This document contains referee comments to bg-2021-327 manuscript: 'Implementation and initial calibration of carbon-13 soil organic matter decomposition in Yasso model' and our responses to these comments. The referee comments are presented as indented text.

Referee 1 comments

> This manuscript describes new stable carbon isotope capabilities added to the Yasso model. The new model capabilities are described clearly. The model updates were parameterized and evaluated using measured datasets in a way that was well described and justified. Overall, I though the manuscript was a clear and concise description of a valuable new model capability. 13C measurements are a common metric for understanding soil organic matter decomposition processes and adding this capability to a SOM model is a valuable advance.

> I did think that in some areas the introduction and conclusions went beyond the scope of the actual results. Specifically, the model developments and testing were entirely focused on 13C fractionation and did not include changes to or evaluation of overall soil C decomposition rates. Therefore, the hypothesis in the introduction about "significant improvements in SOM decomposition predictions" seems broader than is justified. The study does yield improvements in predictions of 13C dynamics, but this was not used to improve overall SOM predictions.

This is a valid critique, and we will reformulate the text to better reflect the work. The main reason we did not examine the actual SOM predictions is to narrow the focus of the paper to 13C fractionation. The Yasso model has also been extensively calibrated with global SOC data in Viskari et al. 2022 (references at the end of the document). Additionally, this work was done as a proof-of-concept and to direct future development of Yasso w.r.t adding multiple soil layers and so forth.

> The first two paragraphs of the introduction (lines 10-20) provides a good justification for improving SOM models. However, the focus in these paragraphs on agricultural soils and carbon monitoring is not well related to the actual model structure and evaluation which only includes litter decomposition and peat systems. Carbon sequestration in mineral soils is sensitive to mineral-organic interactions and mineral-associated organic matter accounts for a large fraction of SOM (e.g., Lugato et al., 2021). However, Yasso does not include mineral interactions and treats humus as a passive pool and was only evaluated using litter and peat decomposition. Therefore, it does not seem justified to introduce the model in the context of agriculture soils. Since the model seems intended to simulate peat systems, I think it would be more reasonable to introduce it in the context of better understanding and predicting carbon dynamics in peatland or organic soils.

> Reference: Lugato, E., Lavallee, J. M., Haddix, M. L., Panagos, P., & Cotrufo, M. F. (2021). Different climate sensitivity of particulate and mineral-associated soil organic matter. Nature Geoscience, 14(5), 295–300. https://doi.org/10.1038/s41561-021-00744-x

The text will be modified on these parts and to shift the focus more on 13C in the context of ESMs as referee 2 has suggested. However, Yasso was originally build and used primarily for C decomposition in mineral soils and C measurements from mineral soils have been an important part of the calibration data (Viskari et al. 2022). Additionally, humus pool decomposition has been part of Yasso calibration on bare fallows (Viskaro et al. 2020). Since Yasso was recently recalibrated, the focus of this work is not on the actual decomposition rates but rather on the relative difference between 12C/13C decomposition. We used collected litter matter that has been left on the ground in litter bags, to calibrate the relative differences because that was the only data available to us at this point. More complicated SOC models do exist, that take into account e.g. microbial influence or nitrogen and phosphorus but the implementation of these processes is hindered by the lack of good quality data.

Other comments:

> Section 2.1: The peat depth profile measurements that were used to validate the model should also be described in this section.

These will be added.

> Figure 1: It would be helpful if the figure axes used the L notation that is used in the text so it is clearer what is being plotted. Is marginal likelihood in these plots the same as L?

Yes, the marginal likelihood is the same as L and we will modify the image as requested.

> Figure 2: Consider using different symbols for the branch and needle data to accommodate red-green colorblindness (which is common) or in the case of printing the paper in grayscale.

This will be amended.

> Line 138: It was not immediately clear to me how relative 13C content can change over time in the default model without any fractionation included. I think this occurs because the initial pools have different isotope ratios and are mixing over time which causes the isotope ratios to change. But a more specific explanation of this would be helpful. It might also be helpful to show a diagram (perhaps in the appendix) of transfers among the different pools so it is more clear what kind of mixing over time can occur.

That is correct, also the flow of matter between the pools differs so both the initial ratios and the different decomposition rates drive this behavior. A more detailed explanation will be added to the text along with a diagram of C transfers.

> Line 145: The actual depths should be included. And I suggest including a more detailed explanation of why the depth sampling was consistent with the 10 year age assumption. Was there evidence from that site that the age difference was actually close to 10 years across depths?

