

Reviewer 1:

The paper by Laub et al presents a newly developed model called SAMM. The model aims at representing the effect of aggregation on soil organic matter (SOM) dynamic. The paper is well written and the subject fits perfectly with the GMD scopes. I appreciated the approach proposed by the authors and in particular the comparison between SAMMnoagg and SAMM. It is fair to recognize that SAMMnoagg performed as well as SAMM when properly calibrated. I think this paper deserves publication after some corrections:

We thank the reviewer for the positive assessment and for the valuable suggestions to improve the manuscript.

1. The authors claimed that this is the first model considering aggregation. It is not totally true, the MIMICS model published by Wieder et al., (2014) consider a physically protected pool that could be similar to the Agg pools presented here though not as detailed as what presented by the authors.

We agree with this. Therefore, we refined the statement that "SAMM is the first model that demonstrated this capability in a field experiment with different litter qualities".

2. The authors evaluated the SAMM model using microbial biomass obtained after fumigation extraction. This is problematic because the chloroform extraction method extracts the full biomass including the dormant one and, in your model, you represent the active one. You can't directly compare both since 90 % of the biomass is dormant (Lennon and Jones, 2011).

While we agree that it would be best to distinguish between active and dormant microbial biomass, this type of data was not (and is generally mostly not) available for the long-term trials and efficient measurability of all major pools was our credo. Further, SAMM links the microbial aggregate formation to microbial growth and not to microbial biomass itself (see Eqs. A17, A19 and A24-25). Therefore, absolute values of microbial biomass (estimated by CFE) should not strongly affect the simulation of aggregate dynamics. Further, the inclusion of a temperature and moisture scalar (s_t and s_w) should be linked to that factors that control microbial dormancy. We think that in the absence of measured dormancy, this is the best alternative simulating the dynamics of active and dormant fractions of microbial biomass. For these reasons, we think that CFE measured microbial biomass is sufficient for the main purpose of this paper and model, i.e. to show that microbial dynamics play a significant role in aggregate formation.

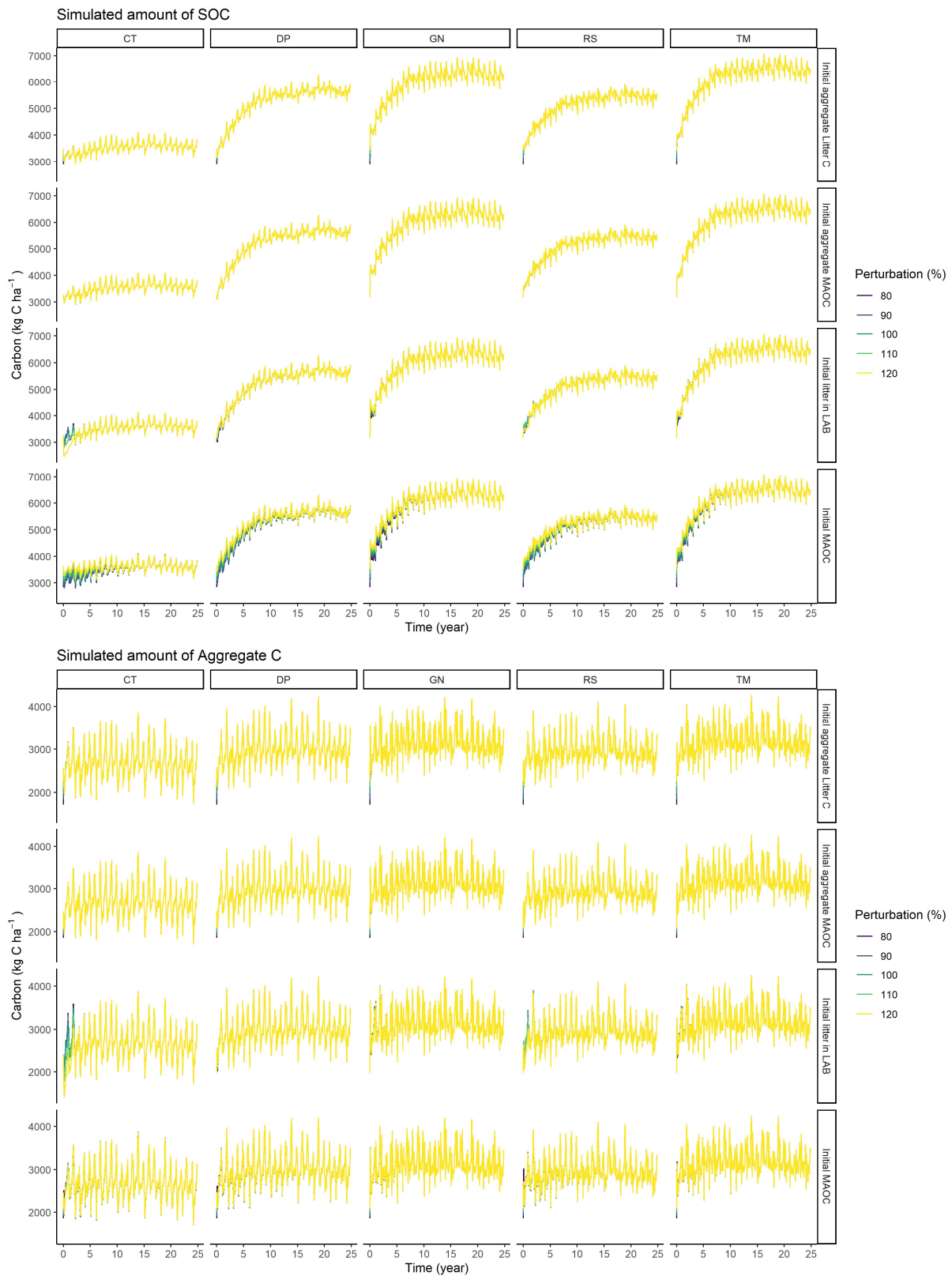
3. The initialization procedure of the model is not detailed enough, does the simulation showed here started after a spinup? What are the consequences of the initialization procedure on the results?

