

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.

90

Line 106: Is there a missing word after "decades"? Maybe "unless $\Delta 14$ data..."

Thanks for pointing out this mistake, we deleted the part about $\Delta^{14}\text{C}$ 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, ~~$\Delta^{14}\text{C}$ data for SOC is included as a calibration constraint~~
110 ~~(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
135 simulations in steady state, every model was calibrated using ~~a Bayesian approach (using the DEzs 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
165 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
175 (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 $\Delta^{14}\text{C}$ 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

195 (Sect. 2.5). This resulted in OC inputs of $349 \text{ g C m}^{-2} \text{ yr}^{-1}$ (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
200 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 ($0.7 \text{ g N m}^{-2} \text{ yr}^{-1}$).

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
380 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
385 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

570 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.

625 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.

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 2

This manuscript addresses a fundamental and timely issue in soil organic matter (SOM) modelling: the mismatch between model complexity and data availability, which leads to parameter non-identifiability, equifinality, and ultimately large uncertainty in model predictions under perturbations. The study is conceptually important and provides a clear demonstration that models calibrated to steady-state observations can produce substantially divergent predictions when subjected to changes in carbon inputs. This has important implications for the reliability of current ecosystem and Earth system models.

We thank the reviewer very much 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.

Below I provide specific comments.

Line 6-8: The statement that the number of identifiable parameters “remained limited to five even under the most data-rich conditions” may be somewhat overstated. The “most data-rich” scenario considered here still represents a relatively limited set of observations, and the result is likely dependent on the specific model structure and data types used. This limitation should be more clearly acknowledged.

Thanks for pointing this out, these specific results are indeed dependent on the model we chose, and the data we considered. To make this clear in the abstract, we changed these sentences as follows.

5 models, the aim of this study is to assess how equifinality affects the variability of predictions made by behavioural soil organic matter models. The results show that ~~the~~ for the model used in this study, the number of identifiable parameters, those that do not compensate for one another, increases with the number of calibration constraints, ~~but~~. However, this number remained limited to five even under the most data-rich ~~conditions~~. ~~Furthermore, the size~~ scenario considered, including the C and N content of particulate organic matter (POM) and mineral-associated organic matter (MAOM), and their ¹⁴C signatures. ~~Furthermore, the~~
10 size of POM and MAOM can only be accurately simulated when data on these pool sizes are used, while the turnover rate of MAOM is reliably simulated only when $\Delta^{14}\text{C}$ data for MAOM are provided. Regardless of the type of mathematical equations

To make it clear that the “most data-rich” scenario applies to our study, we added this sentence to section 2.6, where the data scenarios are described.

that in addition to total C and N, also microbial biomass C and POC were measured. Also for the SOM model, three scenarios
315 of data availability were tested: (1) data on total SOC and N, (2) data on POC, MAOC and their C:N ratios, and (3) data on
POC, MAOC, their C:N ratios, and the $\Delta^{14}\text{C}$ values of POC and MAOC. The choice of these scenarios was based on common
practices in SOM modelling, although data availability may be more extensive in other studies. As the identifiability analysis

Line 35: “Three concepts that are central to achieving this balance are identifiability, equifinality and overparameterisation.” The phrasing “achieving this balance” may be misleading, as identifiability, equifinality and overparameterisation are conceptual tools to describe the problem rather than mechanisms to achieve the balance. This could be clarified.

Thanks for bringing this to our attention. We fully agree with this, and with similar feedback related to this we received from the other reviewer, we changed this part to:

Three concepts that are ~~central to achieving this balance~~ valuable to identify the mismatch between model complexity and data availability are identifiability, equifinality and overparameterisation. The concept of identifiability has a long history in scientific research (Rothenberg, 1971; Reiersøl, 1950), with different aspects of identifiability being defined and studied (Travis and Haddock, 1981; Delforge, 1977; Cobelli et al., 1979; Kleissen et al., 1990; DiStefano and Cobelli, 1980). The two aspects
40 of identifiability that have been most frequently applied to environmental models are structural identifiability and practical

Line 135-141: If these “measurements” are artificially generated, the assumptions should be justified with references.

Thanks for pointing this out. The assumptions about the measurements were made based on meta-analyses of these properties. Now, we made it clear that our aim was not to simulate a soil in a specific location under a specific land use and provided references for studies justifying the numbers.

