

Referee #1

The manuscript submitted by Fernandez-Catinot et al. combines an incubation experiment with SOC fractionation and regular respiration measurements to test whether 2-pool POC-MAOC models can effectively capture both long-term C pool dynamics and short-term respiration fluxes. They find that a third, fast-cycling compartment is needed to accurately capture both simultaneously. The subsequent simulation study shows that transfer rates into MAOC, rather than litter decomposition rates, are the primary driver of predicted C persistence, and that model structure and parameter changes interact non-additively, producing up to 7-fold differences in estimated transit time. The experiment is well designed, and the results are an important contribution to discussions around SOC modeling methodologies.

While I have no major concerns, there are several issues that I believe should be addressed in revision. Most importantly, the control soil comparison rests on an important assumption about the initial Litter C fraction which is currently untested. I would like to see a sensitivity analysis or a discussion of how this choice affects the results. I also encourage the authors to acknowledge the constraints that a 6-month incubation places on steady-state C persistence estimates, and to consider cross-validating fitted parameters between treatments. I provide a few additional comments below.

Response: We thank the reviewer for the careful reading of our manuscript and for the constructive comments. We have revised the manuscript and believe that the changes have improved its overall quality. In particular, we are grateful for the suggestion to perform a sensitivity analysis to determine the proportion of POC represented as Litter C, as it improved the robustness and precision of our results. Below, we provide a point-by-point response to each comment.

L53: "mainly produced by microbial re-synthesis" — consider softening; direct sorption of plant-derived compounds to mineral surfaces is also an important MAOC formation pathway (e.g., Whalen et al. 2022)

The correction has been made in the manuscript.

L72: "despite its components might cycle" — grammar; should be "despite the fact that its components might cycle" or similar

The correction has been made in the manuscript.

L90 (and throughout): When referring to specific numerical comparisons, "mean transit time" and "mean C age" would be more precise, since the full metrics are distributions. This applies here and in several other places (e.g., L28, L35, L334, L353, L397).

The corrections have been made in the manuscript.

L125-132: Please report mass recovery and C recovery for fractions derived from the wet sieving fractionation.

We added the mass recovery and the C recovery from the wet sieving fractionation to the manuscript.

L152-154: It is unclear whether the a_{ij} values reported in Table 2 are actually absolute transfer rates (yr⁻¹) as the footnote states. All reported values fall between 0 and 1, and under the absolute rate interpretation some parameter sets would violate the non-negative respiration constraint (e.g., control 3-pool: $k_2 = 0.121$ but $a_{32} = 0.968$), suggesting the values may be proportions. Please check and correct as needed.

Thank you very much for pointing out this error. In the original version of Table 2, we inadvertently reported the **proportion** of the decomposition rate that is transferred from pool j to pool i rather than the absolute transfer rate a_{ij} yr⁻¹. We have corrected the table so that the reported values now correspond to the actual absolute transfer rates, calculated as the product of the decomposition rate and the corresponding transfer proportion.

L156-160: For control soils, the 3-pool model assigns 10% of POC to a fast-cycling Litter C pool, in line with the author's argument that the POC pool is heterogeneous. However, for the litter-addition treatment, the Litter C pool appears to contain only the added litter, with the native fast-cycling fraction folded back into POC. If POC heterogeneity is a general property of these soils, the litter-addition model's Litter C pool should include both added and native fast-cycling material. I recommend the authors should clarify this choice and discuss whether it affects the fitted parameters or model performance.

We sincerely thank the reviewer for this and the following comment. The reviewer points out a straightforward issue regarding the modelling assumptions that we had overlooked. We addressed these comments, which led to more robust and precise results.

Originally, the decision to define the Litter C pool to contain only the added litter was based on the fact that assigning a proportion of the POC pool to a fast-cycling C pool requires an additional assumption. Because the litter-added C represented a clearly defined C pool, we considered it more parsimonious to model the added litter as the fast-cycling C pool. However, as the reviewer correctly points out, we overlooked that, as in the control samples, the litter-addition samples should also include native fast-cycling C.

Therefore, after both “L156-160” and “L160” comments, we explored models assigning a proportion of POC as Litter C. We evaluated the models performance across a wide range of proportions of POC as Litter C and we report the results below.

