Response to comments from Reviewer 1

Authors:

We would like to thank the reviewer for the positive and perceptive comments, as well as for the questions, which will help us to improve the paper, specifically by giving more details, information and explanations about our data, methods and results.

Reviewer:

The study introduces a parsimonious soil organic carbon (SOC) turnover model for the soil profile that includes key processes controlling carbon persistence. The model specifically incorporates two crucial mechanisms often omitted in simpler models: (1) microbial energy limitation (a form of positive/negative priming where decomposition slows if substrate is scarce) and (2) physical protection via soil aggregation (which protects organic matter from decomposition). The aim was to test this model against long-term field data, identify how well model parameters can be determined (parameter identifiability), and analyze sensitivity of the model to its parameters. However, most of the parameters were not constrained and covariance between the parameters were used as an explanation to parameter unidentifiable.

The manuscript is well-written and please find the specific comments as follows:

Comment 1

Line 34, 36: What soil depths represent topsoil and subsoil? Providing the depth ranges would be helpful for clarity.

Authors: Here, we define the topsoil as the cultivated (tilled) soil layer, rather than in terms of depth ranges.

We will clarify this in the revised version by writing: "*Most model applications to date have focused on the **cultivated** topsoil, which is clearly of major importance with respect to the effects of soil management on SOC and soil health"*

This will also be added at line 54 "Furthermore, these models have almost exclusively been tested using measurements in **cultivated** topsoil"

Comment 2

Line 45-49: The study demonstrated parameter uncertainty and equifinality, despite having fewer parameters than complex models. However, this raises the question: how is the simple model used in this study different from detailed mechanistic models if parameters remain unconstrained? An explanation would help readers understand the trade-offs between complexity and parameter uncertainty.

Authors: Yes, this is a good point. It's true that even the simplest models of organic matter turnover in soil can show equifinality, depending on the type, quantity and quality of the data used to constrain them. This has been demonstrated for models that are even simpler than the one we developed and tested in our study (see e.g. Juston et al., 2010; Luo et al., 2017). We were therefore expecting to encounter the issue, as we wrote at line 233. This is also why we wrote "*may*" on line 48.

However, in contrast to more complex models, we can clearly see why and where the equifinality arises in our relatively simple model: it depends on the model structure, with correlations among only a few parameters, which makes the problem of parameter uncertainty more manageable. This was the case for the model application to the data at the Ultuna long-term soil organic matter experiment presented in the paper: here, simultaneous calibration to data was sufficient to effectively constrain the model parameters in three treatments with strongly contrasting inputs of

OM with respect to both type and amount. This is what we concluded at lines 317-318: *"These strong correlations of k_O, A_Q and k_{till} with ε mean that, in practice, all four parameters are well constrained by the calibration".*

In the revised version, we will clarify these issues by revising and adding to the text at lines 50 to 55:

" …. their reliability in extrapolation (i.e. prediction of independent data) has not yet been convincingly demonstrated (Garsia et al., 2023; Le Noë et al., 2023). This is because these models have often been tested against insufficient datasets (i.e. observations of topsoil C dynamics at a single site and treatment) which increases the likelihood of equifinality despite the small number of parameters (e.g. Juston et al., 2010; Luo et al., 2017). This may be overcome by simultaneous calibration of the model against data for two or more contrasting treatments, for example with respect to the type and quantity of organic matter inputs (e.g. Meurer et al., 2020) or by multi-site calibration at larger scales using data from long-term field trials at locations with contrasting soils and management practices (e.g. Juston et al., 2010; Dechow et al., 2019). Testing model predictions for entire soil profiles remains however difficult and is therefore rarely done, because fewer measurements are made in subsoils and the turnover of organic C in subsoil is very slow, so datasets will rarely be long enough to detect any changes. Additional data sources may also help to alleviate problems arising from equifinality. One possibility is to make use of … "

Comment 3

Line 70: What does ICBM stand for? Additionally, the reference to Andrén and Kätterer (1997) is missing. A schematic diagram illustrating model development over time would help readers visualize how the model has evolved. Similarly, a conceptual diagram of the final model used in this study would be beneficial alongside the mathematical equations.

Authors: ICBM stands for *Introductory Carbon Balance Model*. We apologize for not including the reference to Andrén and Kätterer (1997). We will add this to the revised version in line 70.

We will also include a conceptual diagram of the model to go alongside the equations.

Comment 4

Line 233: Why were parameters in Table 1 fixed, while only parameters in Table 2 were used in the calibration?