In this study~~is~~, artificial SOC data were used to calibrate model parameters. We created "measurements" of the total SOC stock and fractions for one point in time, to mimic the common assumption of an SOC stock in steady state. The intention was not to replicate a soil at a specific location under a specific land use, but to use reasonable values measured across different land uses.
145 The total SOC stock down to 0.2 m depth was calculated assuming an SOC concentration of 2 % and a bulk density of 1.2 g
cm⁻³ (Chen et al., 2024; de Brogniez et al., 2015), resulting in an SOC stock of 4,800 g C m⁻² down to 0.2 m. It was assumed
that 25 % of SOC is POM and microbes in the rhizosphere (i.e., 1,200 g C m⁻², Hansen et al. (2024); Lugato et al. (2021)),
which was divided into 96 % POC (1,152 g C m⁻²) and 4 % microbes (48 g C m⁻²). The remaining SOC was assumed to be
~~MAOC (mineral-associated organic carbon (MAOC, 3,600 g C m⁻²)).~~ The standard deviation of the SOC pools was assumed to
be 10 % of their size.

Line 136: “In this study is, artificial SOC data were”. “is” should be removed?

That is correct, thanks a lot for spotting this typo.

Line 140-141: MAOC should be defined at first occurrence in the main text.

Thanks for pointing this out, this has been corrected

Figure 1. Differences between rhizosphere and SOM models are unclear, and Abbreviations (MAOC, DOM, POM, MIC) are not fully defined.

Thanks for making us aware of this, we agree that the differences between both models could be better indicated. Therefore, we now show a separate conceptual figure for both models. We also explain all abbreviations in the figure caption.