We should clarify that, in this new optimization process, we implemented an algorithm that allowed us to identify parameter sets with lower AIC and MSE values than those reported in the original manuscript. Hence, the fitted parameters and model performance changed. However, as shown below, these changes do not affect the patterns observed or the conclusions discussed in the manuscript.

Assuming different proportions of POC as a fast-cycling C pool for litter-addition samples affected both model performance and the fitted parameters (Fig. 1, Table 1 and Table 2). Interestingly, as observed for the control samples, assigning a fraction of POC to a fast-cycling C pool improved the model performance. Specifically, assuming 30% of POC as a Litter C pool yielded the lowest AIC and MSE values. However, these results had relatively small differences compared to other model assumptions: For example, assumptions of 25%, 20%, and 15%

resulted in AIC values that were 1.6%, 1.7%, and 4.3% higher, respectively, than the value obtained for the 30% assumption, whereas for MSE, these values were 4.6%, 4.8% and 13% higher, respectively (Table 1).

Given these similar performances, we then examined the fitted parameter values. We found that the 30%, 25%, and 20% assumptions, despite yielding low AIC and MSE values, produced some unrealistic parameter estimates, with annual decomposition rates on the order of 10^{-4} to 10^{-6} (Table 2). Therefore, we selected an assumption of 15% of POC as Litter C, as it provided biologically realistic parameter values while maintaining relatively low AIC and MSE values. Furthermore, as shown below, this decision is consistent with the results obtained for the control samples, where exploring different proportions of POC assigned to the Litter C pool also led us to select the 15% assumption. We replaced the original results in the manuscript with those obtained under this assumption.

Fig. 1. AIC and MSE values for litter-addition-samples models assuming different proportions of POC assigned to the Litter C pool.

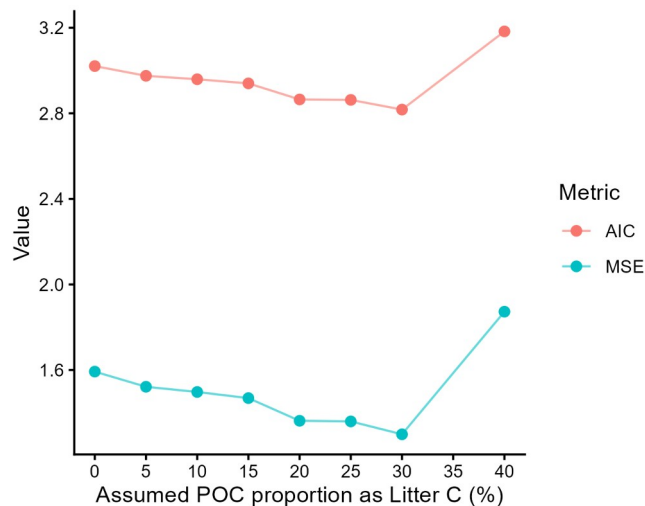


Table 1. AIC and MSE values and their relative values compared with the model showing the lowest AIC and MSE values, for litter-addition samples models assuming different proportions of POC assigned to the Litter C pool.

POC as Litter C assumption (%)	AIC	Relative AIC to best model (%)	MSE	Relative MSE to best model (%)
0	3.021	107.2	1.592	122.5
5	2.975	105.6	1.522	117.1
10	2.959	105.0	1.497	115.2
15	2.940	104.3	1.469	113.0
20	2.865	101.7	1.363	104.8
25	2.863	101.6	1.360	104.6
30	2.817	100	1.299 *	100.0
40	3.183	113.0	1.873	144.1

Table 2. Parameter values for litter-addition-samples models assuming different proportions of POC assigned to the Litter C pool.

