

General words to Referees.

We truly thank both referees for their generally positive comments and highly valuable suggestions toward the manuscript. Here, we greatly appreciate their expertise.

We are providing our responses to these comments point by point as shown below. We have to admit that the corrections we are making are a product of compromise between referee suggestions. Referee 2 points out that the manuscript is too long and suggests significantly shortening and simplifying the manuscript (for example, removing the sensitivity analysis II), while Referee 1 proposes adding extra simulations and new data for testing new ideas. Given such a fact, we tried to make a rational decision by considering multiple factors while keeping the manuscript well-balanced.

We hope the revised version is a good one that can both meet Referees' expectations and the journal's criteria.

PS: The number of page and lines mentioned below correspond to those in the **version with traces**.

- **Referee 2, anonymous**

Referee 2: Mao et al. submitted an interesting manuscript about the evaluation of the Yasso07 model against RENECOFOR dataset a French network of forest plot. The paper is generally well written and the methodology sounds. It also fits well with the Biogeosciences scope.

Authors: Thank you for this positive remarks !

Referee 2: Nevertheless, the main message of the manuscript which seems to be that Yasso07 may not be the best tool to evaluate soil carbon changes in forest when it used outside the context of boreal forest where it has been originally developed is a bit diluted because the manuscript is too long. In particular, I suggest moving the sensitivity analysis in supplementary material (Fig 6 to 8). Regarding the sensitivity analysis, I did not fully understood the Module II and the interest to test effect of simulation length; this should be removed or better explained.

Authors: We are aware of the current length of the manuscript and also the visually complex Figures associated with the Sensitivity Analyses (SA), especially the Module II (although their originality sounds). To better focus on RENECOFOR data fit to Yasso07, now we decide to:

- (i) move the initial Fig. 6 (one of the two figures corresponding to the SA – Module I) and initial Fig. 8 (the only figure corresponding to the SA – Module II) to Supplementary Materials (SM), see the **Fig. S8** and **Fig. S9** in SM1.
- (ii) replace initial Fig. 7 (boxplots corresponding to SA Module I) by a more understandable **Fig. 6**, which shows steady-state carbon quality as a function of initial carbon stock for all the 101 RENECOFOR sites.
- (iii) simplify, accordingly, the descriptions to these results, see **Sect. 3.3, P18** and **Sect. 3.4, P 18**.
- (iv) simplify and clarify the description of SA in Materials and Methods, see **Sect. 2.6.1** and **Sect. 2.6.2, P13**.

Besides, we also made an effort to shorten the manuscript, when it is necessary and possible. For example, we also put the initial Fig. 1 (Yasso07's model structure) in Supple. Mat, see **Fig. S1**. This is because Yasso and Yasso07 are fairly well-known and many papers working on Yasso07 do not necessarily show such a figure. Reducing those above text and figures also provides us opportunities for adding two analyses suggested by Referee 1 without too expanding the length of the manuscript.

Referee 2: Minor comments: P2 L3 I am not sure Yasso represents the whole state of the art.

Authors: Sorry for this ambiguity of expression. We would mean the issues encountered in the use of Yasso07 are representative ones in the current modelling of soil carbon dynamics. So now we rephrase this sentence using "current bottleneck" instead of "the state of the art", see **P2, LN1-6**:
"We revealed, taking YASSO07 as model support, the current bottleneck of soil carbon modelling due to lacking knowledge or data on soil and litter carbon quality and fine root litter quantity, rendering high uncertainties for model inputs."

Referee 2: Some mechanisms are missing and it has a humus pool whereas the humus concept is now criticized (Lehmann, J. & Kleber, M. 2015)

Authors: Yes, but pool based models are still prevailing ones widely used in research and development. We decide to add this information and also the reference to remind readers when introducing the "H" pool, see **P7, LN9-11**.

Referee 2: P6 L3-4: Are that information not available in the ICP forest network?