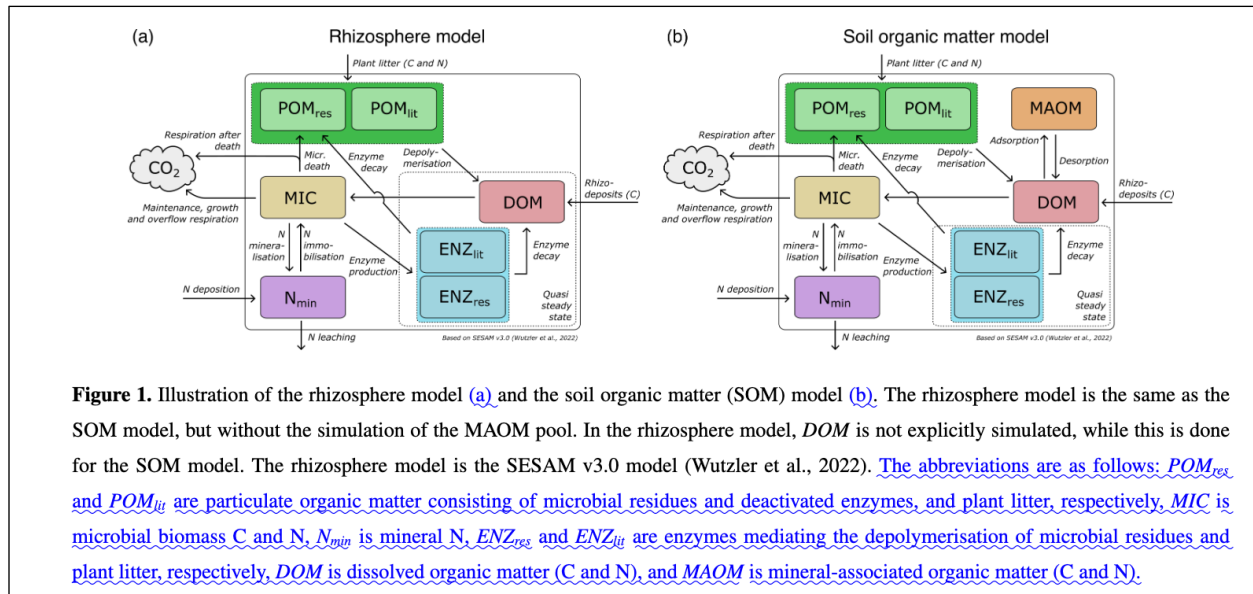
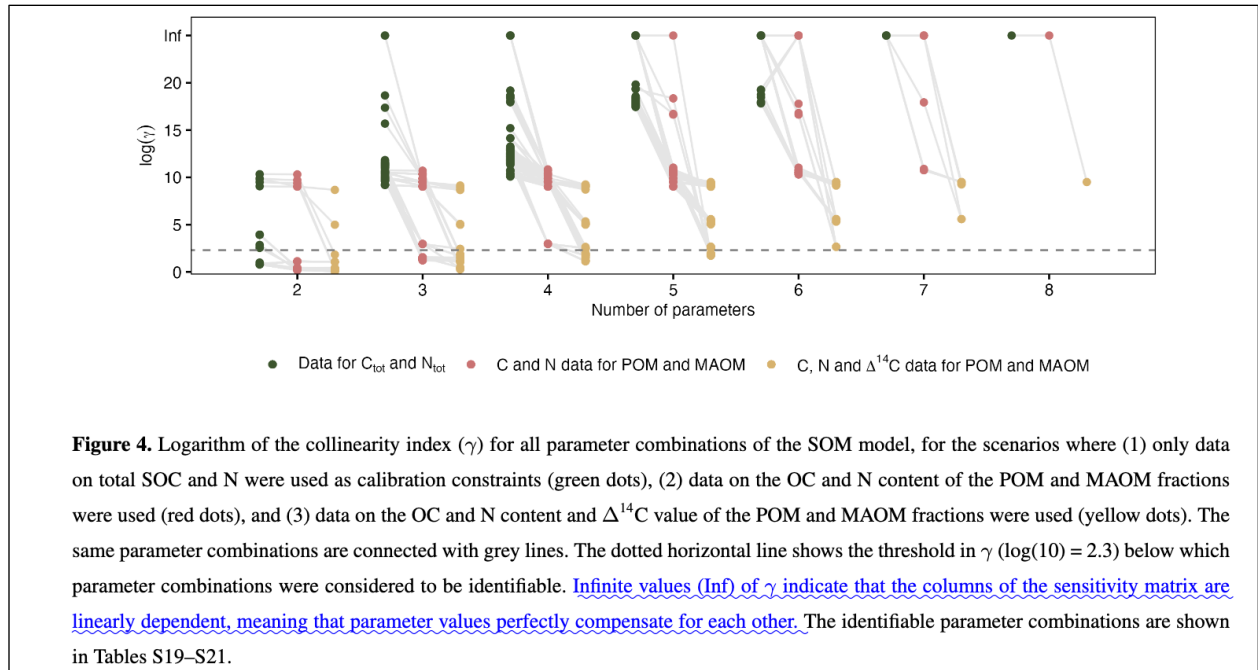
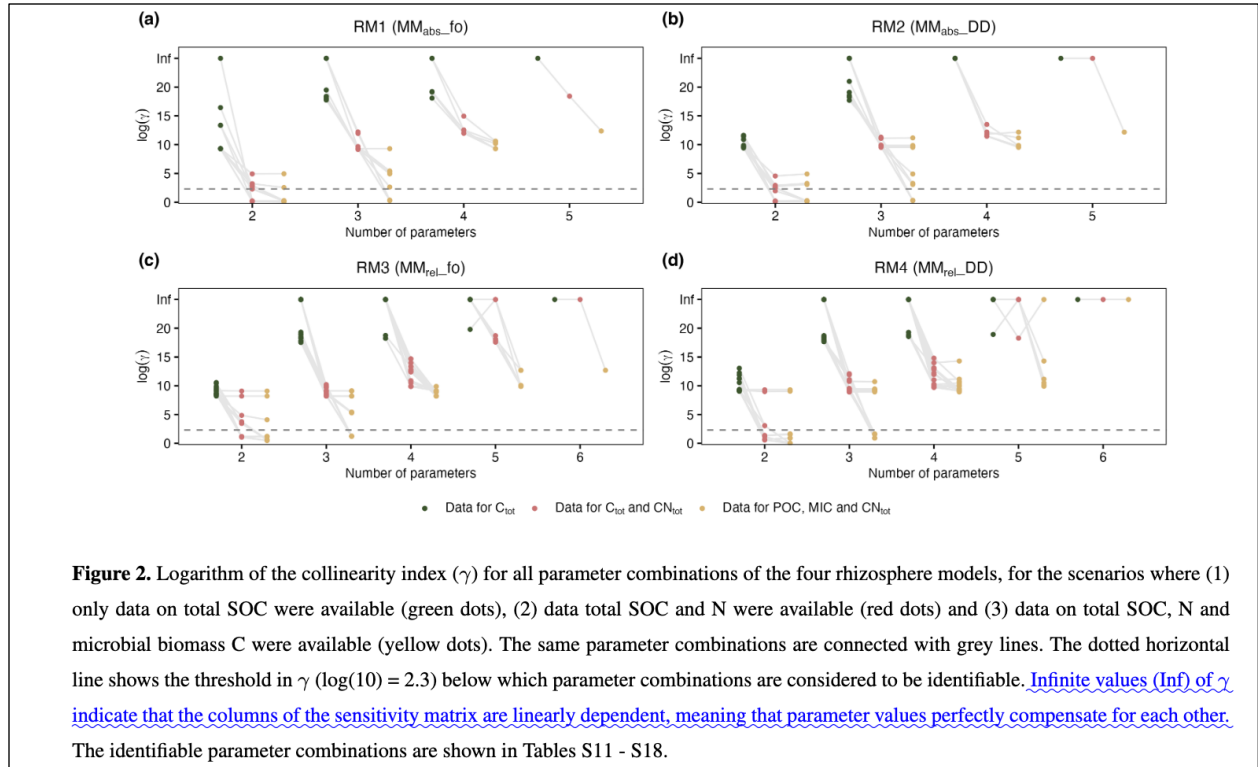