POC as Litter C assumption (%)	k_1	k_2	k_3	a_{21}	a_{32}	a_{31}
0	6.25	4.5 E-15	3.4 E-11	0.027	1.7 E-16	0.049
5	6.16	0.002	1.2 E-7	0.284	0.002	0.275
10	6.32	0.064	0.028	1.28	0.064	0.080
15	6.81	0.018	0.046	1.850	0.017	0.404
20	6.46	5.7 E-5	0.027	2.010	5.6 E-5	0.158
25	6.10	0.015	4.1 E-4	1.745	0.014	0.271
30	6.04	6.69 E-4	1.1 E-6	2.160	6.6 E-4	0.065
40	6.04	0.003	0.003	0.035	0.003	2.614

L160: Additionally, for control soils, the 3-pool model requires assigning a fraction of initial POC to the Litter C pool. In the manuscript, the authors choose an arbitrary value of 10% (L160) but do not report how sensitive the results are to this choice. Since the litter-addition result is acknowledged as "almost self-evident" (L321), the control soil comparison is the stronger, non-trivial test of the 3-pool model's advantage, however, it depends in large part on this assumption. I recommend that the authors either (a) report the sensitivity of the 3-pool model's advantage/parameterization to changes in the Litter C fractions (e.g., 5-20%), or (b) attempt to fit the initial Litter C fraction as an additional parameter in the inverse optimization (if identifiability permits).

Following up on the previous comment, the reviewer points out an important issue regarding the modelling assumptions that we had overlooked. Again, we sincerely thank the reviewer for this comment, which led to more precise results.

We evaluated the model performance across a wide range of assumptions regarding the proportions of POC as Litter C, and we report the results below. Once again, we found that different assumptions resulted in different model performance and fitted parameter values (Fig. 2, Table 3 and Table 4). Our original assumption of assigning 10% of POC to the Litter C pool did not yield the best model performance. Rather, the 20% assumption produced the lowest AIC and MSE values. However, these values differed only slightly from those obtained with the 15% assumption, which resulted in AIC and MSE values that were 8.4% and 10.7% higher, respectively (Table 3).

Given these similar performances, we then examined the fitted parameter values, and the 20% assumption produced an unrealistic yearly POC decomposition rate (8.06×10^{-4}). Therefore, and consistent with the litter-addition samples, we selected an assumption of 15% of POC as Litter C due to its biologically realistic parameter values and relatively low AIC and MSE values. We replaced the original results in the manuscript with those obtained under this assumption.

Fig. 2. AIC and MSE values for control-samples models assuming different proportions of POC assigned to the Litter C pool.

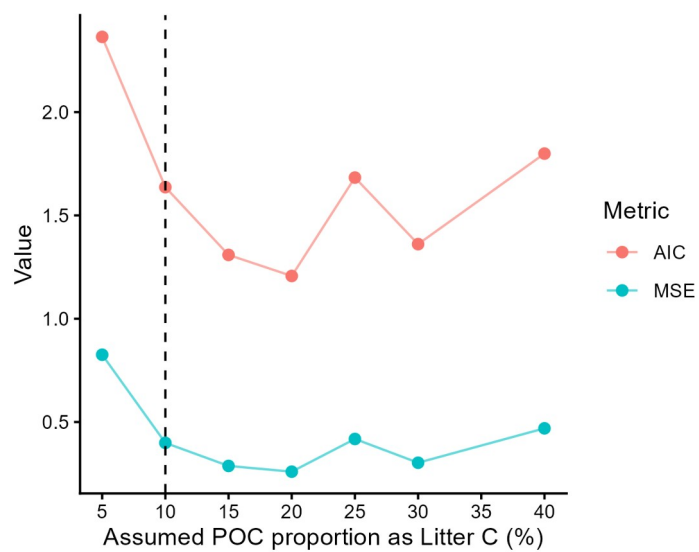


Table 3. AIC and MSE values and their relative values compared with the model showing the lowest AIC and MSE values, for control-samples models assuming different proportions of POC assigned to the Litter C pool.

POC as fast-cycling C assumption (%)	AIC	Relative AIC to best model (%)	MSE	Relative MSE to best model (%)
5	2.364	195.9	0.826	318.1
10	1.637	135.6	0.399	153.7
15	1.309	108.4	0.287	110.7
20	1.207	100.0	0.260	100.0
25	1.683	139.5	0.418	161.0
30	1.361	112.8	0.303	116.7
40	1.799	149.1	0.469	180.8

Table 4. Parameter values for control-samples models assuming different proportions of POC assigned to the Litter C pool.

