temperature. ~~The models have differing response functions to these environmental variables as well as different carbon pool structures, and they are both first-order decay models.~~ This framework allows us to investigate how these above-mentioned differences in soil carbon modes influence atmospheric CO_2. ~~To transfer terrestrial carbon fluxes from the surface to the atmosphere, we~~ We use the transport model TM5 and the anthropogenic and ocean fluxes from an inversion framework (?). In the analysis we also use a simple box model calculation to further understand the main causes in the different outcomes of the models. Specifically, we aim to answer the following questions:

- How do can we use a land surface model together with a transport model to evaluate soil carbon model and what problems do we face when doing that?
- What is the role of soil carbon stocks, the variables driving their decomposition and the functional dependencies of those variables on modelled heterotrophic respiration at global scale and how this leads to differences in the atmospheric CO_2 signal?

2 Materials and Methods

We used the land surface model JSBACH (?) to obtain ~~the~~ net land-atmosphere CO_2 exchange and fed ~~them~~ that, together with ocean, fossil and land use fluxes, into a transport model, TM5, which simulates ~~the~~ resulting atmospheric CO_2 ~~mole fractions~~ at selected surface ~~stations~~ sites as well as column integrated values for comparison to satellite derived column CO_2.

2.1 Model simulations: JSBACH with two soil carbon models

JSBACH is the global land surface model of the Max Planck Institute‘s Earth System Model (?), simulating ~~the~~ terrestrial carbon, energy and water cycles (?). In this study JSBACH was run with two different soil carbon sub-models ~~for soil carbon~~ that are described below. The older model, CBALANCE, has been used in CMIP5 simulations of JSBACH (?)~~and the~~. The newer model, YASSO, has been used in simulations for the annual global carbon budget (**??**) and is used in CMIP6 simulations of JSBACH (Mauritsen et al.,2019). It is ~~furthermore~~ also used in JSBACH4, a re-implementation of JSBACH for the ICON-ESM (**??**).

Independent of the ~~used~~ sub-model used for soil carbon, JSBACH uses three carbon pools for ~~the living vegetationcarbon~~living vegetation: a wood pool, containing woody parts of plants, and ~~a green and a reserve pool containing~~ green and reserve pools that contain the non-woody parts. JSBACH simulates different processes that lead to losses from the vegetation pools, such as grazing, shedding of leaves and natural or anthropogenic disturbances. Depending on the process, some of the vegetation carbon is lost as CO_2 into the atmosphere, while the remaining part is transferred as dead vegetation into the litter and soil pools of the ~~used~~ sub-model for soil carbon, where it is then subject to the internal processes of the soil carbon sub-model. The only process outside of the soil carbon sub-model ~~which~~ that influences dead material ~~are the fireprocesses~~is fire, burning parts of above ground litter carbon.

2.1.1 The soil carbon model CBALANCE

CBALANCE (CBA) is the original soil carbon sub-model of JSBACH (**?**), which has been used in CMIP5. The environmental drivers for decomposition in CBA are soil temperature (at soil depth of 30 to 120 cm below the surface) and ~~a~~ relative soil moisture (α) of the upper-most soil layer, which is 5 cm thick. ~~The~~ α varies between zero and one.

The function for soil temperature dependence, $f_{CBA,T_{soil}}$ of decomposition follows a ~~Q10~~ Q_{10} formulation as

$$f_{CBA,T_{soil}}(T_{soil}) = Q_{10}^{\frac{T_{soil}}{10^\circ C}} \tag{1}$$

with a Q_{10} value of 1.8 and T_{soil} is soil temperature in ~~Celsius~~ °C (shown in Fig. S1a) (**?**). The dependency on relative soil moisture α is linear (Fig. ~~S1 b~~S1b) and it is calculated as

$$f_{CBA,\alpha}(\alpha) = MAX(\alpha_{min}, \frac{\alpha - \alpha_{crit}}{1.0 - \alpha_{crit}}) \tag{2}$$

where α_{crit} is 0.35 and α_{min} is 0.1 (**?**).

Together these functions are modulating the rate of decomposition, so that the R_h from each pool (denoted by i) is

$$R_h(T_{soil}, \alpha) = f_{CBA,\alpha} f_{CBA,T_{soil}} \frac{C_i}{\tau_i} \tag{3}$$

where C_i is the carbon content of each pool and τ_i is the turnover rate of each pool in days. CBA uses five different carbon pools having different turnover times:

- Two green litter pools: one above- and one ~~below ground~~ below-ground in which the non-woody plant parts ~~are decomposed~~ decompose with turnover times between 1.8 and 2.5 years (**?**)
- Two woody litter pools: one above- and one ~~below ground~~ below-ground in which the woody plant parts are decomposed with turnover times of several decades
- One slow pool receiving its input from the four litter pools and its turnover time is ~~in~~ on the order of a century.

2.1.2 The soil carbon model YASSO

The original soil carbon model of JSBACH was replaced by YASSO (YAS) (**??**). JSBACH's YAS implementation is based on the Yasso07 model (**?**). Development of Yasso07 ~~has~~ relied heavily on litter bag and other observational data sets that ~~have been~~ were used to estimate ~~the~~ model parameters (**??**). Owing to its strong connection to experiments, its environmental drivers are quasi-monthly air temperature and precipitation.

The decomposition dependency on air temperature is

$$f_{YAS,T_{air}}(T_{air}) = e^{\beta_1 T_{air} + \beta_2 T_{air}^2} \tag{4}$$

where T_{air} is air temperature (°C), parameter β_1 is $9.5 \times 10^{-2}\ ^{\circ}C^{-1}$ parameter β_2 is $-14 \times 10^{-4}\ ^{\circ}C^{-2}$ (Fig. S1c).

The decomposition depends on precipitation P_a (m) as

$$f_{YAS,P_a}(P_a) = (1 - e^{\gamma P_a}). \tag{5}$$

where γ = -1.21 m^{-1} (Fig. S1d). The environmental drivers for YAS (precipitation and air temperature) are averaged for ~~a~~ 30-day ~~period~~periods.

Similar to CBA, YAS has slowly and ~~fast~~ rapidly decomposing pools, but its pool dynamics are more structured. ~~YAS uses 18 carbon pools, nine for the decomposition of woody litter and nine for the decomposition of~~ First, all the pools are divided into woody and non-woody ~~litter. The only~~ materials. The difference in the calculation of the decomposition rates between non-woody and woody pools is an additional parameter that increases the turnover rates ~~with increasing size~~ of the woody litter~~(?). In addition to the distinction into woody and non-woody litter which is also done in CBA,~~ , dependent on its size parameter (**?**), which is PFT-dependent.

YAS takes the chemical composition of the incoming litter into account. The incoming litter is divided to different chemical pools according to the PFT dependent factors. Information for the PFT dependent factors for the litter decomposition has been derived from observations (**????**). YAS uses four chemically distinct ~~pool kinds~~pools: acid soluble, water soluble, ethanol soluble and non-soluble. For each of these four chemical compositions one ~~above and one below ground pool are~~ above- and one below-ground pool is used. In addition there is one humus pool ~~for woody and one for~~ (divided to woody and non-woody ~~material. The dynamics~~ pools as all the other pools). Dynamics of the YAS carbon pools are described in (**?**) with decomposition fluxes causing redistributions among the ~~different~~ pools or losses to the atmosphere. Each of the pools has a decay constant, which is modified by the environmental dependencies in Eqs. (**??**) and (**??**).

~~Each PFT used in JSBACH has a distinct chemical composition. Furthermore, the branch size of the woody litter is PFT-dependent.~~

2.2 The model simulations: The JSBACH set-up

JSBACH model simulations followed the TRENDY v4 protocol in terms of JSBACH version, simulation protocol and forcing data (**??**). Climate forcing was based on CRUNCEP v6 (**?**) and ~~the~~ global atmospheric CO_2 was obtained from ice core and

~~NOAA monitoring station data~~ National Oceanic and Atmospheric Administration (NOAA) measurements (**?**). For ~~both set-ups~~ each set-up, the model was ~~separately run into~~ run to equilibrium, i.e. until the soil carbon pools of the applied carbon sub-model were at steady-state. The two different transient simulations were then done for ~~the period~~ 1860 to 2014. Anthropogenic land cover change was forced by the LUHv1 dataset (**?**) and was simulated as described in **?**. While fire and windthrow were simulated, natural land cover changes and the nitrogen cycle were not activated. Simulations were done at T63 spatial resolution (approximately 1.9° or, 200 km). For further details on the spin-up and the model version please refer to the SI.