Figure 2: The occurrence of infinite values (“inf”) is not explained. It would be helpful to clarify under which conditions these values arise and how they should be interpreted. In addition, the “inf” values are currently plotted at the boundary of the figure, which is not ideal for interpretation. A clearer representation would improve readability. The same comment applies to Figure 4.

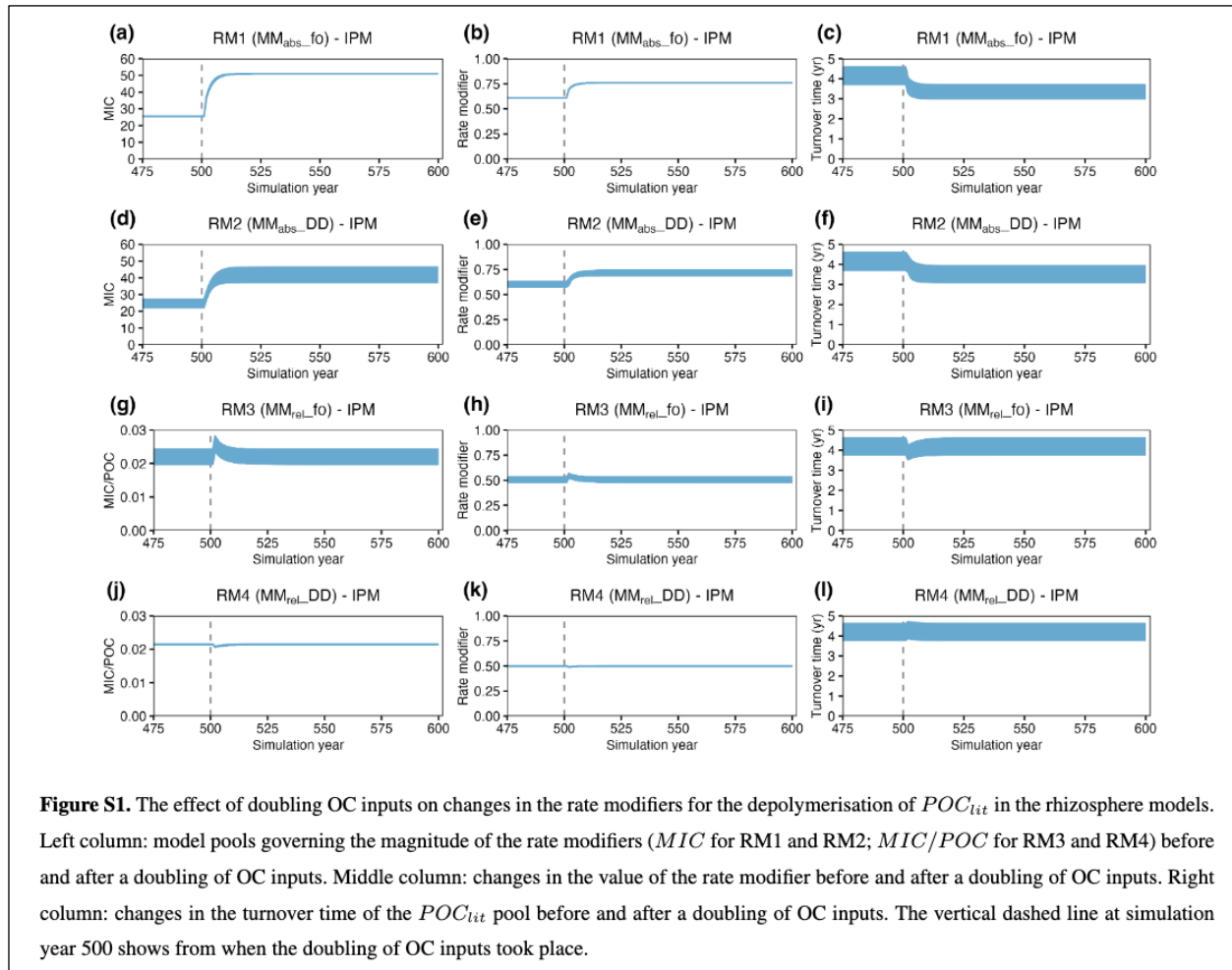
This was indeed not explained in the manuscript, thank you very much for pointing this out. The collinearity index (γ) equals infinity if the columns of the sensitivity matrix are linearly dependent (as explained in the article describing the methods; <https://doi.org/10.18637/jss.v033.i03>, p. 13). This means that their values in the model perfectly compensate for each other. This has now been added to the caption of Figure 2 and 4. In addition, we have put vertical whitespace between the Inf values and the upper border of the graph.