Litter addition assumption (%)	k_1	k_2	k_3	a_{21}	a_{32}	a_{31}
5	13.622	0.05	0.131	0.101	0.045	0.105
10	8.152	0.15	0.086	0.411	0.134	0.123
15	5.896	0.070	0.060	0.257	0.069	0.195
20	4.701	8.06E-07	0.027	0.087	7.93E-07	0.078
25	4.165	2.78E-04	0.004	0.067	2.72E-04	0.209
30	5.170	0.19	0.023	1.678	0.137	0.533
40	5.989	0.06	0.056	0.194	0.047	3.728

L260-267: Given that the C age and transit time estimates are derived from steady-state equations applied to parameters fitted from a 6-month incubation, over which the MAOC pool barely changes, I think the k_{MAOC} may be poorly constrained. Since the steady-state persistence metrics shown in Figure 4 are sensitive to the parameters of the slowest pool, I am concerned that the C age and transit time estimates may carry important uncertainty. I'd recommend the authors report confidence intervals on the fitted parameters and derived metrics to help address this concern. Additionally, the input vector u used to calculate transit time (Eq. 7) should be specified, given that $I = 0$ during the incubation (L145).

It is important to clarify that while the incubation experiment is a transient response to an experimental manipulation, the environmental conditions were kept constant during the entire duration of the experiment. It is under these constant environmental conditions that we derived the parameters for the model. The assumption of steady-state for predicting ages and transit times also implies constant rates, and therefore our age and transit time estimates can be interpreted as the long-term characteristics of a soil where input and decomposition rates are held constant at the same level as in the incubation experiment. In other words, the age and transit time distributions are not representative of the actual incubation, but of the expected long-term behavior if the same constant environment is maintained.

For the derivation of the age and transit time distribution, we assumed a series model structure where all inputs enter the fast cycling pools. This implies that the vector u has as first element the value of 1, and zero for the other elements. This can be deduced from equations 8 and 9.

L342: Also see Curtin et al. 2019 ([https://doi.org/10.1016/S1002-0160\(18\)60049-9](https://doi.org/10.1016/S1002-0160(18)60049-9)) on POM as an organo-mineral composite with physically distinct sub-fractions.

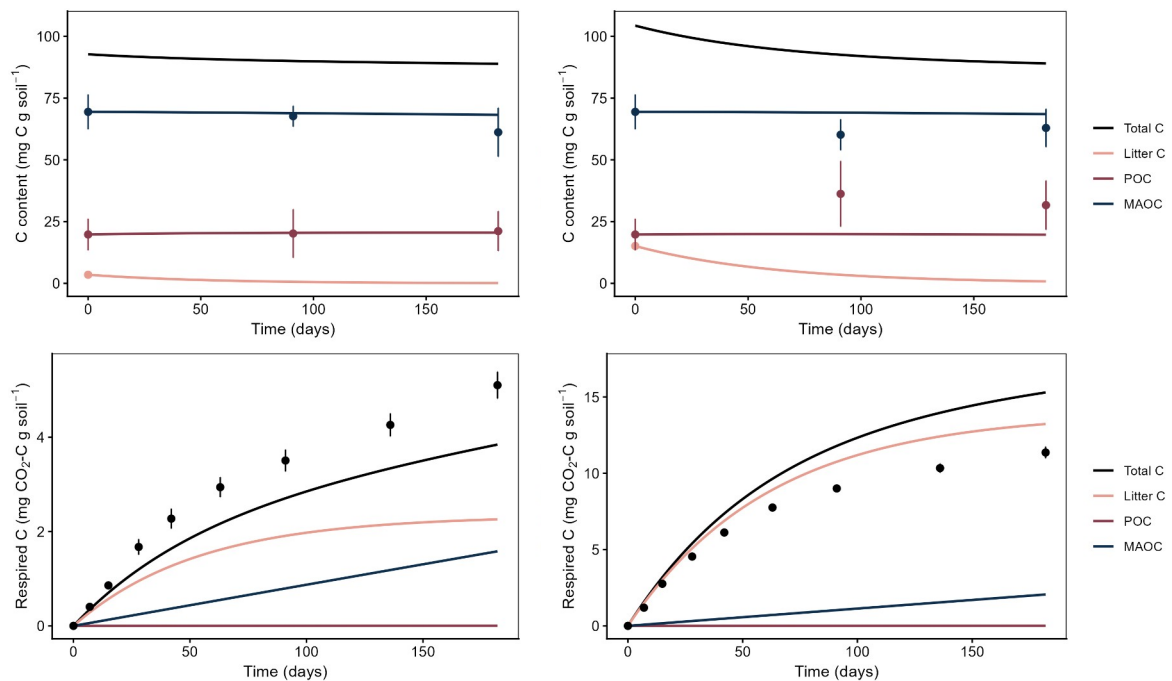
L345-351: The Litter C pool in control soils is a modeling construct that cannot be independently measured through the fractionation used here — chopped litter and native POC both land in the $>53\ \mu\text{m}$ fraction. I think this is worth acknowledging, as it means the 3-pool model's advantage for control soils cannot be validated against measured pool sizes at any time point, similar to the pools of classic process based models like CENTURY.