2.3 The model simulations: TM5

To estimate atmospheric CO_2, we used the global Eulerian atmospheric transport model TM5 (**??**). TM5 was run globally at 6° x 4° (latitude x longitude) resolution with two-way zoom over Europe, where the European domain was run at 1° x 1° resolution. This is the pre-existing set-up that was readily available. The 3-hourly meteorological fields from ECMWF ERA-Interim (**?**) were used as ~~constraints~~forcing to run TM5. Linear interpolation was done to obtain CO_2 estimates at the exact locations and times of the observations.

We fed TM5 ~~with~~ daily biospheric, weekly ocean and annual fossil fuel fluxes to obtain realistic atmospheric CO_2. Values of GPP and total ecosystem respiration were taken from the JSBACH simulations for the two different soil model formulations. Also, carbon release from vegetation and soil owing to land-use change, fires and herbivores were taken from the JSBACH model results as part of terrestrial biosphere carbon fluxes. In addition, we used the posterior biospheric flux estimates from CarbonTracker Europe (CTE2016, later ~~referenced~~ referred to as CTE; **?**) to provide some guidance on the ability of TM5 to represent the individual site observations. The ocean fluxes were the a posterior estimates from the same study.

Fossil fuel emissions are from the EDGAR4.2 Database (**?**) and Carbones project (http://www.carbones.eu), with scaling to global total values as for the Global Carbon Budget as described in **?**. The annual fossil fuel flux to the atmosphere was approximately 8.63 $\mathrm{PgCyr^{-1}}$, and ocean uptake of carbon was approximately 2.33 $\mathrm{PgCyr^{-1}}$ when averaged over ~~years~~ 2001-2014. Annual values are summarized in Table S1.

Simulations with TM5 were done for 2000-2014~~, but the first year was considered as spin-up and omitted from the analysis~~.

2.4 The surface observations

Surface observations of atmospheric CO_2 from NOAA weekly discrete air samples (ObsPack product: GLOBALVIEWplus v2.1; **?**) were used to evaluate the effect of different soil carbon models on tropospheric ~~atmospheric~~ CO_2 seasonal cycles at sites around the globe. The sites used in the evaluation are shown in Fig. **??**. The uncertainties of NOAA flask air measurements for the period of this study are ±0.07 ppm (68% c.i.). From the data, samples reflective of well-mixed background air were selected (based on flag criteria) similar to **?** to minimize the influence on the observation of transport model errors in our analysis.

2.5 The satellite retrievals

GOSAT (Greenhouse Gases Observing Satellite) from Japan Aerospace Exploration Agency (JAXA) was launched in 2009 and observes column XCO_2 with the TANSO-FTS instrument (**?**). These data were used to evaluate the different simulations and to assess ~~the~~ model performance at larger spatial scale.

XCO_2 from the simulation ~~results~~ were calculated using global 4° x 6° x 25 (latitude x longitude x vertical levels) daily average 3-dimensional (3-D) atmospheric CO_2 fields. For each satellite retrieval, the global 3-D daily mean gridded atmospheric CO_2 estimates were horizontally interpolated to the location of the retrievals to create the vertical profile of simulated CO_2. Averaging kernels (AKs) (**?**) were applied to model estimates to ensure reliable comparison with GOSAT retrievals:

$$\hat{C} = c_a + (\boldsymbol{h} \circ \boldsymbol{a})^T (\boldsymbol{x} - \boldsymbol{x}_a), \tag{6}$$

where $\hat{C}$ is XCO_2, scalar c_a is the prior XCO_2 of each retrieval, $\boldsymbol{h}$ is a vertical summation vector, $\boldsymbol{a}$ is an absorber-weighted AK of each retrieval, $\boldsymbol{x}$ is a model profile and $\boldsymbol{x}_a$ is the prior profile of the retrieval. Each retrieval had a prior profile (**?**). The retrievals for different terrestrial TransCom (TC) areas (Fig. **??**) were compared with those calculated from the two model simulations. For comparison with GOSAT XCO_2, the estimates of 3D fields at 6° x 4° resolution were used, but not those from the zoom grids due to technical reasons. Differences in XCO_2 due to model resolution were not significant within the context of this study. In this work GOSAT observations (NIES retrieval V02.21 and V02.31) between July 2009 and the end of 2014 were used.

2.6 Global datasets for evaluating simulated soil carbon and gross primary productivity

For evaluation of the JSBACH model results we additionally used data from two soil carbon databases and the FLUXCOM project (**?**). We used the gross primary production (GPP) produced by FLUXCOM, where eddy covariance flux observations are upscaled using machine learning methods and meteorological and remote sensing data. The FLUXCOM GPP has 0.5 degree spatial resolution and eight-day temporal resolution for 2001-2014. Additionally we used two different soil carbon datasets, SoilGrids (**?**) and one based on Harmonized World Soil Database (HWSD) (**?**). For the soil carbon data we used the preprocessed datasets from **?** providing values for organic soil carbon down to 1 m depth.

3 Results

3.1 Global carbon ~~stocks and~~ fluxes and stocks with the two model formulations

3.1.1 Carbon fluxes

Since the two different model formulations differ only in their soil carbon module formulation, the incoming flux to the ecosystem from photosynthesis is the same in both cases. ~~Analysis of the results was done for 2001-2014~~We analyzed results for 2000-2014, and we show here averaged values for that period.

~~Global simulated GPP of 167 (Table **??**) is highly overestimated when compared to the upscaled data product from FLUXCOM, which is giving a mean value of 126 for this time period (**?**) and having a range of 106-130 for a longer time period. Despite the overestimate of global GPP by the model, the comparison to the FLUXCOM product shows that the seasonal cycles in different latitudinal regions are quite similar, although in the northern boreal region JSBACH reaches maximum GPP values later than the FLUXCOM product (Fig. S4). Vegetation carbon biomass was similar in both model formulations (Table **??**). The small difference is caused by fire fluxes, which have small differences in their magnitude but similar spatial patterns (Fig. S3). The global estimate for total soil carbon by CBA was 4.5-fold larger than by YAS (Table **??**). The global estimate for the litter simulated by YAS pool was larger than that simulated by CBA.~~

~~The net flux shows a slightly larger land sink for YAS than CBA (Table **??**) . Owing to the larger litter pool, fire fluxes are larger in the YAS model formulation by 0.50 . This caused the heterotrophic respiration of YAS to be 0.56 smaller than by CBA, since the model was spun-up to steady state in 1860. This does not completely compensate for the difference and leads to a small discrepancy in net fluxes between the two model formulations.~~

~~Since the soil carbon pools are of very different magnitude, but the annual R_h between the model formulations are similar, the turnover times of the two formulations must differ. The turnover times (τ)of soil carbon pools can be evaluated at both grid scale and from global values. This global value is obtained by dividing the total soil carbon pool (to which both soil and litter carbon stocks are added) by the annual R_h. Calculated from the global values averaged between 2001~~ The main target variable of our analysis is understanding the role of heterotrophic respiration, but to better elucidate how it influences the atmospheric CO_2, we also show net primary production (NPP) and ~~2014, the apparent turnover time for CBA is 51.3 years and for YAS 14.8 years. The anomalies of the turnover times are represented in Fig. S4. These have been calculated from the carbon pools over the whole time period and the mean annual R_h. The models show longer turnover times in northern high latitudes and dry areas. CBA predicts longer turnover times to southern Europe, whereas YAS does not. On the other hand, YAS predicts longer turnover times close to Saharan desert, different to CBA. YAS also consistently predicts longer turnover times to northern latitudes, but CBA does not do this for the northern European region.~~

~~The global distribution of soil carbon is very different between the model formulations (Fig. **??**). Overall the CBA values are higher, with the highest values reaching over 105 , being four times larger than the values predicted by YAS. The CBA model has large values of soil carbon in the mid-latitudes of the Northern Hemisphere. YAS predicts larger values in the temperate region of the Northern Hemisphere, but the highest values of soil carbon are located in arctic regions~~net ecosystem exchange (NEE). NPP is obtained from the gross primary production (GPP) by subtracting autotrophic respiration. NEE is obtained by further subtracting from GPP total ecosystem respiration, autotrophic respiration, direct land cover change, fire, harvest and herbivory fluxes, as shown in Table **??**.