Line 400: “This difference is due to the formulation of the rate modifiers for depolymerisation”. If I understand correctly, the different non-linear equations lead to differences in decomposition rates through the “rate modifier” part. However, these differences ultimately reflect variations in the effective decomposition rates or residence times of carbon in different pools under increased C inputs. While the current explanation

in terms of rate modifiers is mathematically correct, it would be more intuitive to present or discuss the corresponding decomposition rates or residence times of the pools.

Thanks for this suggestion, this will indeed improve the interpretation for the readers. To do so, we have added a graph of the turnover time of the POM_{lit} pool to figure S1.



We also added a brief interpretation of this to the results section.

pool, which doubled upon a doubling of OC inputs for RM1 and increased by 67.2–70.7 % for RM2 (Fig. S1 a,c). This caused an increase in the rate modifier for depolymerisation of POC_{lit} upon a doubling of OC inputs (Fig. S1 b,d), leading to a higher fraction of POC_{lit} being depolymerised compared to the initial OC inputs. [This is also reflected in a decrease in the turnover time of POC after C inputs are doubled, resulting from the faster turnover \(Fig. S1 c,f\).](#) In contrast, the pools controlling the value of the rate modifier for POC_{lit} in RM3 and RM4 (the ratio of MIC to POC_{lit}) remained constant upon a doubling of OC inputs, causing the rate modifier to be identical before and after a doubling of OC inputs (Fig. S1 e–h). As a result,

the same fraction of POC_{lit} was depolymerised per time step upon a doubling of OC inputs, as is also evidenced by the lack of a change in the turnover rates of POC after OC inputs were doubled (Fig. S1, i,l). This difference shows the effect of mathematical formulations on model predictions, using either absolute or relative Michaelis-Menten rate modifiers.

Figure 3: the figure lacks sufficient clarity to distinguish between MIC and full parameter model results. In particular, the large differences in prediction range (42× for RM1 and 7× for RM2) are not readily apparent.

Thanks for pointing this out. In Fig. 3, the evolution of MIC is difficult to see because the value of this pool is much lower compared to POC. We added to the caption that the evolution of the MIC pools is better shown in Fig. S1.