To initialize the pools, we used the mean measured fractions of SOC in the rice straw treatment, which had not experienced changes in SOC over the time of the experiment. In the absence of data on historic plant input quantities and qualities, this was considered the best option. Ideally, SOC fractions at the start of any experiment would be measured.

We did a small sensitivity analysis on the influence of the initialization assumptions. For this, we perturbed the following initialization assumptions. 1) the fraction of initial litter C that was labile, 2) how much of the initial SOC was MAOC, 3) how much of initial litter was protected in aggregates, and

4) how much of the initial MAOC was protected in aggregates (See top to lowest panel, below). We assessed the effect on SOC and on simulated C in aggregates. The perturbation was from 80% to 120% of the initial values. As can be seen, the effects are small for SOC and disappear completely after the first 5 to 10 years, and for simulated C in aggregates, they disappear after less than 3 years.

We added this explanation to the main text.



4. I don't understand what is the rationale behind the ProtLAB pools, how the presence of structural litter can protect the labile pool. It needs to be more justified.

Thanks for the hint. We realized that it was not mentioned in the main text. We added a short description in the main text and refined the description in the appendix. In short, this was a way for us to simulate soluble, non-lignin cell-wall and lignin cell-wall components with three instead of two pools.

5. It is not clear what the time step of the model is, please clarify.

Thanks for pointing this out. We added the clarification that the model is coded in R and run with the deSolve package. It can thus be run at any time step. We used a daily time step with the optimized rk4() solver, after confirming that the results for this were the same than using an ode() solver, which makes time steps infinitely small and has no numerical errors.

6. In the main text, the information on the boundary's conditions is not clear. A sentence refereeing the appendix would help the reader to find the information.

We were not 100% sure what boundary's conditions you referred to. We interpreted this statement in the context of the upper and lower boundary for MIC pool's C:N ratio. We thus added the following sentence "Further, MICC&N can immobilize or release N, to maintain their C:N ratio (see appendix A1.4)."

7. Since the model is newly developed a mass balance calculation showing that the mass balance is closed is necessary to trust the model behavior.

We agree and in fact we have included this, we just did not mention it in the text (See the model code that we published on Zenodo contains a mass balance equation, causing the model to stop if the mass balance is not closed). We now made this explicit "we added a mass balance equation to stop the model with an error message, if the mass balance is not closed."