Authors: We should say it is hard and extremely rare to obtain such a national scale dataset that contains such complete information (climate, soil with time series, fine and coarse litterfalls with time series) which are usually done on very instrumented sites (not national networks). Many countries involved in ICP Forests have such data but the main strength of the RENECOFOR network is to have, for two soil surveys, data obtained with exactly the same methods making estimation of SOC change possible. As far as we know, similar data also exist in some countries out of Europe (e.g. in China), but still remain inaccessible to us.

To avoid being too absolute, we decide to delete the sentence, see **P6, LN5-6**, since we have already highlighted the rarity of the dataset before, see **P5, LN30**.

Referee 2: P8 L12-13: In the original dataset to calibrate the model is there some data coming from RENECOFOR sites?

Authors: no, because Yasso07 was first published in the year of 2009 (Tuomi et al., 2009), i.e., the year when the RENECOFOR's 2nd soil inventory campaign was still ongoing. The dataset was first published in 2017 (see Jonard et al. 2017) and this is the first time that the dataset is used for testing Yasso07.

Referee 2: P13 in eq. 7 the second line of the equation should be $ACC_{sim} = (C_{sim,t2} - C_{sim,t1}) / (t2 - t1)$, right? If not please better explained, if yes please check that this only a typo mistake and the calculation were made the good way.

Authors: After checking, our equation should be the right one, because we used the observed C stock at t1 ($C_{sobs, t1}$) as the input to simulate the C stock at t2 ($C_{sim,t2}$).

Now, following the suggestion given by Referee 1, we also performed simulations to calculate the stock until 1 meter and had this steady-state stock value (CS_{steady-state}) compared with CS_{obs,t1} down to a depth of 1 meter. Please see below and also the text. See also **Fig. S4**.

Referee 2: Table 2: is 'ignorable' the good terms do you mean negligible?

Authors: Done. "negligible" is now used, see **P35, LN10**.

Referee 2: Fig. 3: Please don't call the non-hydrolysable compounds N. It is a misleading acronym since it is more used for nitrogen.

Authors: we did notice this potentially misleading term, but we think that it is more important to follow the Yasso07 inventors' given terminology. This allows keeping consistency among studies working on Yasso07 and facilitating inter-study comparisons. Moreover, in the case of this paper, we don't think the use of "N" can be really misleading, as "nitrogen" was always fully spelled when appearing in the main text.

We decide to add a note in the table of "Nomenclature and abbreviations", saying that in none of case "N" means nitrogen in this paper and when nitrogen is mentioned (for example in Figure. 5 and Figure S7), we used "nitrogen" , not "N". See **Page 3**. In order to avoid too many acronyms, we checked the text and kept using "carbon" instead of "C."

- **Referee 1 (R1), T. Wutzler (twutz@bgc-jena.mpg.de)**

Referee 1 : The study presents a model- data comparison at multi-site scale of forest sites which are relevant for management policies and accounting for global climate negotiations. The presentation is good and I could understand what has been done. Especially the litter quality database part is already valuable to other scientists.

Authors: Thank you Thomas for your positive remarks!

For the model-data comparison I have several remarks of what should be done additionally/differently, that potentially could alter the conclusion quite severely. Because of the paper did not change much compared to the pre-public-discussion, I repeat my comments in the this public discussion.

Authors: Sorry for the delayed responses, as it took us some time to obtain the new soil data from the network and to perform additional analyses.

Referee 1: 1) Steady state and observed stocks: The authors computed litter quality (percentages) from steady state computations and then scaled all pools down so that the sum matched observed initial stocks. Assuming that lower stocks resulted by recovery from disturbance, however, the composition of the faster pools should be closer to steady state than the slow pools. I recommend repeating the simulations with an additional initialization procedure according to Wutzler 2007.