Even though annual total global values of heterotrophic respiration are similar between the different model formulations (Table **??**), their global seasonal cycles are different (Fig. **??**c). The YAS version has a 66% larger variation of R_h during the year than CBA. Both model versions have their minimum value of R_h in February. While CBA has a maximum in August, YAS reaches its maximum value one month earlier, and global R_h also stays high during August. YAS clearly has a steeper increase and decline in its seasonal cycle than CBA. The ~~R_h seasonal cycles show clear differences in their seasonal pattern between~~

~~the two model formulationsin different latitudinal regions~~ higher peak of heterotrophic respiration by the YAS model leads to higher global NEE values during June and July (Fig. **??** e). In the first four months of the year, NEE is higher in the simulations of the CBA model, caused by the higher heterotrophic respiration values at this time (Fig. **??** e). Autotrophic respiration (which, as explained above, like GPP and NPP is the same for both model formulations) and has its highest values in July and August (Fig. ~~**??**). The magnitude of heterotrophic respirationis quite similar in the different latitudinal zones~~S2a). During 2000-2014 both CBA and YAS predict increases in heterotrophic respiration, but only YAS has a significantly increasing trend (p-value $<$ 0.005) (Fig. **??**). CBA has a larger standard deviation in the annual values (0.87 PgC) than YAS (0.73 PgC). The annual NEE time series do not have significant trends and CBA has larger interannual variability (standard deviation of 0.84 PgC vs. 0.79 PgC by YAS).

In addition to the comparison of the global results, we looked how the two soil modules differed for broad latitudinal separated regions. As for the global values total magnitudes of R_h are comparable, while the seasonal cycles show clear differences and similar behaviour is also noticed in different latitudinal regions (Fig. **??**c, d). The YAS model shows~~however~~, however, a larger amplitude in the seasonal variation in all of the regions.

~~The seasonal cycle is quite different between the model formulations in the tropics. At 0° N-30° N, where YAS has seasonal cycle shifted earlier compared to CBA. In this region YAS has 42 % larger seasonal amplitude of R_h than CBA~~The NPP is the same in the different latitudinal regions (Fig. **??**a, b). In the two ~~more~~ most northern regions in the Northern Hemisphere the amplitude in R_h of YAS is approximately twice the amplitude of CBA. In both of these regions YAS has clear maximum values of R_h in July and August, while the seasonal cycles of CBA are more shallow and do not include such clear maximums.

~~In the southern hemisphere in~~ The seasonal cycle of R_h is quite different between the model formulations in the tropics. At 0°N-30°N, YAS has a seasonal cycle shifted earlier compared to CBA. In this region YAS has a 42% larger seasonal amplitude for R_h than CBA. In the Southern Hemisphere regions 0°S -10°S and 10°S -30°S~~the~~, CBA predicts higher values of R_h during the first months of the year after which it stays lower until the end of the year, whereas YAS shows a clear lowering between June and September. In the region 10°S -30 °S YAS has 54% larger amplitude in R_h than CBA. The differences in heterotrophic respiration lead to pronounced differences in the NEE within the tropical region (Fig. **??**e, f).

The variation in R_h seasonal dynamics of these two model formulations can be linked to the differences in their environmental drivers and functions. In Table **??** the correlation between heterotrophic respiration and the environmental drivers of each specific model formulation are shown for the different latitudinal regions. Figures ~~S5-S8~~ S3-S7 in the supplementary material show these same relationships. The R_h from CBA has a strong correlation with ~~the~~ soil moisture α in the tropical region (30°S-30°N) and a high correlation with ~~the~~ soil temperature T_{soil} in the northern high latitudes (30°N-90°N) and lower correlation in southern high latitudes (30°S-60°S). ~~In other regions the r^2~~ For other combinations of regions and drivers the r values are low ~~between R_h from CBA and its environmental drivers and~~ for CBA and in two regions the dependency between α and R_h is negative. ~~On~~ For the YAS model, on the other hand, ~~the~~ R_h ~~predicted by the YAS model~~ shows strong correlation to its environmental drivers (Table **??**). The ~~r^2~~ r values between R_h and precipitation are over ~~0.80~~ 0.90 in all regions except region 30°S-60°S. Between ~~the~~ air temperature and R_h the results are similar, with the only ~~lower r^2 value taking place in tropics~~ small r value in the Southern Hemisphere tropics. The seasonal cycle of R_h predicted by the YAS model does ~~not show~~

~~any positive correlation to~~ correlate positively with the soil moisture variable α in any of these regions (Table **??** and Fig. ~~S9).~~ S7). This is not unexpected as such, since α is not the driver of the YAS model. In the tropical region ~~the~~ soil moisture for CBA and precipitation for YAS are more important drivers compared to soil and air temperatures. ~~In the high latitudes the temperature has~~ At high latitudes temperature has a larger effect on R_h ~~with~~ in the results of both models, even though in the Northern Hemisphere ~~also the precipitation~~ precipitation also has a significant role for YAS.

~~To assess~~We also investigated, whether the ~~higher amplitude of the seasonal cycle~~ seasonal cycle of the heterotrophic respiration is correlated with litter fall. The only significant correlation occurred at 30°N-60°N for both model versions. This was caused because both have similar annual cycles of R_h and litter fall, but the seasonal cycle of R_h precedes litter fall (Fig. S8).

Global simulated GPP of 167 PgCyr^{-1} (Table **??**) is highly overestimated when compared to the up-scaled data product from FLUXCOM, which is giving a mean value of 126 PgCyr^{-1} for this time period (**?**) and having a range of 106-130 PgCyr^{-1} for a longer time period. Despite the overestimate of global GPP by the model, the comparison to the FLUXCOM product shows that the seasonal cycles in different latitudinal regions are quite similar, although in the northern boreal region JSBACH reaches maximum GPP values later than the FLUXCOM product (Fig. S9).

The annual net CO_2 flux shows a slightly larger land sink for YAS than CBA (Table **??**). Owing to the larger litter pool, fire fluxes are larger in the YAS model formulation by 0.50 PgCyr^{-1}, however they have similar spatial patterns (Fig. S10). This caused the heterotrophic respiration of YAS to be 0.56 PgCyr^{-1} smaller than by CBA, since the model was spun-up to steady state in 1860 and thus leads to a small discrepancy in net CO_2 fluxes between the two model formulations.

3.1.2 Carbon stocks

The soil carbon stocks predicted be the two models differed in magnitude and also their latitudinal distributions differed. The global estimate for total soil carbon by CBA was 4.5-fold larger than by YAS (Table **??**). The global estimate for litter simulated by the YAS model was larger than that simulated by CBA. Vegetation carbon biomass was similar in both model formulations (Table **??**).

The global distribution of soil carbon is very different between the model formulations (Fig. S11c, d, Fig. S12). The CBA model has large values of soil carbon in the mid-latitudes of the Northern Hemisphere. YAS predicts larger values in the temperate region of the Northern Hemisphere, but the highest values of soil carbon are located in arctic regions. The data based estimates from SoilGrids and HWSD also predict the highest values at high northern latitudes (Fig. S11a, b and Fig. S12). The CBA model predicts higher values and differing latitudinal pattern south of 60°N compared to the data based values (Fig. S12). The YAS model shows very similar behaviour to the HWSD latitudinal pattern and magnitude south of 60°N. The r^2 and the root mean square errors are generally better for the YAS model than the CBA model when comparing the values along the latitudinal gradient against the data-based products (Table S2).