Authors: Some reasons were already given at lines 234 to 239. However, we realize now that this description was insufficient. We will add some new text as well as supporting references in tables 1 and 2 in the revised version to mention these additional reasons. At line 231, the text will be revised to:

"*The model was simultaneously calibrated to the measurements from the three treatments using the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven, 2006; Beven and Binley 2014; Juston et al., 2010) …. Six of the fifteen parameters were included in this analysis (Table 1) as they were judged to be uncertain and sensitive. The remaining nine parameters were set to fixed values (Table 2) as they could be estimated from measurements (e.g. f_{clay}, f_{agg}, F_p) or they were not expected to be sensitive (e.g. k_y, Andrén and Kätterer, 1997; Juston et al., 2010; Meurer et al., 2020), or both (e.g. ψ_{ae}, ψ_{mic}, ϕ_{min}, γ_o, γ_m)*".

Comment 5

Line 253-254: Why was the mean model efficiency (EF) across all three treatments used to identify acceptable parameter sets?

Authors: Because we wanted to obtain a common parametrization for all three treatments. We will clarify this in the revised version. We will add text at L254: "*This was done in order to obtain a more robust parameterization of the model by selecting only parameter sets that simultaneously fitted all three treatments well.*"

- Does this mean that the same parameter set was used for all treatments after calibration?

  Authors: Yes. This was implied earlier at lines 231 to 232. However, in the revised version, we will clarify this by writing explicitly that we wanted to obtain a common parametrization for all three treatments.

- Why not use treatment-specific parameter sets?

  Authors: This is as stated above to obtain a robust model parameterization of the model, i.e. parameters sets that should be valid for a wide range of conditions. This is also a way to reduce issues with parameter equifinality.

  Also, if parameter values must be changed to account for different treatments (in this case, the amount and type of organic matter inputs), it is a sure sign that something important is missing or wrong in the model, with respect to process descriptions. This means, in turn, that predictions made for contrasting conditions (re. OM inputs) using those calibrated parameter sets may be wildly wrong.

  We wanted to critically test the model to see whether it could match the data from the three treatments with a common parameterization. It passed this test.

- Wouldn't taking the mean EF lose treatment-specific information that could be valuable for refining the model?

  Authors: In principle, yes, it would, but only if the model had performed poorly. However, as shown in figure 2, and as we wrote at lines 308 to 310 (and in the abstract at lines 15 to 20), we were able to get excellent calibrated fits to the data from the three treatments with exactly the same parameterization. This demonstration of the capability of such a simple model is an important result of this study.

Comment 6

Line 257: 15 model parameters (Table 3?)

Authors: Yes, the 15 parameters in table 3 were included in the sensitivity analysis. We will refer to Table 3 in the text.

Line 320, 329,332, 346, 352: Graphs texts are too small and difficult to read.

Authors: OK, yes, we will make revised versions of these figures with more legible text

Comment 7

Line 314: Given that most parameters were not well constrained, could parameter covariance be a model artifact or a coincidence?

Authors: It is a consequence of the model structure, so not a coincidence. We feel that this is already stated quite clearly at lines 314 to 316.

(note that we would not like to call this a model *artifact*, because it is inherent to the model and not something that occurs by chance as a consequence of the calibration methods applied)

- Could the current dataset be insufficient to constrain these parameters?

  Authors: This is a good point: to some extent perhaps. For example, if we had the same kind of data from tilled and untilled soils (so 6 treatments at the site, three different OM inputs, with and without tillage), then we would probably have been able to more clearly identify the parameter k_{till} but only under the condition that the model describe the effects of tillage in a reasonable way (which we don't know yet).

- Would incorporating additional data sources (e.g., isotope data, incubation experiments) help resolve this issue? If yes, how can this modeling work be robust?

  Authors: We mentioned in the introduction (L55-58) that *in situ* isotope data is also a possible way to reduce some of the model parameter uncertainty, under the condition that they would be prove to be sensitive. We are more doubtful about incubation data based on disturbed/sieved soil samples. We think that the approach adopted here is anyway rather robust, as it included data from several treatments simultaneously.

  It is also worth noting that this is the first application of a new model. We wrote at line 404 in the section "*Concluding remarks*" that the tests of the model in this paper suggest that it "*shows promise*". We don't think this claim is unreasonably strong. And as we wrote at lines 407 to 409, a greater degree of confidence in the robustness of the model can, of course, be established over time by showing that it produces acceptable results when repeatedly tested against different data sets.