Authors: The alternative method, i.e. the relaxed equilibrium assumption (REA) method, proposed in Wutzler (2007) is indeed very interesting and should definitely be better highlighted in our manuscript (see below). However, we do have concerns of applying such a method to this manuscript. We don't think that, until now, we've really have enough information to repeat the simulations using such an approach. How can we properly choose the current rate of assimilation ($\Delta C_c / \Delta t$ in Eq. 4) that might be site- or specific dependent? Shouldn't we still need to make some critical assumptions? With a changed AWENH composition, the results would probably be different (as showed in our sensitivity analysis Fig. 8, now Fig. S9 in Supple. Mat.), but would they be more reliable?

Even though we can do extra- sensitivity analyses to justify all the above things, but wouldn't all this make the manuscript's objective too diffuse, even shifted? For us, testing the regular and REA methods (just like the work performed in Wutzler (2007)) can be totally an independent study which corresponds to a new paper. When saying this, here we should add that, actually, we are indeed conceiving a new paper project tackling the issue of soil carbon quality initiation. Specifically, we aim to re-simulate the RENECOFOR sites' C dynamics in Yasso07, by using the site- and depth- dependant composition of carbon in different ages (determined by the 14C method, analysis still ongoing), instead of the regular initialisation methods. This project follows the idea of the newly published paper (Balesdent et al. (2018) Nature 559, p.599–602) that showed vertical heterogeneity of composition of carbon age along soil profiles. Also, this project's idea is in line with the hypothesis of the REA method, i.e., soil carbon quality may not be set as that at its perfectly steady state in theory.

Despite such a choice of not doing REA simulations and associated sensitivity analyses, we've decided to **expand our discussion** regarding this point. First, we cited this work and highlighted the existence of this method that merits more attention. Thus, we proposed therefore to perform an independent

study on the test of different initialisation methods by using different pool-based carbon models (Yasso, Yasso07, RothC etc.), as no such work has been done so far.

Additionally, we further pointed out that solely testing different methods of model initialisation, does not allow radically solving the uncertainty issue. We propose therefore considering specific or generic curves of carbon age ~ soil depth + ecosystem type in the future carbon dynamics modelling, following the key message of Balesdent et al. (2018).

Please see these added discussions in **Sect. 4.2, P22, LN 15 –26** and **P23, LN 7 –12**.

We hope you can understand such a decision we made with compromises and appreciate the improvements in the current version.

Referee 1 : Comparing different soil depths: The authors argue that stock changes are less susceptible to differences in soil depth than stocks, because the more stable pools reside in deeper layers. However, they did not account for this effect on initialization of stock qualities. I suggest instead transforming the observations (down to 1m) to the depth assumed by the YASSO model (0.4m) before comparison. This should be possible, because several depths were measured, e.g. by fitting a function to the depth distribution of bulk density and carbon concentrations and computing the cumulative stock up to a certain depth.

Authors: Thank you for this suggestion.

We contacted RENECOFOR and, fortunately, obtained the ground truth data of soil density for the depth of 40-100 cm for each site, although these data do not have the 5-subplot replicates as the 0-40 cm ones. Now we have estimated the carbon stock until a depth of 1 m based on some these additional data.

Following your suggestion, we now are able to compare the observed C and simulated carbon stock until a depth of 1 meter. Because of the length of the manuscript (which is the major criticism of Referee 2) and absence of replicates for 40-100 cm, we still would like to focus mainly on carbon change (ACC) as our major objective rather than on carbon stock (CS). But the latter can indeed be considered a good way of checking Yasso07's theoretical prediction. Running this simulation also gives us a good opportunity to show RENECOFOR site-dependent steady-state carbon quality, which is shown in a new **Fig. 6** replacing the old boxplot Fig. 7.

Accordingly, we put the plot related to carbon stock in Supple. Materials (see **Fig. S4**) and gave descriptions in Results (see **P16, LN26-30** and **P18, LN16-19**). Certainly, we also added related information in M&M on observational data of 40-100 cm (see **Sect. 2.2.1, P9, LN2-17**) and simulation (see **P14, LN22-27**).