The turnover times of the two formulations must differ, since the soil carbon pools are of very different magnitude, but the annual R_h between the model formulations are similar. The turnover times (τ) of soil carbon pools can be evaluated at both grid scale and from global values. This global value is obtained by dividing the total soil carbon pool (to which both soil and litter carbon stocks are added) by the annual R_h. Calculated from the global values averaged for 15 years, the apparent turnover time for CBA is 51.3 years and for YAS 14.8 years. The anomalies of the turnover times are represented in Fig. ??. These have been calculated from the carbon pools over the whole time period and the mean annual R_h. The models show longer turnover times in northern high latitudes and dry areas. CBA shows a larger spread of turnover times within different temperature regimes than YAS (Fig. ??). The turnover times of CBA are generally longer and show a large spread across different temperatures. The YAS model shows a large spread of turnover rates at warmer temperatures but below 0°C the range is narrower (Fig. ?? b). Both models predict the fastest turnover rates in moist and warm conditions.

3.1.3 Box model

To assess whether the larger seasonal cycle amplitude in R_h by YAS is caused by the larger litter pool or the environmental response functions, a simple box model calculation was performed (~~a detailed description~~ detailed description is given in Appendix).

~~When the~~ When global respiration was calculated with the turnover times and soil carbon pools of the YAS model, but using the environmental responses and drivers of the CBA model, the annual magnitude decreased by 29% compared to the original YAS model (Table ??). However, the yearly maximum value did not change ~~by~~ much. When the opposite was done, and the turnover time and soil carbon pools of CBA were used with the environmental responses and inputs of the YAS model, the magnitude of ~~the~~ global heterotrophic respiration ~~was increased~~ increased by approximately 1.4-fold (Fig. ??). The increase in the amplitude was 83% (Table ??). Therefore, this simple analysis suggests that the environmental variables and their response functions cause the larger global amplitude of R_h in the YAS model formulation. To further disentangle ~~,~~ whether this change was caused by the different environmental drivers or their functional dependencies, we made additional tests.

The amplitudes of the seasonal cycle of R_h (difference between the maximum and minimum values) are shown in Table ??. For the YAS model, there happens a strong decrease in the amplitude when both driver variables and the response functions are changed. When only driver variables are changed, ~~there occurs~~ only a slight decrease occurs. When the response functions are changed, the decrease in the amplitude is more pronounced with 21%. The amplitude predicted by the CBA model increases, when the driving variables and response functions are changed (Table ??). This increase ~~is~~ occurs when either driving variables or response functions are changed individually. However, with the change of the response functions the change in ~~the~~ amplitude is larger ~~,~~ (74%~~. Therefore~~). In summary, the response functions have a more pronounced role in the changes than ~~only~~ the driving variables alone, and this was true for both ~~of the~~ models.

3.2 Evaluation against surface observations

Seasonal cycle amplitudes ~~(SCAs)~~ of atmospheric CO_2 are successfully simulated by the modeling framework across different latitudes (Fig. ~~5a~~ ??a). The r^2 values of the observed seasonal cycle and the model estimates are high across latitudes, despite some lower values in mid-latitudes of the Northern Hemisphere (Fig. ??b). Averaged over all ~~the~~ latitudes the r^2 value, calculated as linear correlation of simulated and observed averaged annual cycles, was 0.93 for ~~the~~ CTE, 0.90 for CBA and 0.87 for YAS.

The capability of the model formulations to simulate the amplitude of the seasonal cycle differs within latitudinal regions (Fig. **??**). The CBA model is able to capture the timing of the seasonal cycle in northern latitudes, but has a tendency to overestimate the ~~SCA by~~ seasonal cycle amplitude by about 30% north of 45°N. In this region the underestimation of ~~SCA~~ seasonal cycle amplitude by CTE is approximately 5% and by YAS 14%. In the region 0°N-45°N ~~the~~ YAS underestimates the ~~SCA in averageapproximately by~~ seasonal cycle amplitude, on average, by approximately 32%, whereas CTE underestimates it by 4% and CBA overestimates it ~~only~~ by 1%. The agreement between estimated atmospheric CO_2 and observations was worse in YAS than in CBA when considering the r^2 value and the seasonal cycle. Overall, the magnitude of the ~~SCA~~ seasonal cycle amplitude predicted by YAS had less bias north from 45°N compared to CBA, but large underestimation in latitudes 0°N-45°N, where CBA was very successful in ~~attaining the right SCA.~~ simulating the right seasonal cycle amplitude.

~~This behavior is further illustrated from comparisons of the detrended seasonal cycle at four stations~~ Four surface observation sites in the Northern Hemisphere illustrate similar behaviour of the seasonal cycle and its amplitudes as described above (Fig. **??** and Table ~~S2~~S3). To confirm the general quality of the TM5 model used for both YAS and CBA we plotted its biospheric posterior fluxes from CarbonTracker Europe 2016 (CTE); indeed, deviations ~~of CTE to~~ between CTE and observations are much smaller than from the JSBACH model at all sites. At the high-latitude sites, Alert and Pallas (Fig. **??**a, ~~b~~e), CBA overestimates the seasonal cycle amplitude, while YAS shows some phase-shift of the cycle. The observed ~~SCAs~~ seasonal cycle amplitudes are smaller at the two more southern sites, Niwot Ridge and Mauna Loa. For those sites, CBA is generally successful in capturing their magnitude (Table ~~S2~~S3), whereas YAS underestimates them strongly. YAS is also having difficulty capturing the seasonal pattern at Niwot Ridge. This was happening generally in the temperate region, as is also seen in the lower r^2 values of the YAS model at the different sites (Fig. **??**).

In addition to the seasonal cycle the temporal development of the seasonal cycle amplitude for the four sites is displayed in Fig. **??**b, d, f, h. We show this development for relative values of the seasonal cycle amplitudes to make the temporal development visible, since the values between the two different model formulations are so different. The correlation coefficients between observed and the different modelled time series are shown in Table S4. CTE better captured the interannual variation of the seasonal cycle amplitude than the CBA and YAS models, which perform comparably. The YAS model shows stronger interannual variation at Niwot Ridge (Fig. **??**d) and this is caused by the small magnitude of the seasonal cycle amplitude by YAS at this site.

When comparing the overall bias in atmospheric CO_2 at these four sites between the observations and the model simulations, CBA overestimated CO_2 by 3.65 ppm and YAS by 2.27 ppm, when averaged over all the measurements within the study period. A closer look at the bias at Mauna Loa (Fig. ~~S10) revealed a trend in this bias between year 2000 and 2014. The~~ S13) revealed biases in the 2000-2014 trends for CBA and YAS, whereas CTE shows no bias in trend. The CBA ~~is overestimating~~ overestimates CO_2 by 1.76 ppm in the beginning and by 3.74 ppm in 2014. The ~~YAS has lower overestimation, being~~ overestimates by YAS are smaller, 1.12 ppm in 2000 and 3.14 in 2014.

The results at surface sites show that CBA largely overestimated ~~SCA~~ seasonal cycle amplitude at high northern latitudes, whereas YAS almost consistently underestimated the ~~SCA~~ seasonal cycle amplitude in the Northern Hemisphere. CBA captured the seasonal cycle patterns better than YAS across different latitudes. Overall, the YAS model showed biases in the atmospheric CO_2 cycle at temperate latitudes in the Northern Hemisphere, whereas the CBA model had biases in the high latitudes in the Northern Hemisphere.

3.3 Column XCO_2 comparisons for TransCom regions

This evaluation of the two soil modules against satellite column XCO_2 was carried out for the different TransCom (TC) regions (Fig. **??**). The comparison was based on seasonal cycle amplitudes and r^2 values similar to the surface site evaluation. Not all the TC regions show a clear seasonal cycle, such as regions in South America (TC regions 3 and 4), northern part of Africa (TC=5) and Australia (TC=10). For completeness we show the analysis also for these regions in Table ~~S3~~S5. For regions with clear seasonal ~~cycle~~ cycles we used the ccgcrv curve fitting procedure available from NOAA (https://www.esrl.noaa.gov/gmd/ccgg/mbl/crvfit/crvfit.html, (**?**)), but for regions with missing data or no clear seasonal cycle, we averaged over all years of data.