8. From Fig. 7 it is not clear whether the prediction of SAMMnoAgg recalibrated are different from SAMM. It might be interesting to test through a statistical analysis if the two models give predictions that are significantly different.

As you correctly pointed out, it cannot be seen in Fig. 7, but the evaluations statistics of Table 5 show that overall SAMM performs significantly better than SAMMnoAgg for the joint evaluation of litterbag C, microbial N and SOC. For SOC, there is almost no difference, though.

9. L46-49. You should write "One of the important processes..." not only CUE matters

We agree that CUE is one of the most important processes and therefore added "one of the key factors of SOC stabilization (Cotrufo et al., 2013)." to the sentence.

10. L95: The data should be show in supp mat because you may have no significant changes for 2 mains reasons :1. There is indeed no or a very limited effect or 2. The variance between plots is so high that the statistical power of your setup is not strong enough to detect any change.

Thanks for this suggestion. We now display the bulk density data for the years where it was available in the supplement (see below). The only effect was difference between different sampling years. The most realistic explanation for this was that differences were caused by varying personal conducting the sampling. Given the high sand content of the site, we are also not surprised to only find a limited effect on soil structure. Since we used a mixed model with a nested random effect, random variability in the field based on position should be accounted for and we think the reason is the first that you specified.

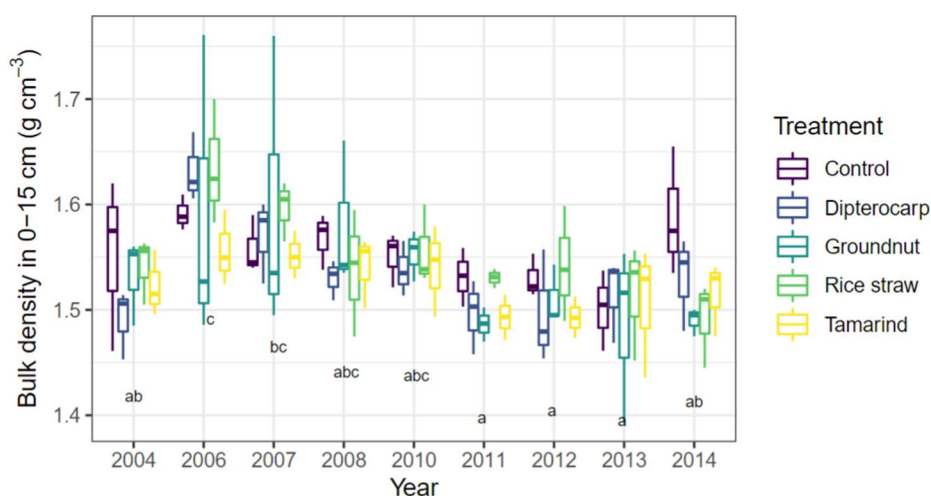


Figure A5. Comparison of measured bulk densities in 0-15 cm in the years with available data. Treatment differences were not significant but a significant effect of year existed. This was however not considered to be any temporal trend but rather an effect arising from different people conducting the sampling. All statistical test conducted with a mixed linear effects model, containing a random intercept per subplot nested in the experimental block.

11. L307: This comparison is not totally fair because you are comparing with 1st order kinetics models, you should compare with Millennial and MIMICS.

We agree that this would be better, but we did not find any info on the parameter correlation of these models. We thus added "It could, however, also be due to the superiority of Michaelis Menten to first order kinetics."

References cited

Lennon, J. T. and Jones, S. E.: Microbial seed banks: the ecological and evolutionary implications of dormancy., *Nat. Rev. Microbiol.*, 9, 119–130, <https://doi.org/10.1038/nrmicro2504>, 2011.

Wieder, W. R., Grandy, A. S., Kallenbach, C. M., and Bonan, G. B.: Integrating microbial physiology and physio-chemical principles in soils with the MIcrobial-MIneral Carbon Stabilization (MIMICS) model, *Biogeosciences*, 11, 3899–3917, <https://doi.org/10.5194/bg-11-3899-2014>, 2014.

