In the document, feedback from the reviewers is written in italic face, while our responses are written in green. Screenshots of parts of the manuscript that have been changed are shown in boxes.

Reviewer 1

This manuscript discusses the issues of parameter identifiability and equifinality in the context of soil organic matter (SOM) decomposition models, and uses multiple model formulations and parameter optimizations to estimate variability across model structures and parameterizations and demonstrate how much predictive uncertainty may remain after models have been optimized using steady-state values.

I thought this was a very informative paper and a very useful study for highlighting challenges with modeling soil carbon cycle processes. The introduction explains the complex issues of parameter identifiability and related uncertainties clearly and makes a good case for paying more attention to these issues in soil models. The model simulations and approaches are well described and provide a clear demonstration of the key concepts of the paper, including how multiple model structures and parameter values can provide similar accuracy with respect to steady-state values while diverging when the steady state changes, and demonstrating how unidentifiable parameter pairs manifest in these types of models. Overall, the paper does a great job making an important argument and backs it up with strong modeling results.

We thank the reviewer for taking the time to read our manuscript and for providing detailed and very constructive feedback. This, together with your indications of parts of the manuscript that weren't properly described, significantly improved the quality of the manuscript. This is greatly appreciated. Please find our responses to the feedback below.

I have a few specific comments:

Line 84-86: This is true for non-microbial models as well

Thanks for mentioning this, our aim was not to state that this is not the case for non-microbial models. We therefore changed these sentences accordingly.

> in the field (Treseder et al., 2012). While SOM models thus became more mechanistic, their parameters are often "effective parameters" that represent multiple processes and cannot be directly measured (Beven, 2002), as is the case for non-microbial first-order models. Therefore, despite the mechanistic character of these models, multiple parameters need to be calibrated rather than being derived from measurements.

Line 106: Is there a missing word after "decades"? Maybe "unless Δ14 data..."

Thanks for pointing out this mistake, we deleted the part about $\Delta^{14}C$ as this was out of place.

> fect' behavioural models) can widely diverge (Luo et al., 2016, 2017; Guo et al., 2022). For example, it has been shown that models generally overpredict the turnover time of soil organic carbon (SOC) and thereby overestimate the potential of soil to increase their organic carbon (OC) stocks over the coming decades ~~, $\Delta^{14}C$ data for SOC is included as a calibration constraint (He et al., 2016; Luo et al., 2019)~~(He et al., 2016; Wang et al., 2019). The optimisation of non-identifiable parameters thus in-

Line 122: Typo in "assess"

Thanks for pointing this out, this typo has been corrected.

Line 129: Provide a reference here for the DEzs algorithm)

Thanks for this suggestion. While revising this part of the text, we noticed we incorrectly stated that the frequentist calibration was performed using the DEzs algorithm, while this should have been the DE algorithm. We corrected this mistake and added references for both algorithms. We note that the R packages used to perform these analyses are cited in the corresponding sections in the Methods.

> rameters compensating for each other, realistic values for all parameters need to be perturbed by a very small amount. Such values were obtained for all models by performing a frequentist calibration using ~~a Differential Evolution Markov Chain with snooker updater (DEzs) algorithm~~ the Differential Evolution (DE) algorithm (Storn and Price, 1997), given as many constraints on simulated pools as realistically possible. Next, to assess how different amounts of available calibration data affect model simulations in steady state, every model was calibrated using ~~a Bayesian approach (using the DE_{zs} algorithm)~~ the Differential Evolution Markov Chain with snooker updater (DEzs) algorithm (ter Braak and Vrugt, 2008) for parameter sets which were either identifiable (termed the identifiable parameter model; IPM) or non-identifiable (termed the full parameter model; FPM).

Line 166: The SOM model actually does keep track of the DOM pool size

Thanks for pointing this out, this was indeed not clear in the text. We have now clarified this in this section.

> The second model, termed the SOM model, is identical to the SESAM model but additionally simulates mineral protection of OC (Fig. 1) and keeps track of the size of the DOM pool. The processes are identical to the rhizosphere model, with the addi-

> enzymes (POM_{res}) and plant litter (POM_{lit}), microbes (MIC) and mineral N (N_{min}). In addition, in the rhizosphere model dissolved organic matter (DOM) and enzymes mediating the depolymerisation of plant litter (ENZ_{lit}) and microbial residues (ENZ_{res}) are simulated without explicitly keeping track of their size, assuming they are in quasi-steady-state (i.e., their size determines the rate of depolymerisation, but they are instantly transferred to other pools upon their creation). The SOM model, however, explicitly tracks the size of the DOM pool to simulate competition for DOM between microbes and mineral surfaces.