To further illustrate the results from this comparison, we show data for two regions having a clear seasonal cycle~~and reflecting the same behaviour that was noticed at the four surface observation sites that were looked deeper into in the previous section~~. In TC region 2, the southern part of North America, CBA is more successful in capturing the observed ~~SCA~~ seasonal cycle amplitude than YAS (Fig. **??**a), even though CBA reaches the minimum XCO_2 later than observations. YAS underestimates ~~SCA~~ the seasonal cycle amplitude by 56% and has a different seasonal pattern than observations, so the minimum is reached earlier than in the observations and also the shape during the summer period ~~is different to~~ differs from the observations. In Europe, TC region 11, ~~Europe,~~ both models capture the ~~SCA~~ seasonal cycle amplitude (Fig. **??**~~, Table S3~~c, Table S5) and the seasonal cycle in the first part of the year. The increase of CO_2 is not as well captured by the simulations. The time series of seasonal cycle amplitudes predicted by the CBA and YAS models(**??**c, d) do not correlate significantly with the observations.

Overall, observed and simulated XCO_2 differ from each other in ways similar to the surface site observations. Estimates of ~~SCA~~ seasonal cycle amplitude by YAS are too small in mid-latitudes (Fig. **??**a) and in TCs 2, 5 and 8 compared to the observations, and CBA is better ~~in~~ at capturing the observed annual cycles. At TC=1 (the northern part of North America), CBA overestimates the ~~SCA~~seasonal cycle amplitude, while YAS better captures it. However, the seasonal cycle pattern is better captured with CBA (Table ~~S2~~S5) than with YAS. Generally YAS had smaller ~~SCAs~~ seasonal cycle amplitudes than the observations and CBA was more consistent with the observations in most ~~TC regions (Table S3~~TC-regions (Table S4). CBA is also better than YAS in capturing the seasonal pattern of XCO_2 in ~~the~~ all TC regions (Table ~~S3~~S5).

There ~~occurs bias between the space-born observations~~ is bias in absolute XCO_2 between the GOSAT retrievals and the model simulations. When averaged over the time period used and the TC regions, CBA overestimates the GOSAT observations by 3.37 ppm and YAS by 2.33 ppm. These values were in line with ~~the bias~~ bias in absolute CO_2 estimates at the four surface sites.

4 Discussion

In this work our aim was to use atmospheric observations to ~~benchmark~~ assess whether soil carbon models of a land surface model can be evaluated with this kind of framework. Our main finding was that the two models predicted ~~differently the annual cycle of the~~ different annual cycles of global R_h, with the YAS model having a larger amplitude. This in turn ~~lead to pronounced~~ leads to clear differences in the model predictions of seasonal cycles of the atmospheric $\mathrm{CO_2}$ ~~mole fractions. To be able to~~ abundance. To attribute the differences between the two models to a specific cause, we need to compare their results from their different aspects and to also judge whether our model simulations are reasonable in the light of previous research.

~~Similarly to the earlier studies (**??**), in our results the YAS model was more successful than CBA in estimating the observed global soil carbon stocks, which is approximately 1500 (**?**) including large uncertainties. The distribution of soil carbon stocks was also more realistic in YAS than in CBA. The large soil carbon stocks in the mid-latitudes predicted by CBA (Fig. **??** a) are unrealistic compared to current estimates of global soil carbon distribution (**?**). The large carbon stocks at high latitudes predicted by the YAS model (Fig. **??**) are more in line with the observations. However, the version of JSBACH used does not include peatlands and is modelling only mineral soils, therefore the large carbon reservoirs of peatlands are not captured by the model, as they are now described as mineral soils~~

4.1 Evaluation of carbon fluxes

Annual heterotrophic respiration was 66.1 $\mathrm{PgCyr^{-1}}$ for CBA and 65.5 $\mathrm{PgCyr^{-1}}$ for YAS (Table **??**), which falls in the range of estimates from Earth System Models (41.3-71.6 $\mathrm{PgCyr^{-1}}$) and is close to the observation based estimates of 60 $\mathrm{PgCyr^{-1}}$ (**?**). Part of the difference is caused by the fire fluxes. The YAS model ~~is widely used in different applications at smaller scale and its performance to estimate soil carbon stocks has been found to be good (**?**). Comparability between the model-calculated and the observed carbon stocks is relevant for any analyses of the carbon fluxes because in the both models the fluxes are proportional to the stocks (flux = decomposition rate * stock).~~

~~The global turnover rate of soil carbon by CBA was somewhat larger than in an earlier study, where it was estimated to be 40.8 years (**?**). This value was in the higher end of the CMIP5 models. The global turnover rate value from YAS , which was 14.8 years, is more in the range of the other CMIP5 models (**?**). The spatial distribution of the turnover rate anomalies show the the differences caused by the environmental drivers and their dependencies at annual timescales. When comparing these overall turnover rates of the total soil carbon, it is important to keep in mind that both models consisted of carbon pools that had widely varying turnover rates. For example, despite of the higher overall turnover rate, the turnover rate of the most recalcitrant carbon pool of YAS was an order of magnitude lower than that of CBA~~has a larger litter pool that behaves as fuel for fires. Therefore, to have the system at steady state, global heterotrophic respiration by YAS must be less. Moreover, the simulation time of 140 years before the beginning of the analysis might cause some divergence between the model runs.

Moving to ~~the~~ monthly time scales, we can see that the global seasonal R_h cycle had a larger amplitude with YAS than with CBA (Fig. **??**) and a simple box model calculation found ~~the~~ that environmental drivers and their response functions ~~to be the causeof this instead~~ are the cause, not the large litter pool in the YAS model. It is anyhow challenging to further disentangle whether this larger amplitude is mainly caused by the differing environmental drivers of the soil carbon models or if ~~their functional dependencies~~ the functional dependencies of those drivers would play a bigger role. The analysis by the box model suggested a stronger role of the response functions compared to the driving variables at monthly timescales, but strong conclusions cannot be drawn from such a simple analysis. Also other studies have showed that the response functions themselves lead to pronounced differences between soil carbon models (**?**).

~~The annual heterotrophic respiration was 66.1 for CBA and 65.5 for YAS (Table **??**), which falls in the range of estimates from the Earth System Models (41.3-71.6) and is close the observation based estimates of 60 (**?**). The similar values of R_h by YAS and CBA are caused by the way the models are run into steady-state in the beginning of the simulation in 1860. Part of this difference is caused by the fire fluxes. The YAS model has a larger litter pool that behaves as fuel for fires. Therefore, to have the system at steady state, global heterotrophic respiration by YAS must be less. Also, the simulation time of 140 years before the beginning of the analysis might also cause some divergence between the model runs.~~

~~When R_h~~ When heterotrophic respiration is compared by latitudinal zones, differences between the model formulations are visible in the variability and timing of the seasonal cycles in many regions (Fig. **??**). R_h ~~showed strong correlations to~~ correlates strongly with the environmental drivers of the models in different latitudinal zones (Table **??**). Both models are largely influenced by their moisture dependency in the tropical region (Table **??**). CBA is driven by soil moisture with a linear dependence and YAS is driven by precipitation. ~~While at annual timescales these two variables (air vs. soil temperature and precipitation vs. soil moisture) are similar, since precipitation is affecting soil moisture and on longer timescales air temperature determines soil temperature in the top soil layers, the seasonal cycles of the variables are different.~~ At annual timescales, at which the YAS model ~~has been~~ was originally developed, ~~the dynamics of these variables are not likely to be as differentas at these shorter timescales~~precipitation and soil moisture behave similarly. However, the seasonal cycles of the two variables are different. Precipitation begins earlier in the season in the tropical region, and it causes YAS to reach yearly maximum ~~R_h~~ heterotrophic respiration earlier than CBA, which is driven by soil moisture in this region. ~~Also~~ Similarly, air and soil temperatures are more similar on the long term as for short periods. Particularly in the temperate region, where the temperature has a larger role, the air temperature has larger variability than soil temperature and this leads to different kind of seasonal pattern of the R_h predictions by the two different soil models.