These will be added and we have included here an excerpt from Hilasvuori et al. 2013, where the interval estimate originates: "The $F^{14}C$ values in the peat profile peaked in 26-32 cm (Fig. 2), suggesting strongly that the accumulation of this layer had coincided with the high atmospheric $F^{14}CO_2$ values in the 1960s. Based on this peak, we were able to approximate that the peat in the uppermost 20 cm had accumulated during the last 30 years and the 20-26 cm and 26-32 cm layers represented approximately 10 years of peat accumulation each." We will also shift the first index to year 15 and the others also accordingly.

> Figure 3: I suggest splitting this figure into separate panels as in Figure 2. The large number of lines and colors makes the figure difficult to interpret. Also, can bulk 13C in the model be calculated to compare with the bulk 13C measurement from peat?

The figure will be split into multiple panels, and we will add the bulk 13C values.

> Line 162: The negative parameter values are consistent with the theoretical expectation of slower 13C decomposition rate (as described in the introduction) which is a good result for the model and would be valuable to point out more explicitly.

Agreed.

> Line 167: "This situation is not ideal" – why not? Is it inconsistent with measurements or theoretical expectations? It doesn't seem particularly unreasonable to me.

This was meant to underline that the outcome (one positive parameter value) goes against the initial hypothesis, but we do agree that the wording here is not “ideal” and will be changed. Likely, we will merely state precisely what was written here.

> Line 179-180: It’s not clear to me how the results demonstrate improvement to SOM model accuracy and predictability since they were not used to inform any changes to the overall C decomposition rate or structure. Improvements were limited to 13C dynamics.

This was meant more as a general statement and will be modified to reflect that better. The idea is that by including and examining 13C we are adding more constraints to the model that can be used to improve the model and estimates. For example, we can analyse flux data to get more information about the state of the C in the soil (especially if we do not have any measurements from the soil) and how different managements have altered these values; we are able to estimate long term carbon stocks and transfer rates based on input/output C fractionation and we can evaluate how well the model works in different environments.

> Line 189: Similarly, it’s not clear that the study made improvements to SOM decomposition in general outside the direct comparisons to 13C content of organic matter pools.

We will refrace this sentence as well.

> **Referee 2 comments**
>
> The preprint manuscript “Implementation and initial calibration of carbon-13 soil organic matter decomposition in Yasso model” describes calibration of the Yasso model to 13C data collected from a litterbag decomposition experiment. The model was calibrated using 13C values measured on sequential extracts of pine litter and branch samples from a 4-year litterbag experiment. The decomposition parameter matrix of the Yasso model was modified to account for 13C using simple scalars. After optimization, three out of 4 scalars were negative, which was consistent with the hypothesis that 13C is preferentially retained in decomposing organic matter. The optimized model was applied to data from a peat core and produced more realistic predictions than the default model.
>
> This manuscript is clear and concise. However, I think this manuscript should be framed differently to better showcase the results. The manuscript is framed narrowly in terms of soil carbon sequestration as a climate mitigation tool. However, the analyses and results are not directly relevant to soil carbon sequestration efforts. Specifically:
>
> - The study system is unmanaged and focused on C cycling in litter and organic soils, and has no obvious connection to the agricultural soil carbon management strategies listed in the introduction.
> - The 13C calibrated model performs no better at predicting changes in bulk C, hence its relevance to soil carbon measurement and verification efforts are unclear or at the very least indirect.

This critique is on point and on similar lines as Referee 1 and the text will be amended accordingly. We will reframe the work more in the concept of ESMs (as is suggested in the next referee comment). Additionally, the proof-of-concept itself (that we get the model to work with these modifications) is valuable as there are ongoing measurement campaigns in Finland on managed soils and hopefully, at some later date, we will be able to expand the experiment.

> Later in the manuscript the significance of the 13C calibrated Yasso model is described differently, in terms of integration with 13C enabled ESMs. This seems like a much clearer justification for the calibration effort. Taken at face value, the results presented here are nearly trivial: calibrated the Yasso model to 13C data results in a better fit to 13C data. As a technical result, this is to be expected. What is the concrete significance of this incremental

advance for our understanding of soil carbon cycling? What can the calibrated model eventually tell us about the cycling of the bulk C pool or the broader functioning of soil beyond fractionation of 13C?