Line 182: The calibration procedure for the OC inputs was not explained clearly. If they were calibrated, does this make them essentially another parameter of the model? I think *having a precise value for carbon inputs is actually quite optimistic for comparison with "real" field data, since estimating litter inputs (especially for root and root exudation fluxes) is very difficult*

Thanks for pointing this out, we have better explained how the optimization of OC inputs took place and added a note on the consequences of fixing this parameter value.

> Annual total OC inputs were calibrated for the SOM model ~~(see Sect. 2.5),~~ to obtain a correct combination of the size of the combined *POC* pools and its turnover rate~~, determined by optimising its $\Delta^{14}C$ value .~~. This was done by optimising the value for OC inputs together with the other model parameters of the SOM model using a frequentist calibration approach

> (Sect. 2.5). This resulted in OC inputs of 349 g C m^{-2} yr^{-1} (Table S8), of which 80 % was assumed to enter the soil as plant litter, and the remaining 20 % as rhizodeposits. The difficulty in correctly estimating this parameter in field situations (e.g., Hirte et al., 2018; Pausch and Kuzyakov, 2018) and resulting effect on SOC model simulations (Keel et al., 2017; Taghiza therefore means that its uncertainty generally leads to an additional potential cause for equifinality on top of the uncertainty of model parameters necessary to calculate fluxes of OC between model pools. For example, if another value for OC inputs was obtained, the value of parameters governing OC losses would likely have been different after optimisation. We have, however, chosen not to assess this additional uncertainty to limit the complexity of the presented results. Organic N enters the soil through plant litter, assuming a constant C:N ratio of 30, while rhizodeposits are assumed to only contain C. Mineral N is added to the soil through atmospheric N deposition (0.7 g N m^{-2} yr^{-1}).

Line 296: Following on my previous comment, I think the normal situation with real measurements is actually worse than this, because you typically don't have accurate knowledge of total litter inputs. In a steady state system, knowing the inputs is equivalent to having soil heterotrophic respiration measurements, assuming there is no net leaching of DOM. And heterotrophic respiration is difficult to measure accurately if there are living roots. So, if anything this setup might be optimistic compared to a real study where carbon pool data is available.

That is absolutely correct, and we fully agree with this. In our response to your previous comment, we have included references to studies addressing uncertainties in measuring C inputs in the field, and studies that have assessed the effect of this uncertainty on model simulations. We hope that this makes it clear to the reader that the situation is even more complex and uncertain when simulating real soils, compared to the analyses we performed.

Line 361: Could say explicitly here that the identifiable parameters were picked based on the identifiability analysis for each model structure, which is implied but wasn't clear until I looked at the supplementary tables

Thanks for letting us know that this wasn't clear, we have added this information.

finality on model predictions. To do so, each of the rhizosphere models was calibrated twice. In a first scenario (termed the

13

full parameter model; FPM) all parameters for which no realistic estimates could be made or found in the literature were optimised (either 5 or 6 parameters, depending on the model; Table S9). These parameter sets were non-identifiable. In a second scenario (termed the Identifiable parameter model; IPM), only identifiable parameters for the assumed data were optimised: $V_{max,lit}$ and $V_{max,res}$ (see Tables S11, S13, S15 and S17). These parameters were identified using the parameter identifiability analysis for each model separately, as described in Sect. 2.6. In addition to these parameter values, also the rate modifiers for the depolymerisation of POM_{lit} and POM_{res} (see Eq. 1 and 2) were optimised to be as close as possible to 0.5. For both

I had trouble figuring out how the values for the non-optimized parameters were picked, since they had to be fixed but the premise of the study is that the values are not well constrained. I guess the values come from the deterministic parameter calibration? I found that part a bit confusing

Thanks for identifying this imprecise description of this part of the methods. We added a couple of sentences at the end of this paragraph to better explain this, but to also clarify the consequences of having to fix those values.