It is true that the litter C pool in the control treatment cannot be measured directly, but our results show that they are not a modeling construct. The fast respiratory response observed in the incubation experiment can only be explained by a fast response pool that would differ from the measurable POC pool. This is one of the main points that we would like to highlight in this article, that even though two measurable fractions are used to characterize SOM, they do not account for fast respiratory responses, and only a fast non-measurable fraction can explain these short timescale responses. Models help us to determine the size of this fast response fraction, and it is not useful to regard them as modeling constructs given that there is good empirical evidence that demonstrate their existence.

Section 4.1: Since the 2- and 3-pool models are fitted independently to each treatment, I'm curious whether the 3-pool parameters fitted to the litter-addition treatment would also predict the control soil dynamics (or vice versa), with only the Litter C initial condition changing. If so, that would be strong evidence that the 3-pool structure captures a real property of these soils rather than reflecting additional fitting flexibility. This may be beyond the scope of the current revision, but it could act as a useful cross validation and I encourage the authors to consider it.

We agree that litter-addition parameters fitted to control samples data and vice versa may represent a useful cross-validation for the 3-pool model structure, providing additional evidence that the models capture intrinsic properties of these soil C dynamics. However, we believe that this is beyond the scope of the current study. These analyses would considerably increase the amount of results to present, making the presentation of our main results less tractable. Nevertheless, after the reviewer comment, we explored this possibility, and we did not find a good fitting when interchanging parameters between control and litter-addition soils, which is shown below (Fig. 3). We observed the differences mainly for the respiration fluxes. We believe that this is because a litter amendment to the soil may trigger processes that result in different C dynamics compared to soils that did not receive litter. Hence, one set of parameters do not necessarily fit the other dataset adequately. For example, litter-addition soils yielded faster litter C decomposition rates (k_1 , Table 2), but also a higher transference from Litter to POC and MAOC (a_{21} , a_{31} , Table 2), which resulted in poor predictions for respiration fluxes.

Fig. 3. Left panels: Control soils fitted with litter-addition parameters. Right panels: Litter-addition soils fitted with control parameters (note that the y-axis have different scales).



Referee #2

This is a well executed study combining an incubation experiment with inverse modeling to compare 2 pool and 3 pool SOC models using both carbon fractions and respiration data. The dataset is useful and the modeling approach is standard and sound. I do not see fundamental technical errors. However, I have a major concern with how the results are framed and interpreted.

The paper is motivated by the idea that POC and MAOC are chemically and functionally diverse and that current models may be too simple. This sets up a fairly deep problem about process representation. However, what is actually tested is much narrower. The study compares a 2 pool model with a 3 pool model and shows that the 3 pool model fits both respiration and pool data better. This result is expected because the 3 pool model has more flexibility. The manuscript then interprets this as evidence that POC needs to be split into multiple pools and that SOC dynamics require multiple timescales. These conclusions are not really new. They follow from the model structure rather than from independent evidence in the data. There is therefore a gap between the problem as framed and the result that is actually shown.