Referee 1: Effects of mineralogy and potential stocks: The authors did not explain variation in residuals well by studied explanatory variables. I suggest including some soil mineralogy measures. Additionally, one could include potential stocks as derived from mineralogy by Feng 2013 and Beare 2014 or the indicators by Rasmussen 2018 to include a measure of distance to potential.

Authors: Indeed, soil texture and mineralogy greatly affect soil biogeochemical cycling and carbon stock. Follow this idea and your valuable suggestion, we contacted RENECOFOR and obtained a new dataset including soil physical (texture) and chemical (pH, stocks of total nitrogen, total phosphorus, exchangeable Al, K, Ca and Mg) of the 101 sites.

We added these variables to the residual analyses. We added a new table in Supple. Mat. For the linear regression results for all of the 11 variables (See **Table S2**). The associated PCA in Supple. Mat. has been updated (see **Fig. S7**). Further, in the main text, we added a new plot about effect of soil properties on residuals as **Fig. 5**. Associated result descriptions and discussions concerning these added results can be found in the main text, see **P16, LN32-P17, LN5, P20, LN1 -21 and P27, LN20-22 and P27, LN29-31**.

General comments (locations refer to the pre-public-discussion version)

Referee 1: p3125: The authors claim that at annual time aggregation, first order decomposition is adequate. However, largest criticism of first order comes from interaction among pools, like priming instead of time aggregation (Wutzler 2013)

Authors: Adding pool interactions will alter Yasso07's fundamental configuration and this is no more the major purpose of the manuscript. So we highlighted this point in the text by citing this work to draw future readers' attention, see **P4, LN21-23**.

Referee 1: p415: The authors claim to be first study of larger scale YASSO application. I know that YASSO is the soil model of the MPI earth system model implemented by Tea Thum, and suspect that there should be also larger scale studies.

Authors: in **P4LN15**, we've used the word "rarely" to avoid to being too absolute. We also deleted the statement to avoid confusion, see **P6, LN5-6**.

Referee 1: Sect. 3.4 and complicated figure 8 express the simple fact that there are initially high changes and later on slower changes in recovering C-Stocks. They can be shortened very much.

Authors: We now have decided to move this figure to Supple. Mat, following the suggestion given by Referee 2. Accordingly, the Section 3.4 are shortened, see **Sect. 3.4, P18-19**.

Referee 1:

References

Beare M, McNeill S, Curtin D, Parfitt R, Jones H, Dodd M & Sharp J (2014) Estimating the organic carbon stabilisation capacity and saturation deficit of soils: a New Zealand case study. *Biogeochemistry*, Springer Science + Business Media, 10.1007/s10533-014-9982-1

Feng W, Plante A & Six J (2013) Improving estimates of maximal organic carbon stabilization by fine soil particles. *Biogeochemistry*, Springer Science + Business Media, 112, 81-93 10.1007/s10533-011-9679-7

Rasmussen C, Heckman K, Wieder W, Keiluweit M, Lawrence C, Berhe A, Blankinship J, Crow S, Druhan J, Pries C, Marin-Spiotta E, Plante A, Schädel C, Schimel J, SierraC, Thompson A & Wagai R (2018) Beyond clay: towards an improved set of variables for predicting soil organic matter content. *Biogeochemistry*, Springer Nature, 137, 297-306 10.1007/s10533-018-0424-3

Wutzler T & Reichstein M (2007) Soils apart from equilibrium – consequences for soil carbon balance modelling. *Biogeosciences*, 4, 125-136 10.5194/bg-4-125-2007.

Wutzler T & Reichstein M (2013) Priming and substrate quality interactions in soil organic matter models. *Biogeosciences*, 10, 2089-2103 10.5194/bg-10-2089-2013.

Authors: We cited Wutzler and Reichstein M (2007), Wutzler and Reichstein M (2013), Beare et al., (2014) and Rasmussen et al. (2018), i.e. the four of the five references in this revised version and thank you again for your time and effort for the improvement of the manuscript.