The observations show that litterfall has strong influences on heterotrophic respiration (**?**), but this process is not included in the models, so at seasonal timescales in the different latitudinal zones no clear influence of litterfall driving the heterotrophic respiration was seen. However, changes in the chemical composition of litterfall is considered to be a potential reason for changes in the amplitude of atmospheric CO_2 (**?**) and this is something we could study with the YAS model.

Different moisture dependencies of R_h have earlier been found to be important (**?**). At the global level **?** recommended using parabolic soil moisture functions in preference to functions based on mean annual precipitation. Their study considered soil respiration, i.e., autotrophic respiration by roots was also included. **?** evaluated the YAS model against R_h observations at two coniferous sites in southern Finland and found problems in capturing the seasonality in the observations and the variability in the summertime fluxes. One reason ~~for this they mention is the~~ they mention for this is response of the simulated R_h to soil moisture conditions, since R_h is not attenuated in very moist conditions and they found a need to improve the moisture dependency of the YAS model. This is in line with our findings, that a model that has been parameterized at annual time scales requires further development before it can be reliably applied at shorter timescales. Precipitation ~~has been~~ was originally used in the YAS model as a proxy for soil moisture, since enough accurate soil moisture observations for model development ~~haven't been~~ were not available. Clearly, this idea needs reconsideration as our results show that at zonal spatial scales and monthly temporal scale~~the~~ , R_h from YAS is not at all correlated to ~~soil moisture variable α~~the soil moisture.

~~The global~~ Global GPP, being 165 $\mathrm{PgCyr^{-1}}$ in this study, was overestimated, compared to the FLUXCOM estimate. Different FLUXCOM products give estimates between 106 and 130 $\mathrm{PgCyr^{-1}}$ for ~~period~~ 2008-2010 (**?**). There ~~has been also~~ have also been other estimates for ~~the~~ global GPP. The Carbon Cycle Data Assimilation system ~~provides a value~~ estimated of 146 (± 19) ~~(**?**) and the~~ $\mathrm{PgCyr^{-1}}$ (for 1980-1999) (**?**) and estimates based on isotope observations ~~have given estimates of~~ are 150 to 175 $\mathrm{PgCyr^{-1}}$ ~~(**?**). The~~ (for 1980-2009) (**?**). That GPP of JSBACH is ~~relatively high , but it~~ biased high compared to observations is likely of secondary importance to our study comparing two model formulations, because GPP was the same for both ~~of the model formulations and only contributed to~~ formulations and the GPP bias did not lead to strong biases in the seasonal cycle predictions in different latitudinal zones were (Fig. S9). However, to assess the absolute skill of each model formulation in terms of net ecosystem exchange, GPP biases need to be reduced. Furthermore, the high GPP values predicted in the current run would likely be lower, if the nutrient cycles of nitrogen and phosphorus were included in the used version of JSBACH. Beside using a JSBACH version with nutrient cycles, further development work in the phenological cycle could improve the estimated GPP. The difference of the modelled GPP to the FLUXCOM product (Fig. S9) suggests that the maximum leaf area index might be overestimated in the tropics. Also, the timing of the phenological cycle north of 60°N might benefit from re-parametrization.

4.2 Evaluation of carbon stocks and turnover times

The two soil models predicted different global soil carbon stocks (Table **??**) with different latitudinal distributions (Fig. S12). Similar to earlier studies (**??**), in our results the YAS model was more successful than CBA in estimating global soil carbon stocks similar to estimates from observations, approximately 1500 PgC including large uncertainties (from 504 to 3000 PgC) (**?**), as can be seen in the ~~amount of litter fall. Therefore we do not expect the variation of its magnitude to have substantial influence to our results~~different estimates from HSWD (1578 PgC) and SoilGrids (2870 PgC) (see also **?**). The YAS model is widely used in different applications at smaller scale and its performance to estimate soil carbon stocks has been found to be good (**?**). Comparability between the model-calculated and the observed carbon stocks is relevant for any analyses of carbon fluxes because in both models investigated here the fluxes are proportional to the stocks (flux = decomposition rate * stock). Modelled global vegetation carbon was within the observation-based estimate of 442 ± 146 PgC by **?**.

The distribution of soil carbon stocks was also more realistic in YAS than in CBA (Fig. S12, Table S2). The large soil carbon stocks in the mid-latitudes predicted by CBA (Figs. S11c, ~~bearing also in mind that the seasonal cycle in different latitudinal zones was~~ S12) are unrealistic compared to current estimates of the global soil carbon distribution (Fig. S12). The large carbon stocks at high latitudes predicted by the YAS model (Figs. S11d, S12) are more in line with the observations, but miss the high values observed from peatlands and permafrost in high latitude regions. The version of JSBACH used does not include peatlands and is modelling only mineral soils. Therefore, the large carbon reservoirs of peatlands are not captured by the ~~JSBACH model~~ model. This JSBACH version also didn’t have permafrost described. If permafrost would be modelled, the seasonal cycle of heterotrophic respiration at high latitudes would likely be dampened, as the depth of the active layer determines the amount of soil capable of respiring. The YAS model has been used in a JSBACH version containing permafrost in a study concentrating on the Russian Far East (**?**). Both, CBA and YAS, were originally developed for mineral soils and for applications with organic soil, so model development and testing at smaller than global scale could be useful.

The global turnover time of soil carbon by CBA was somewhat larger than in an earlier study, where it was estimated to be 40.8 years (**?**). This value was in the higher end of the CMIP5 models. The global turnover time from YAS, which was 14.8 years, is more in the range of the other CMIP5 models (**?**). The spatial distribution of the turnover time anomalies show differences caused by the environmental drivers and their dependencies at annual timescales. When comparing these overall turnover times of total soil carbon, it is important to keep in mind that both models consisted of carbon pools that had widely varying turnover times. For example, despite the higher overall turnover time, the turnover time of the most recalcitrant carbon pool of YAS was an order of magnitude smaller than that of CBA.

The environmental responses of the turnover rate have quite different forms for the two soil carbon models (Fig. ~~S2).~~ **??**. The CBA model shows a wide distribution of turnover rates across the whole temperature range, whereas the YAS model shows a larger spread in the tropical temperature range. This large spread in warm conditions is also observed (**?**) and is caused by the saturating temperature function of the YAS model, as shown in Fig. S1c. The large spread in turnover times as predicted by the CBA model might be caused by the fact that CBA is driven by soil temperature in one soil layer. The environmental responses of the turnover rates at annual time scales behave similarly as at monthly time scales, so that moisture is a more important driver in warm regions and temperature in cold regions, as was seen in Table **??**.

4.3 Evaluation using atmospheric CO_2

The differences ~~by~~ between the two models in the seasonal cycle of atmospheric CO_2 were strong. CBA better reproduced the seasonal cycle amplitudes capturing the shape of the seasonal cycle both for surface sites and comparisons in the TC regions, even though its soil carbon distribution had ~~lower~~ worse performance compared to YAS. CBA exaggerated the seasonal cycle amplitudes at high northern latitudes, as has been found earlier (**?**).

~~The biases between from satellite retrievals and~~ It is important to keep in mind that this study was done within a land surface model and modelled GPP was biased. The simulated GPP had a larger magnitude and some bias in its seasonal cycle, and therefore its evaluation against atmospheric CO_2 observations is influenced by it. Even though the atmospheric observations provide a valuable and informative comparison for the model results~~originating from the JSBACH simulations are relatively large and this is likely~~, their use as a benchmark metric needs careful consideration.