The benefits of adding 13C into the model will be made clearer but we need 13C to be able to use this isotope for learning about the processes involved. We can, e.g. and in addition to the referee suggestions in the next comment, analyze site-level flux data (respired carbon isotopes) to get more information about the state of the C in the soil (especially if we do not have any measurements from the soil); we are able to estimate long term carbon stocks and transfer rates based on input/output C fractionation and we can (hopefully and eventually) assess the effect of soil management practices based on the impact on respired C isotopes. Additionally, one goal is to assess the stability of long-term soil C storages under changing climate.

If the 13C modifiers are generalizable to other systems (which may or may not be the case), I can see how they might enable the Yasso model so that it could be calibrated based on tracer experiments or in cases where the d13C of vegetation has shifted, or how it might be useful for interpreting time series of 13CO2 data to attribute fluxes to different soil C pools. These sorts of application are alluded to, but perhaps the manuscript would stand on its own more clearly if it was framed more clearly as an intermediate step towards these larger scientific goals.

Thank you for the comment and we will be making these changes to the manuscript.

Detailed comments:

Abstract: Details of the calibration dataset are not given in the abstract – consider including them.

These will be added.

Line 1; Line 10: I agree that strategies for increasing soil carbon as a climate mitigation strategy have received increasing attention over the years. However, I think this initial framing is an innapropriate place to start this manuscript (see broader comments above). Carbon cycling in soil is a fundamental aspect of terrestrial ecosystem function. Soil carbon influences the climate system and a whole range of global biogeochemical cycles regardless of how we try to manage it.

We will modify the text and frame the work more in the context of land surface or Earth System models.

Line 7: I suggest deleting “despite of their simplicity”, as it implies that we expect that simple modifications will not generate improvements.

Agreed, this will either be deleted or modified. On a side note, this alludes to a modelling paradigm where simpler models are preferred to, e.g., avoid overfitting.

Lines 21-32: This paragraph begins by addressing the challenge of deciding which processes to include in models, but the application for 13C seems to mostly relate to parametrization. Is 13C useful for both determining model structure and fitting parameters? Are these distinct challenges?

The answer to this question depends on your point of view but we tend to regard these as one issue. The model structure should reflect what the model is used for or what is the underlying scientific question. In this sense, 13C is a natural addition to the Yasso model (and for many other soil carbon models as well), e.g., for verification purposes. The parameters reflect our current understanding of (often latent) variables and processes. Fitting the parameters is an open-ended quest as new innovations, data, and model modifications all require parameter tuning.

Line 28: Writing edit -- delete “By” before “estimating”.

Will be corrected.

Lines 114-115: In other words, the precipitation and temperature dependence was the same for both isotopes? These factors are included in the original “alpha” term?

Yes, to the first question. The temperature (T) and precipitation (P) dependency is slightly more complicated, and we will add these explanations to the manuscript if needed. We have added here equation 5 from Viskari et al. 2022 (reference at the end of the document) that shows the decomposition rate functions (k_i), where the original “alpha” is multiplied by *h(d)*, decomposition rate response to litter diameter, a term containing precipitation dependency and lastly the sum including temperature dependency (all with additional parameters).

$$k_i(\theta, c) = \frac{\alpha_i}{J} h(d)\left(1 - e^{\gamma_i P}\right) \sum_{j=1}^{J} e^{\beta_{i,1} T_j + \beta_{i,2} T_j^2}$$

Line 126: how were the parameter “grid” and increment refocused? Was this done in a systematic way?

Yes, and multiple times with different initial states. The supporting material contains one run with the starting point at the origin (we end up with same results regardless of starting point). We run the model with all combinations from the “grid”, where each parameter was varied similarly; then we choose the local optima as the new middle point for the grid, decrease the value increment and run everything again etc. We can modify the text to clarify this.

Figure 1: What do the color gradients represent? Likelihoods, presumably? In the panels situated along the diagonal, does the vertical axis on each panel show the likelihood? What do the vertical lines represent – parameter values at maximum likelihood? This caption needs to be expanded to clarify.

Yes, these are unnormalised likelihoods and the vertical line is on the MAP estimates. Clearly the caption needs to be clarified.

Figures 2-3: Why does d13C change over time in the default case? The default parameters are identical for 12C and 13C, correct? In this case, shouldn’t the 12C:13C ratio be preserved in all transformations, and the d13C value remain the same over time?

The different pools have different initial content of 13C, and they have different decomposition rates and the flow of matter (flow rates) between the pools are different. All these combined drives the d13C change over time in the default case. This explanation will be added to the manuscript.