Comment 8

Line 316: Why were only the 30 best parameter sets selected?

Authors: Because their predictions were sufficient to cover the range of variability observed in the measurements which is a criteria for the GLUE method. This was implied in the text at lines 291 to 293, but it was not explained. We will mention this criteria in the M&M section after lines 231-232.

"The model was simultaneously calibrated to the measurements from the three treatments using the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven, 2006; Beven and Binley 2014; Juston et al., 2010). The number of acceptable parameter sets was determined such that the range of variation of their predictions approximately covered the variations observed in the measurements. With this criterion, 30 of the 12000 parameter sets were identified as acceptable".

- What was the acceptance rate of parameter sets out of 12,000 simulations?

  Authors: 30 out of 12000 = 0.25%. This small value is a consequence of the inefficient sampling which is inherent in the GLUE method: note that it says nothing about the quality of the model. We can state this in a follow-up sentence:

  "Note that this low acceptance rate is a consequence of the inefficient sampling inherent to the GLUE method and says nothing about the quality of the model."

Comment 9

Line 317: The phrase "strong correlation" is used, but no statistical analysis (e.g., correlation coefficients, p-values) is provided to support this claim. Including quantitative analysis would strengthen this statement.

Authors: Yes, we will add R^2 and p-values in a revised version of the figure. All four relationships are highly significant ($p<0.002$) with R^2 values varying between 0.30 and 0.76.

Comment 10

Line 338: Figures 5 and 6 are difficult to interpret.

- A more detailed explanation of what these figures represent would improve clarity.
- What key insights should the reader take from these figures?

Authors: Yes, we agree that this was not well explained. In addition to what we wrote at lines 336 to 338, these figures also suggest that the results of the sensitivity analysis should be "*reasonably well grounded in reality*". We wrote this at line 343 in connection with figure 7, but this conclusion should also be based on figures 5 and 6. In the revised version, we will modify the text at lines 342 to 343 to make this clearer.

Comment 11

Line 345: Minor inconsistency: (e.g., consistently use "Figure X" or "Fig. X rather than mixing "Fig. X" and "Figure X").

Authors: Yes, we will fix this in the revised version

References cited (in answers to both reviewer 1 and 2)

Andrén, O., Kätterer, T. 1997. ICBM: the introductory carbon balance model for exploration of soil carbon balances. Ecological Applications, 7, 1226-1236.

Chen, S., Zou, J., Hu, Z., Lu, Y. 2020. Temporal and spatial variations in the mean residence time of soil organic carbon and their relationship with climatic, soil and vegetation drivers. Global and Planetary Change, 195, 103359.

Coucheney, E., Kätterer, T., Meurer, K.H.E., Jarvis, N. 2024. Improving the sustainability of arable cropping systems by modifying root traits: a modelling study for winter wheat. European Journal of Soil Science, 75, e13524.

Dechow, R., Franko, U., Kätterer, T., Kolbe, H. 2019. Evaluation of the RothC model as a prognostic tool for the prediction of SOC trends in response to management practices on arable land. Geoderma, 337, 463-478.

Kätterer, T., Bolinder, M. 2024. Response of maize yield to changes in soil organic matter in a Swedish long-term experiment. European Journal of Soil Science, 75, e13482.

Juston, J., Andrén, O., Kätterer, T., Jansson, P-E. 2010. Uncertainty analyses for calibrating a soil carbon balance model to agricultural field trial data in Sweden and Kenya. Ecological Modelling, 221, 1880-1888.

Luo, Z., Wang, E., Sun, O. 2017. Uncertain future soil carbon dynamics under global change predicted by models constrained by total carbon measurements. Ecological Applications, 27, 1001-1009.

Meurer, K., Chenu, C., Coucheney, E., Herrmann, A., Keller, T., Kätterer, T., Nimblad Svensson, D., Jarvis, N. 2020. Modelling dynamic interactions between soil structure and the storage and turnover of soil organic matter. Biogeosciences, 17, 5025-5042

Poeplau, C., Don, A., Schneider, F. 2021. Roots are key to increasing the mean residence time of organic carbon entering temperate agricultural soils. Global Change Biology, 27, 4921–4934.

Pold, G., MacDonald, E., Braun S., Herrmann, A. M., 2025. Soil and vegetation property data from the Ultuna R3-RAM56 long-term soil amendment experiment, 1956-2023

Wutzler, T., Reichstein, M. 2013. Priming and substrate quality interactions in soil organic matter models. Biogeosciences, 10, 2089-2103.