$V_{max,lit}$ and $V_{max,res}$ (see Tables S11, S13, S15 and S17). These parameters were identified using the parameter identifiability analysis for each model separately, as described in Sect. 2.6. In addition to these parameter values, also the rate modifiers for the depolymerisation of POM_{lit} and POM_{res} (see Eq. 1 and 2) were optimised to be as close as possible to 0.5. For both optimisation scenarios, it was assumed that only measurements of total SOC and the C:N ratio were available, as these data are most commonly available. The values of the parameters that were not optimised were fixed at the values obtained through the deterministic calibration (Sect. 2.5). This means there is an additional hidden uncertainty, as the values of the fixed parameters could not be confidently determined, while choosing different values would likely have led to other calibrated values for the optimised identifiable parameters. This uncertainty is, however, not assessed in the present study.

Line 455: Are the predictions more reliable? Or is the uncertainty underestimated? If the values of the non-optimized parameters are unknown but need to be fixed, this is adding a hidden uncertainty to the model (as mentioned in the Discussion), so I'm not sure it is actually more reliable. In Figure 5a and 5b, there is certainly a narrower distribution of predictions in the IPM approach, but this results from fixing the value of some unknown parameters. Couldn't this approach just as easily lead to a narrow but wrong result? So, perhaps 5b is a more accurate depiction of the actual predictive uncertainty in this situation where data is very limited, unless there are other constraints on the parameter values.

Thanks for pointing this out, we fully agree with this. One of the reasons for not going in more detail about this is not to further complicate the (already long) manuscript. We hope that pointing this out multiple times throughout the manuscript makes it clear to the reader that this is an additional layer of uncertainty. The authors hope to find time and resources to focus on this in a future manuscript, as this would lead to a more comprehensive assessment of model uncertainty caused by equifinality. Now, we have also pointed out this issue at the end of section 2.3.2.

> In addition, similar to the rhizosphere models, they show that the performance of models in steady state (i.e., the behavioural models) is not a sufficient indicator for the performance when making predictions. One has to keep in mind, however, that multiple uncertain model parameters had to be fixed in the IPM models, leading to hidden uncertainty about the accuracy of the simulations.

Line 461: This alone is Insufficient

Thanks for spotting this typo.

Line 463: Did it reduce the uncertainty, or underestimate the uncertainty?

That's a good point, and following the discussions above, the uncertainty is most likely still underestimated. We updated this sentence to make that clear to the reader.

> turnover rates reduced prediction uncertainty; and (4) optimising only identifiable model parameters similarly reduced the uncertainty of predictions, while not eliminating it due to uncertainties caused by the values of the fixed parameter values. The discussion concludes with making recommendations for incorporating parameter identifiability analysis into the model development and evaluation process.

Section 4.2: I think this paragraph makes an important point very clearly

Thanks for this supportive statement

Line 521-522: I really like the "hidden uncertainty" phrasing here, which makes an important point about how to interpret these results

Also here, thank you very much for the positive feedback.

Line 533: Again, is it correct to say that uncertainty is reduced? I'm not sure this section is even making a point about reducing uncertainty. I think it mostly makes a strong argument that identifiability is an important analysis to do. I think part of the challenge here for the community is that identifiability analysis identifies a problem but does not really provide a solution for reducing uncertainty

Thanks for making this fair point. Identifiability analysis by itself indeed doesn't reduce uncertainty, but it makes modelers (and interest experimentalists) aware of which parameters can be jointly optimized, and for which parameters reliable fixed values need to be obtained. Perhaps more importantly, it illustrates the trade-offs between data availability and model complexity, which is especially important for experimentalists. As a result, it allows modelers to assess how much additional uncertainty in predictions is caused by the fact that they are optimizing more parameters than they should, given available data; the latter to be addressed by the experimentalists. From this, it follows that identifiability analysis rather serves to assess and quantify uncertainty. Therefore, the title of this section has been changed accordingly.

4.4 ~~Reducing~~ Assessing uncertainty in model predictions through identifiability analysis

The importance of practical identifiability and equifinality has long been recognised across environmental disciplines that use

Line 567: This is the big challenge, right? Because if the unidentifiable parameter values could be well constrained with observations or experiments they would not need to be optimized in the first place. In that sense, the identifiability issue is the beginning of the conversation, not the end of it

We fully agree, particularly with the last sentence. With your permission, we would like to end the manuscript by quoting this and putting the emphasis on the role experimentalists can or have to play.

procedure when developing and applying SOM models. This will remove the hidden uncertainty in model predictions caused by equifinality, ~~thereby increasing confidence in SOM models~~and support future research into which simulated SOM properties need to be better understood and parameterised because, to quote one of the reviewers of this manuscript, the identifiability issue is the beginning of the conversation, not the end of it.