Dear Editor,

Attached is the new version of the manuscript “Calibrating the soil organic carbon model Yasso20 with multiple datasets” with the requested minor revisions. We hope that these revisions will address the remaining issues before the final acceptance for publication.

The responses to the specific requests are below and the lines given are for the revised manuscript.

“Dear editor, dear authors,

I was pleased to find that the revision done by the authors addresses all critical points that were raised in the previous round of reviews. I do not see any further critical issues that would require changes.

Before a possible publication, I would only recommend the following (optional) minor changes:”

Our gratitude once again for the helpful suggestions and for finding the resulting changes to be acceptable. We have also made additions to the manuscript according to the recommendations.

“Abstract: I would make it clear that this is a new Yasso version, so write NEW Yasso 20, and I would mention that you now also compare to the older model version”

We adjusted the abstract based on this and starting from line 10, the abstract now reads:

‘**Abstract.** Soil Organic Carbon (SOC) models are important tools for assessing global SOC distributions and how carbon stocks are affected by climate change. Their performances are, however, affected by data and methods used to calibrate them. Here we study how a new version of Yasso SOC model, here named Yasso20, performs if calibrated individually or with multiple datasets and how the chosen calibration method affected the parameter estimation. We also compare Yasso20 to the previous version of the Yasso model. We found that when calibrated with multiple datasets, the model showed a better global performance compared to a single dataset calibration. Furthermore, our results show that more advanced calibration algorithms should be used for SOC models due to multiple local maxima in the likelihood space. The comparison showed that the resulting model performed better with the validation data than the previous version of Yasso.’

“Likelihood: I accept your explanation regarding the fixed sd, and that it doesn’t make a large difference, but would still argue that it is fundamentally more appropriate to adjust the sd parameter during the calibration, rather than fixing it to the empirical standard error.”

As was the case during the previous revision round, we inherently agree with the reviewer on this. Even though we tested adjusting the sd parameters during calibration and found this to be a very useful exercise, we feel there needs to be a more thorough analysis of the uncertainty calibration as a separate study. For example, CIDET is a much more consistent dataset than LIDET in terms of measurement noise. During this exercise we found that calibrating the uncertainties would reduce the RMSE with LIDET validation dataset, but not with CIDET. We would like to study these interesting findings before we report them as the reviewer accepts our reasoning behind fixing the sd.

Accordingly, we added a part to the Discussion about the need of this kind of research. Starting from line 465, the manuscript now reads:

‘Our results show that simultaneously using multiple datasets from different environments improves the general applicability of the SOC model even when the simplistic leaching factor approach had to be used to be able to compare different litter bag datasets and detailed uncertainty estimates were lacking, confirming our first hypothesis. This is in line with prior studies arguing for larger representation in the calibration data (Zhang et al., 2020). Furthermore, a more detailed analysis of different calibrations shows (Figure 2) that the information from multiple datasets is in truth even necessary for the calibration as when calibrating only with one dataset, the decomposition parameter uncertainty ranges either were large or, in the case of the more nuanced EuroDeco dataset, don’t even appear to converge. Something that was not examined in this study was how the uncertainties for the different datasets should be defined. Even if the assigned measurement uncertainties were correct for each dataset, combining them introduces structural uncertainties that should also be accounted for (MacBean et al., 2016). A potential method to address would be to estimate the dataset uncertainties along with the model parameters, as done for example in Cailleret et al. (2020) but applying this approach to the SOC system will require a more thorough analysis in order to assess how it impacts the results.’