The differences in absolute CO_2 and XCO_2 levels against the surface observations and the satellite retrievals, respectively, with modelled CO_2 are caused by the ~~use of a posteriori ocean fluxes from the~~ modelling system, but this bias does not influence the analysis performed. We obtained the land surface fluxes (GPP, respiration, fire, herbivory fluxes, land-use change emissions) from JSBACH and together with the rest of fluxes from CarbonTracker Europe2016 (CTE2016~~. The~~), we used

TM5 to obtain atmospheric $\mathrm{CO_2}$ values. Fossil fuel emissions have not been optimized in CTE2016. Therefore we obtained ocean fluxes that had been optimized with the land carbon cycle of CTE2016, that differ from the JSBACH estimate. The land carbon cycle of CTE2016 is modelled by the SiBCASA-GFED4 model (**?**) and fire emission that were estimated from satellite observed burned area (**?**). The net global a posteriori land sink ~~(including biomass burning emissions) of CTE~~ of CTE2016 is approximately -2.0 (± 1.1) $\mathrm{PgCyr^{-1}}$ ~~in time period~~ for 2001-2014. ~~The global land sink of JSBACH~~ On the other hand, the JSBACH estimate for the net land sink is approximately -1.7 $\mathrm{PgCyr^{-1}}$ (Table **??**) and is therefore ~~lower~~ smaller than the land sink by CTE2016. ~~Since the ocean fluxes are a posteriori fluxes from~~ The fire flux of JSBACH is modelled, whereas the estimate in CTE2016 ~~simulation, they cause a bias to the simulated atmospheric~~ is based on data. As shown in Fig. S13 for Mauna Loa, the bias in the $\mathrm{CO_2}$ ~~mole fraction when used together with the land fluxes from the JSBACH simulation. The~~ develops during the study period and the plot shows consistency so that YAS, which predicts a net land sink ~~of YAS version is slightly~~ closer to CTE2016 ~~value, and this leads to lower bias also at Mauna Loa (Fig. S10).~~

~~Additionally, the spaceborn observations also contain bias. GOSAT retrievals were evaluated against ground-based Total Carbon Column Observing Network (TCCON) , and is biased low by approximately 1.48 ppm (**?**).~~ than CBA, has smaller bias at the end of the time period. We concentrated the analysis on the averaged seasonal cycles, that are not influenced by this linear increase. We show also some time series for the seasonal cycle amplitudes, but these have been calculated from detrended time series.

~~In this study we concentrated on analysing the SCAs and the pattern of~~ The space-borne observations give a similar message as the surface observations in TransCom regions, which showed clear seasonal cycle. Niwot Ridge is located in TransCom region 2 (southern part of North America) and also there YAS showed too low amplitude and CBA performed better, similarly as seen in the Fig. **??**. The Pallas site is located in TransCom region 11 (Europe) and at Pallas the seasonal cycle ~~and emphasized the differences between the two different soil carbon models . Therefore we do not consider this bias to play a big role in these analysis. Also the transport model contains biases (**?**), but since only one transport model was used~~ was more pronounced than in Europe as whole, but similarly for the surface observations at Pallas and TransCom region 11, the models both perform acceptably. Using large TransCom regions helped to interpret the signal despite the larger variability than in the surface observations (comparing grey shaded regions in Figs. 7 and 8) and it has been recommended to use the information content of the satellites on continental scales (**?**).

The transport model itself also brings uncertainty to the result. Modelling of atmospheric transport is a challenging task as open scientific questions in the field remain (**?**) and the models contain biases (**?**). The errors in atmospheric transport models cause a substantial difference in the inverse $\mathrm{CO_2}$ model flux estimates (**?**). However, in this study ~~, it~~ we only used one atmospheric transport model. It is expected that the biases, as only one transport model was used, are similar between the two soil model runs and are not the cause for the large differences seen in the two ~~different~~ simulations.

5 Conclusions

We demonstrated how atmospheric CO_2 observations can be used to ~~benchmark~~ evaluate two soil carbon models ~~and that it is important to benchmark models across several different variables~~within the same land surface model and the different viewpoints offered by several variables considered. We used two different soil carbon models within one land surface model and used a three-dimensional transport model to obtain atmospheric CO_2, while obtaining the anthropogenic and ocean fluxes from CarbonTracker Europe framework. We evaluated the carbon stocks of the soil models and compared seasonal cycles calculated with soil carbon fluxes from the soil models to atmospheric CO_2 results from both surface and space-born observations. This work highlighted how the changes in the heterotrophic respiration transfer to the net ecosystem exchange estimates and further to the atmospheric CO_2 signal. We also discussed the importance of the model drivers and their functional dependencies, which differed for the two soil carbon models we studied. When considering both surface- and space-based observations, it is not straightforward to say which of the two soil carbon models performed better. Also, the evaluation was done within a land surface model that is biased in its GPP predictions when compared to an upscaled GPP product and this hampers the use of atmospheric CO_2 as a numeric benchmark.

The comparison of the two soil carbon models revealed large differences in their predictions. The YAS model better captured the magnitude and spatial distribution of soil carbon stocks globally~~and resulted in similar global turnover rate compared to other Earth System Models, in comparison to the much higher turnover rate by the CBA model.~~

~~The YAS model showed biases in the~~. However, it was biased in its atmospheric CO_2 cycle at temperate latitudes in the Northern Hemisphere. The CBA model, on the other hand, showed better performance in capturing the seasonal cycle pattern of ~~the~~ atmospheric CO_2~~mole fraction~~, but it ~~had biases in the~~ is biased at high latitudes in the Northern Hemisphere. ~~When considering both land based and atmospheric based observations, it is not straightforward to say which model performed better. However, the~~ R_h from the YAS model showed misalignment with soil water content in ~~the~~ tropical regions, as they were negatively correlated with each other. This suggests that use of precipitation as a proxy for soil moisture might not be sensible ~~in~~ at sub-annual time scales ~~.~~

~~In addition to the surface observations of , also space-born observations were used. They were providing a larger-scale confirmation for the results obtained from the surface observations and thus worked as a complimentary information source~~and calls for improvement in the parameterization of the YAS model.

Soil carbon models have several development needs (**??**) that are now partly being answered with ~~the~~ next generation models including more mechanistic representation of several below ground processes (**??**). The development of moisture dependency from simple ~~empiric~~ empirical relationships is moving towards mechanistic approaches, which may yield more reliable results in the long term (**?**). Our results confirm that the moisture dependency of heterotrophic respiration plays on important role in the whole global carbon cycle.

In this study we used space-born XCO_2 observations in addition to the surface observations of CO_2. They were providing a larger-scale confirmation for the results obtained from the surface observations and thus provided complimentary information.

The number of satellite observations of column XCO$_2$ are increasing at a fast pace, e.g., OCO-2 observations started in 2014, and they possess high potential for carbon cycle studies (**?**).

Code and data availability. The site level data from Global Atmospheric Watch -network is available via Obspack (2016) (https://doi.org/10.15138/G3059Z). The EDGAR4.2 emission database is available at http://edgar.jrc.ec.europa.eu. The GOSAT data are from GOSAT Data Archive Service (GDAS) (https://data2.gosat.nies.go.jp/index_en.html). The CRUNCEP data is available from Viovy (2010) (https://vesg.ipsl.upmc.fr/thredds/catalog/store/p529viov/cruncep/V7_1901_2015/catalog.html). The JSBACH model can be obtained from the Max Planck Institute for Meteorology, and it is available for the scientific community under the MPI-M Sofware License Agreement (http://www.mpimet.mpg.de/en/science/models/license/, last access: 16 September 2019). The CarbonTracker Europe code is continuously updated and available through a GIT repository at Wageningen University and Research: https://git.wur.nl/ctdas. For further details, see also: www.carbontracker.eu. The transport model TM5 is available via https://svn.knmi.nl/svn/TM5. For the curve fitting for the atmospheric CO$_2$ data we used scripts available from ERSL NOAA at https://www.esrl.noaa.gov/gmd/ccgg/mbl/crvfit/crvfit.html.