Reviewer 2:

The manuscript by Laub et al. presents a new, advanced soil carbon dynamics model featuring a measurable-pools structure, which includes an explicit aggregate formation process and its connection with microbial growth. The model parameters were calibrated against measurements in a long-term experiment at a tropical site, showing low parameter correlations that indicate a parsimonious model structure. Their model results could reasonably reproduce the observed microbial biomass and soil carbon changes after litter addition, and highlighted the role of aggregate protection which accounted for about half of soil carbon stabilization at the tested site. Overall, the manuscript is well written and logically organized. Model limitations are also well discussed. I have only some minor comments that need to be addressed/clarified.

Thanks a lot for this overall very positive assessment.

For the Bayesian calibration of the parameters, it is not clear what data from the observations were used for the optimization. Do you use all the observations including the time series of carbon changes for different pools after each litter addition? If so, the model evaluation metrics actually represent the potentially highest level that the model can reach, which would be expected to degrade when applied to other sites.

Thanks for pointing this out. We now specify that “To calibrate SAMM and SAMMnoAgg, we used all available data of litterbag C, microbial N, SOC, while data of aggregate C and free MAOC was only used to calibrate SAMM.” We further added the following sentence to the discussion “It is likely that across a range of sites, SAMM model performance will be lower, and that the calibration to the single site of this study may have resulted in an overfitting of some parameters.”

It is not clear how the initial state of the model was derived. Was a spin-up process employed to reach equilibrium, or were initial values prescribed for each pool?

To initialize the pools, we used the mean measured fractions of SOC in the rice straw treatment, which had not experienced changes in SOC over the time of the experiment. In the absence of data on historic plant input quantities and qualities, this was considered the best option. Ideally, SOC fractions at the start of any experiment would be measured.

Line 222: Is there an explanation for the 1~2 months delay in the peak of MICc compared to the peak of LMWc?

Most likely this is related to a slower death rate of microbes than uptake of LMWc.

Figure 2: It would be helpful to add a panel showing changes of the total SOC. Besides, line colors for the different pools are a bit difficult to distinguish, please consider using more distinct colors. “STRUc” in the legend should be “STRc”.

Thanks - we have included SOC into this figure, now and changed to STRc. The choice of the color scheme (Magma of viridis package; <https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html>) was a decision to adhere to the colorblind-friendly requirement of GMD. It is

purposefully aligned with the colors of Figure 2 and, based on different options we compared, the best option to display that many pools in a colorblind-friendly way.

There are a few typos in the current manuscript, such as “MAOc” being written as “MOAc” in some places, “depolimerization”, “One the one hand”. Please check carefully throughout the text.

Thanks for the hint – we have corrected the mentioned mistakes and went carefully through the whole manuscript again with this in mind.

The current abstract is not a very concise and engaging summary of the study, please refine it.

We rewrote major parts of the abstract with the goal of refining it. We hope that you consider the current form of the abstract (see below) to be more concise and engaging. We also added the track-changed version.

Maintaining soil organic matter (SOM) is crucial for healthy and productive agricultural soils and requires understanding at the process level, including the role of SOM protection by soil aggregates and the connection between microbial growth and aggregate formation. We developed the Soil Aggregation through Microbial Mediation (SAMM) model, to represent this important connection. The pools of SAMM are fully measurable and we calibrated and evaluated it against data from a long term bare-fallow experiment in a tropical sandy soil. This experiment received plant litter additions of different compositions, leading to twice the soil carbon stocks in the best treatment compared to the control (about 8 vs 4 t C ha⁻¹ in 0-15cm soil depth) after 25 years. As hypothesized, the SAMM model effectively represented the microbial growth response after litter addition and the ensuing formation and later destabilization of aggregates. The low correlations between different calibrated model parameters ($r < 0.5$ for all parameters; $r > 0.4$ for only 4 of 22) showed that SAMM is parsimonious. SAMM was able to capture differences between treatments in soil organic carbon (Nash-Sutcliffe modelling efficiency (EF) of 0.68), microbial nitrogen (EF of 0.24) and litter carbon (EF of 0.80). The amount of carbon inside aggregates (EF of 0.60) and in the free silt and clay fraction (EF of 0.24) was also simulated very well to satisfactory. Our model results suggested that, despite the sandy soil, up to 50% of carbon stocks were stabilized through aggregate protection mechanisms; and that microbial- and physical aggregate formation coexist. A SAMM model version without aggregate protection (SAMMnoAgg) initially failed to stabilize soil organic carbon (EF dropped to -3.68) and the simulation of microbial nitrogen worsened (EF of 0.13). By re-calibrating SAMMnoAgg, it was possible to partly correct for the lack of aggregate protection by reducing the decomposition rate of mineral attached carbon by about 85% (EF of 0.68, 0.75 and 0.18 for SOC, litter carbon and microbial nitrogen, respectively). However, the slightly better evaluation statistics of SAMM (e.g., Akaike information criterion of 5351 vs 5554) suggest that representing aggregate dynamics within SOM models can be beneficial and is necessary to understand the mechanism behind SOM dynamics. Our results indicate that current models without aggregate formation partly compensate for the absence of aggregate protection by lowering turnover rates of other pools. Thus, they remain suitable options where data on aggregate associated carbon is not available.