Methods section: Please include details about the computing methods. How were these procedures implemented? What computing environment was used (e.g., Python, R, Matlab)? Were any R packages used to assist with fitting?

All of this will be added. We used the R version of Yasso, which has some R package dependencies, and an additional R script (with R.utils) to iteratively handle the model runs – no additional packages were used with the fitting since in this limited experiment there was no need for e.g., MCMC or importance sampling-based approaches. All experiments were run on a 8-core laptop.

Lines 131-132: I believe there are formal methods for evaluating collinearity between parameters. Computing a “collinearity index” might be useful for determining whether the parameters are identifiable (although such indices still reduce to qualitative rules of thumb). There are methods in R for this sort of analysis (package “FME” might be useful).

Parameter (multi)colinearity is not usually reported. It means that we interpret the parameters as independent variables (predictors) and the likelihood as the dependent variable and model their relationship *via* linear regression; then colinearity means that there (nearly) exists a linear mapping between two parameters and multicolinearity means that one predictor is close to a linear combination of the others – so we can (nearly) exactly predict one parameter value from the rest. If needed, we will add condition indices to measure multicolinearity.

> Lines 151 – 152: Here the emphasis is on incorporation into ESMs, not MRV for soil carbon sequestration.

Yes, as stated before we will refocus the work more on LSM/ESMs.

> Lines 145-146: So depth and time have been exchanged? Is this based on an assumption that the peat is accreting linearly? How was the conversion between depth and time parametrized? Why 10 year intervals, why not 20 or 50 years? More justification/expanation is needed here.

Yes, depth and time have been exchanged and our assumptions about peat age come from Hilasvuori el al. (2013) - the interval estimate originates: “The $F^{14}C$ values in the peat profile peaked in 26-32 cm (Fig. 2), suggesting strongly that the accumulation of this layer had coincided with the high atmospheric $F^{14}CO_2$ values in the 1960s. Based on this peak, we were able to approximate that the peat in the uppermost 20 cm had accumulated during the last 30 years and the 20-26 cm and 26-32 cm layers represented approximately 10 years of peat accumulation each.” We will also shift the first index to year 15 and the others accordingly as well as extend the timeseries a bit.

> Lines 167-169: I do not follow this reasoning. Is the non-ideal finding that the parameter for the N pool is positive? How does the lack of depth resolution explain this?

This was meant to underline that the positive N-pool related parameter value goes against the initial hypothesis, but we do agree that the wording here is not “ideal” and can be changed. However, we do provide an explanation for this behavior which is related to the flow of matter between the soil carbon pools. The flow itself is dependent on temperature and respiration, which should affect the model behavior differently, based on the soil depth. Adding soil layers would enable us to take these processes into account in more detail – it may or may not “fix” the issue. The positive value itself is not really a problem, it is an artifact of model structure.

> Lines 179 – 180: The results presented here indicate that calibration of 13C parameters to 13C data improves accuracy and predictive power for 13C. However, they do not show how this improves the skill of the model with respect to bulk C pools or fluxes. What can these results tell us beyond 13C fractionation?

Our view of this work is that it is the first step in 13C modelling with Yasso. The modified model can be used in various ways beyond C13 fractionation, e.g., increasing understanding of soil carbon storages by examining CO2 flux data, but the results presented here are not meant to demonstrate that. However, we will add explanations about potential use cases to the manuscript.

Hilasvuori, E., Akujärvi, A., Fritze, H., Karhu, K., Laiho, R., Mäkiranta, P., Oinonen, M., Palonen, V., Vanhala, P., and Liski, J.: Temperature sensitivity of decomposition in a peat, Soil Biol. Biochem., 67, 47–54, https://doi.org/10.1016/j.soilbio.2013.08.009, 2013.

Viskari, T., Laine, M., Kulmala, L., Mäkelä, J., Fer, I., and Liski, J.: Improving Yasso15 soil carbon model estimates with ensemble adjustment Kalman filter state data assimilation, Geosci. Model Dev., 13, 5959–5971, https://doi.org/10.5194/gmd-13-5959-2020, 2020.

Viskari, T., Pusa, J., Fer, I., Repo, A., Vira, J., and Liski, J.: Calibrating the soil organic carbon model Yasso20 with multiple datasets, Geosci. Model Dev. 15, 1735–1752. https://doi.org/10.5194/gmd-15-1735-2022, 2022.