Appendix A: ~~Box~~ Description of the box model

A simple box model calculation was performed to evaluate the importance of the dependencies ~~on~~ of environmental drivers and the soil carbon pool sizes on the larger global seasonal cycle amplitude in R_h as predicted by YAS. In this box model, we assume that ~~heterotophic~~ heterotrophic respiration R_h is a product of environmental dependencies and the turnover time as

$$R_{h,YAS} = b * f_{YAS,T_{air}}(T_{air}) * f_{YAS,P_a}(P_a) * \frac{C_{soil,YAS}}{\tau_{YAS}}, \text{where} \qquad b = \frac{\Sigma f_{CBA,T_{soil}}(T_{soil}) f_{CBA,\alpha}(\alpha)}{\Sigma f_{YAS,T_{air}}(T_{air}) f_{YAS,P_a}(P_a)}, \tag{A1}$$

where $R_{h,YAS}$ is the heterotophic respiration of model YAS, b is a scalar that takes into account the different magnitudes of the response functions, T_{air} is ~~the~~ air temperature, P_a is ~~the~~ annual precipitation, $C_{soil,YAS}$ are the total soil carbon pools and τ_{YAS} is the turnover time of the total soil carbon pools. T_{soil} is ~~the~~ soil temperature and α is the relative soil moisture. This formulation in **??** refers to the YAS model. The response functions are as shown in Section **??**. For the CBA model the formulation is as

$$R_{h,CBA} = \frac{1}{b} * f_{CBA,T_{soil}}(T_{soil}) * f_{CBA,\alpha}(\alpha) * \frac{C_{soil,CBA}}{\tau_{CBA}}. \tag{A2}$$

These response ~~have been~~ were introduced in Section **??**.

The equations ~~were used for monthly data averaged over the years 2001-2014 of~~ used monthly heterotrophic respiration, environmental drivers and soil carbon stocks averaged over 2001-2014 to estimate the turnover times for each grid point for YAS using eq. **??** and for CBA using eq. **??**. Using these turnover times, we calculated ~~the~~ global R_h with the turnover times and soil carbon pools of each model by making different tests. First, we used the environmental responses and drivers of the other model (lines B in Table A1). Additionally we changed the driving variables, but kept the original response functions (lines C in Table A1). Then we changed only the response functions of the original model while keeping the original driving variables (lines D in Table A1).

Since the driving variables of soil moisture and annual precipitation differed in ~~magnitudes~~ magnitudea by approximately four-fold, ~~the~~ soil moisture was multiplied by four when using the function for annual precipitation (f_{YAS,P_a}) and when annual precipitation was used in the function for soil moisture ($f_{CBA,\alpha}$) it was divided by four. The annual cycles of R_h are shown in Fig. **??** and the amplitudes in Table **??**.

Author contributions. TT designed the experiment with the help of SZ. JESM performed the JSBACH model simulations. AT did the CarbonTracker Europe (CTE2016) runs with the JSBACH biospheric fluxes, with the $\mathrm{CO_2}$ fields provided by ITK. ITK provided the CarbonTracker Europe (CTE2016) results used for comparison at the surface stations. TT performed the analysis with help from SZ, AT and TM. TT wrote the first version of the draft and all the authors contributed to the manuscript.

Competing interests. Dr. Sönke Zaehle is an associate editor for Biogeosciences.

Disclaimer. TEXT

Acknowledgements. TT was funded by Academy of Finland (grant no. 266803). TT and SZ were funded by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (QUINCY; grant no. 647204). SZ was furthermore supported by the European Union's Horizon 2020 Project funded under the programme SC5-01-2014 (CRESCENDO, grant No. 641816). ITL received funding from Netherlands Organisation for Scientific Research (NWO) under contract no. 016.Veni.171.095. JEMSN and JP were supported by the German Research Foundation's Emmy Noether Program (PO1751/1-1). JSBACH simulations were conducted at the German Climate Computing Center (DKRZ; allocation bm0891). We thank Dr. Janne Hakkarainen for helping in analysing the GOSAT data and averaging kernel calculation. We thank Dr. Martin Jung for access to the FLUXCOM results and the FLUXCOM initiative. We are grateful for Naixin Fan for sharing the preprocessed SoilGrids and WHSD data with us. We thank Prof. Dr. Wouter Peters for constructive comments on an earlier version of this manuscript. We thank two anonymous reviewers whose constructive comments improved this manuscript.

Table 1. Global C storage in the two different model formulations averaged over 2001-2014. For the YAS model the eight above ground pools are summed to obtain the litter pool, while the remaining 10 pools (below ground and humus) represent the soil pool.

C pool (Pg C)	CBA	YAS
Litter C	171	263
Soil C	3217	703
Vegetation C	432	432

Table 2. Global terrestrial C fluxes from the two different model formulations averaged over 2001-2014.

Row	Flux ($\mathrm{PgC year^{-1}}$)	CBA	YAS
A	Net CO_2 flux (A = -B + E + G + H + I + J)	-1.68	-1.75
B	GPP	167	same
C	Heterotrophic resp. R_h	66.1	65.5
D	Autotrophic resp. R_a	89.9	same
E	TER (E = C + D)	156	~~156~~ 155
F	NPP (F = B - D)	77.4	same
G	Direct land cover change	2.30	same
H	Fire	1.60	2.10
I	Harvest	0.23	same
J	Herbivory	5.54	same

Table 3. ~~r^2~~ The Pearson correlation r values for the different latitudinal zones between modelled heterotrophic respiration and the environmental drivers of the CBALANCE (CBA) and YASSO (YAS) models. The environmental drivers are all calculated as monthly means for the latitudinal zones. ~~The values denoted by asterisk~~ Significant correlation (~~*~~p-value < 0.05) ~~show a negative correlation. The positive relationships with the r^2 value of 0.80 or more~~ have been written in bold. α is the relative soil moisture, T_{soil} and T_{air} are soil and air temperature, and P_a is the precipitation.

Lat. zone	CBA vs. α	CBA vs. T_{soil}	YAS vs. P_a	YAS vs. T_{air}	YAS vs. α
60°N -90°N	~~0.04~~ -0.22	~~**0.92**~~ **0.96**	~~**0.91**~~ **0.95**	~~**0.80**~~ **0.90**	~~0.23~~ -0.48
30°N -60°N	~~0.65*~~ **–0.81**	~~**0.97**~~ **0.99**	~~**0.95**~~ **0.98**	~~**0.90**~~ **0.95**	~~0.85*~~ -0.92
0°N -30°N	~~**0.92**~~ **0.96**	~~0.24~~ 0.49	~~**0.92**~~ **0.96**	~~**0.86**~~ **0.93**	~~0.34~~ 0.58
0°S -10°S	~~**0.84**~~ **0.92**	~~0.00~~ 0.03	~~**0.86**~~ **0.93**	~~0.27~~ 0.52	~~0.20~~ 0.46
10°S -30°S	~~**0.88**~~ **0.94**	~~0.14~~ 0.38	~~**0.87**~~ **0.93**	~~**0.82**~~ **0.92**	~~0.22~~ 0.48
30°S -60°S	~~0.21*~~ -0.46	~~0.58~~ **0.76**	~~0.60~~ **0.78**	~~**0.91**~~ **0.95**	~~0.83*~~ **–0.91**

Figure A1. Different annual cycles of the heterotrophic respiration (R_h) predicted by the YASSO (a) and CBALANCE (b) model and the different alternatives from the box model calculation.

Table A1. The amplitude of global heterotrophic respiration within a year in different box model formulations. The input variables or functions that differ from the original model formulation are in bold letters.

Line	Option	Amplitude (PgCyear^{-1})
A)	YAS - Original model	3.8
B)	YAS with inputs $\mathbf{T_{soil}}$ and α and functions $\mathbf{f_{CBA,T_{soil}}}$ and $\mathbf{f_{CBA,\alpha}}$	2.7
C)	YAS with inputs $\mathbf{T_{soil}}$ and α and functions $f_{YAS,T_{air}}$ and f_{YAS,P_a}	3.7
D)	YAS with inputs T_{air} and P_a and functions $\mathbf{f_{CBA,T_{soil}}}$ and $\mathbf{f_{CBA,\alpha}}$	3.0
A)	CBA - Original model	2.3
B)	CBA with inputs $\mathbf{T_{air}}$ and $\mathbf{P_a}$ and functions $\mathbf{f_{YAS,T_{air}}}$ and $\mathbf{f_{YAS,P_a}}$	4.2
C)	CBA with inputs $\mathbf{T_{air}}$ and $\mathbf{P_a}$ and functions $f_{CBA,T_{soil}}$ and $f_{CBA,\alpha}$	3.2
D)	CBA with inputs T_{soil} and α and functions $\mathbf{f_{YAS,T_{air}}}$ and $\mathbf{f_{YAS,P_a}}$	4.0