Abstract. ~~In light of the large role that Maintaining soil organic matter (SOM) plays in maintaining is crucial for healthy and productive agricultural soils, it is crucial to understand the processes of SOM protection and requires understanding at the process level, including the role of soil aggregate protection. Yet, few numerical process models include aggregate formation and even fewer represent the important SOM protection by soil aggregates and the connection between microbial growth and aggregate formation. Here, we propose a model of We developed the Soil Aggregation through Microbial Mediation (SAMM); which consist of measurable pools and couples soil aggregate formation to microbial growth. The model was evaluated model, to represent this important connection. The pools of SAMM are fully measurable and we calibrated and evaluated it against data from a long term bare-fallow experiment in a tropical sandy soil, subject to. This experiment received plant litter additions of different compositions, leading to twice the soil carbon stocks in the best treatment compared to the control (about 8 vs 4 t C ha⁻¹ in 0-15cm soil depth) after 25 years. The As hypothesized, the SAMM model effectively represented the microbial growth response after litter addition and the following ensuing formation and later disruption destabilization of aggregates. Model parameter correlation was low (all The low correlations between different calibrated model parameters ($r < 0.5$ for all parameters; $r > 0.4$ for only 4 of 22 parameters) showing-) showed that SAMM is well-parameterized. Differences between treatments resulting from different litter compositions could be captured by SAMM for parsimonious. SAMM was able to capture differences between treatments in soil organic carbon (Nash-Sutcliffe modelling efficiency (EF) of 0.68), microbial nitrogen (EF of 0.24) and litter carbon (EF of 0.80). Aggregate-related fractions, i.e., The amount of carbon inside aggregates (EF of 0.60) and also carbon in the free silt and clay fraction (EF of 0.24) were was also simulated very well to satisfactory. Analysis of model parameters led to further noteworthy insights. For example, model Our model results suggested that, despite the sandy soil, up to 50% of carbon in the soil is stocks were stabilized through aggregate protection, even in a sandy soil, and that both microbial activity mechanisms; and that microbial- and physical aggregate formation coexist. When aggregate formation was deactivated, the model A SAMM model version without aggregate protection (SAMMnoAgg) initially failed to stabilize soil organic carbon (EF dropped to -3.68) and microbial nitrogen was represented less well the simulation of microbial nitrogen worsened (EF of 0.13). By re-calibrating the model version with deactivated aggregates SAMMnoAgg, it was possible to partly correct for removing the aggregate formation, i.e., the lack of aggregate protection by reducing the decomposition rate of mineral attached carbon by about 85% (EF of 0.68, 0.75 and 0.18 for SOC, litter carbon and microbial nitrogen, respectively). Yet, the overall However, the slightly better evaluation statistics of SAMM (e.g., Akaike information critereon-criterion of 5351 vs 5554) show the potential importance of suggest that representing aggregate dynamics within SOM models can be beneficial and is necessary to understand the mechanism behind SOM dynamics. Our results indicate that current models without aggregate formation partly compensate the missing protection effect for the absence of aggregate protection by lowering turnover rates of other pools and thus may still be. Thus, they remain suitable options where data on aggregate associated carbon is not available.~~