We thank the reviewer for the critical comments and thoughtful interpretation of our results. However, we would like to stress that our main aim is not to show that three-pool models are better to represent SOC dynamics. This is not new and a large number of previous studies have shown that 3-pool models perform well in predicting different types of experimental and observational data. Our aim is rather to show that the currently popular conceptual framework of characterizing SOC into POC and MAOC is too simple to explain results from measurements and experiments operating at different timescales. In other words, our critique is not about the mathematical models, but rather to the conceptual simplification of the POC-MAOC framework for characterizing SOC dynamics.

Nevertheless, we believe that we can make our point better by bringing up this point earlier in the article, and not just in the discussion as in the previous version. In the introduction of the revised version, we frame the problem in terms of the current POC-MAOC framework and how it simplifies the dynamics of SOC to two timescales, while multiple timescales of responses are often observed in experimental studies. We also modified the title, to make clear that the point is not about the use of 2 or 3-pool models, but rather on the limitations of the POC-MAOC framework to address multiple timescales. We also changed parts of the abstract, discussion and conclusion. We hope that these changes more clearly highlight the intention of our study. However, if the reviewer feels that further revisions are needed, we would be happy to continue refining the manuscript accordingly.

The initial motivation around chemical diversity of POC and MAOC is also not carried through. In the discussion, the focus shifts to timescales and model structure, not chemistry or mechanisms. There is no clear link between chemical heterogeneity and the added pool in the model. If the problem is truly chemical diversity, there are already more detailed models in the literature that address this, but the paper does not engage with them. This makes the overall story feel inconsistent. I would strongly encourage the authors to reposition the paper in the literature. Many SOC models already represent multiple pools and multiple timescales (see refs below). The paper would benefit from stating clearly that the contribution here is not the idea of multiple timescales itself, but the use of simultaneous pool and respiration constraints to test a minimal POC-MAOC structure. That is the part that feels most defensible.

As mentioned in the previous answer, our intention is not to demonstrate that three-pool models are better. The reviewer is right in that there are multiple three-pool models in the literature and this has been known for some time. Although we do not go into detail regarding the main chemical and biophysical mechanisms around the POC-MAOC characterization, we emphasize on the topic of timescales because they encompass the large diversity of mechanisms that may operate simultaneously and that are difficult to disentangle in empirical studies. While we may not be able to study all chemical and biological transformations of SOC in detail, we are able to observe distinct timescales of system response to experimental manipulations. These timescales are useful for representing aggregated processes in models, and they should be consistent with conceptualizations of SOC in overall frameworks such as the POC-MAOC framework.

The discussion should better distinguish what comes from the data and what comes from the model assumptions. The age and transit time analysis is mathematically correct but does not add new insight. The results mainly reflect the chosen model structure. Adding a fast pool will always lead to younger carbon and shorter transit times, so this part reinforces the model assumptions rather than providing independent evidence.

The discussion does make a distinction between the experimental results and those obtained from the models, and we try to improve this further in the new version. However, we disagree with the statement from the reviewer that ‘adding a fast pool will always lead to younger carbon and shorter transit time’. The outcomes can be more complex than this. For example, if the fast pool doesn’t have a release flux but has a transfer flux to slow cycling pools, the transit times would actually increase. The final outcome depends on the parameters of the model, which are informed by the obtained data. For this reason, it is not completely possible to separate experimental from modeling results because our models are informed by the experimental data.

Overall, we believe our discussion on transit times is important because these transit times are aggregated metrics that inform us about holistic system behavior independent on the actual structure of the model. For example, if the value of the decomposition rate of the slow pool changes from a 2-pool to a 3-pool structure, the change in dynamics can be composed by changes in other parameters of the models. Therefore, one needs one single aggregated metric for the comparisons and to make inferences on system behavior. It is for this reason that we use transit time as this aggregated metric that we can use across different model structures.

Please moderate the broader conceptual claims in the conclusion. It is reasonable to say that SOC models may need to represent processes operating across multiple timescales. It is not reasonable to imply that this study establishes a broadly new conceptual view of SOC dynamics. The data do not go that far.

We agree that this study alone does not establish a new conceptual view of SOC. However it serves as an additional evidence that points out problems with the current view of SOC as composed exclusively on POC and MAOC. We hope that our study helps to improve over this current conceptual framework, and eventually leads to a new improved framework based on the main characteristic timescales of SOC dynamics.