Figure 3. The effect of parameter identifiability and resulting equifinality on the prediction of SOC by the four rhizosphere models (RMs). Each row shows the results for a different model (see Table 1). The first column shows the results from a Bayesian optimisation of two identifiable parameters (the Identifiable Parameter Model: IPM), while the second column shows the results from a Bayesian optimisation of five (RM1 and RM2) or six (RM3 and RM4) parameters (the Full Parameter Model: FPM). Note that the lines for total SOC are not shown, as these overlapped with *POC*. The evolution of the MIC pool for the IPM models is enlarged in Figure S1. The last column shows frequency diagrams of the relative increase in SOC after a doubling of OC inputs after simulation year 500 (as shown by the vertical dashed line). The text reports the median increase, with the range between square brackets. The black dots show the average OC measurement in simulation year 500, and vertical black bars show the standard deviation.

We also clarified to the reader how these differences in prediction range (42 and 5 times) can be seen in the figure.

435 Also the number of optimised parameters had an effect on predictions by the rhizosphere models. For RM1 and RM2 (with absolute Michaelis-Menten kinetics), the average relative increase in OC upon a doubling of inputs was lower for the full parameter models than for the identifiable parameter models (Fig. 3). In contrast, for RM3 and RM4 (with relative Michaelis-Menten kinetics), overparameterisation did not affect the average increase in OC, which was ca. 100 %. Moreover, the range in predictions was ca. 42 times larger for RM1 for the full parameter models (a range of 34 %, i.e., between 34.5 and 68.5 %, Fig. 3c) compared to the identifiable parameter models ~~the~~ (a range of 0.8 %, i.e., between 61.1 and 61.9 %, Fig. 3c). For RM2, the range in predictions was ca. 6 times larger for the full parameter models showed a prediction range approximately 42 times larger for RM1 and 7 times larger for RM2 (a range of 21.2 %, i.e., between 54 and 75.2 %, Fig. 3f) compared to the identifiable parameter models (a range of 3.5 %, i.e., between 67.2 and 70.7 %, Fig. 3f). These results show that the accurate prediction of steady-state stocks of SOC by behavioural models is not a sufficient criterion to evaluate the predictive capabilities of such models.

445

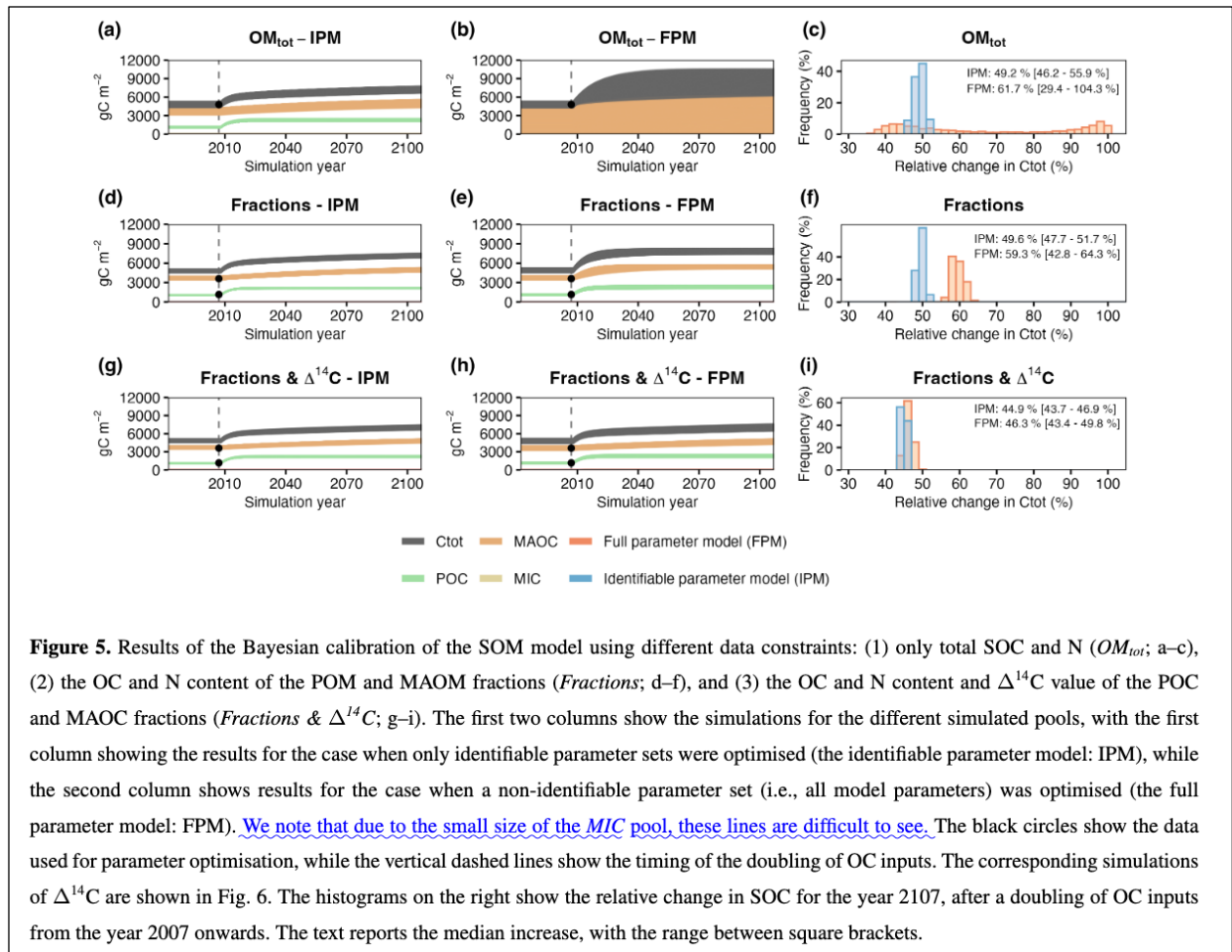
Section 3.2.1: It is unclear which figures support the results described in this section, as no figure is explicitly cited. Given that this is part of the Results section, appropriate figure references should be provided to support the statements. In addition, Figure 4 is not clearly introduced or discussed in the Results section before the manuscript proceeds to Figure 5. This disrupts the logical flow of the results. Please ensure that all figures are properly introduced and described in sequence.

Thanks for pointing out this oversight. This section is based on the results presented in Fig. 4, which is now cited in this section to improve the flow of the manuscript.

The parameter identifiability analysis for the SOM model showed that, similar to the results for the rhizosphere models, the number of identifiable parameters increased with an increasing quantity of calibration data (Fig. 4). When only data on total SOC and N were used, at most two parameters were jointly identifiable, while three parameters were identifiable together when

Figure 5: The color scheme is difficult to interpret, particularly because MAOC and MIC appear very similar, and the meaning of some colors (e.g., dark orange) is unclear. This also makes it difficult to support the statement in Lines 427–429 that SOC can be dominated by either POC or MAOC (Fig. 5b), as the contribution of POC is not clearly visible. Improving the color scheme and legend would enhance clarity and consistency between text and figure.

Thanks for noticing the similar colors between MAOC and MIC in the legend. This was caused by plotting the lines transparent, which also caused the legend colors to be transparent. This has been fixed now, so that it's clear from the legend which colors are for MAOC and MIC. We note that also in this plot, the lines for MIC are very difficult to see because of the small size of this pool. We clarified this in the caption.



Regarding the statement that SOC can be dominated by either POC or MAOC (Fig. 5b), we now made it clear that the lines for POC are covered by the lines for MAOC. This is unfortunately inevitable, as both POC and MAOC compensate for each other, and span the entire range of potential values from a very small pool size to almost all SOC present in these pools.

calibration scenario had a substantial impact on the internal dynamics of the model. For example, when model parameters were optimised using only data on total SOC and N while calibrating all model parameters, the majority of SOC could be either in *POC* or *MAOC* (Fig. 5b, note that the lines for the *POC* pool are covered by the lines for the *MAOC* pool, as both pools span the entire range of potential values from a very small size to almost all SOC present in these pools; their density distribution is shown in Fig. S3), clearly demonstrating equifinality resulting from overparameterisation. Further evidence for

To make this better clear to the reader, now we also show a plot of the density distribution of the amount of POC and MAOC at steady state for this model in the supplement, and refer to this plot in the main text (see textbox above).

