

## **List of all relevant changes:**

We have substantially revised the manuscript to address the reviewers' comments and to improve the clarity, transparency and robustness of the study. The main changes introduced in the revised version are summarized below:

1. We have revised the model initialization procedure and plant C input assumptions. We revised the modelling framework to better account for non-equilibrium initial conditions and the perennial nature of the olive grove system. In particular, we incorporated a historically driven spin-up approach and explored the use of annual aboveground biomass estimates derived from remote sensing as a proxy to vary plant-derived C inputs over time. Also, we clarified how the initial SOC baseline was defined.
2. We have changed the soil moisture rate modifying factor according to Farina et al. (2013) to improve the representation of soil moisture effects under semi-arid conditions.
3. We included irrigation as an approximate water input in the simulations. We clarified that, in the revised manuscript, irrigation water was incorporated into the model as an additional water input, together with precipitation, in order to improve the representation of soil moisture conditions.
4. We revised figures, tables and supplementary material. We have corrected them according to the revised results.
5. We have expanded and revised the discussion of the mismatch between measured and modelled SOC, especially for biochar. Also, we expanded the discussion of structural limitations of RothC, particularly regarding semi-arid conditions, biochar-related processes and the model simplification in modelling the mixture as two independent amendments.
6. We have improved methodological detail. We added further information on model inputs, simulation settings, parameterization choices and performance metrics. It has improved the reproducibility of results.
7. We have included the site history and field management. We added further details on land-use history before the experiment, the establishment of the olive grove, residue management practices and the handling of pruning residues.
8. We revised the methodological description of the MAD outlier filtering, clarified the selected threshold and added supplementary information reporting the number and proportion of

excluded observations. We also strengthened the discussion of how this issue may particularly affect biochar due to its heterogeneous and particulate nature.

### **Author's responses:**

The author's responses are provided in colour blue below

### **Reviewer 1**

#### **General comment**

This interesting and well-structured manuscript is about the challenge of matching soil carbon modelling results with soil carbon measurements of a field experiment with organic amendments (biochar, compost). Specifically, it compares soil carbon dynamics that were modelled with a modified version of the RothC model with measurements from an 11-year old field experiment on an olive grove. The authors found a large mismatch for the biochar treatment, and an acceptable match for the compost treatment. The manuscript is well written and covers many important aspects. Thus, it has the potential to be an important contribution to the discussion around using SOC models for MRV systems in carbon accounting frameworks.

However, I see one larger issue that is not discussed well enough at the moment: The biochar treatment included a massive addition of C of up to 13.46 Mg C ha<sup>-1</sup> per application (L.114). This does not seem to be reflected well in the measured SOC over time (Fig. 4). In the biochar facet of Fig. 4, we basically see a step from the first measurement (that is, the mean SOC from the control plot only) to the second measurement (that is, the first real measurement from the biochar plots), and no change afterwards anymore, even though there were four additional biochar application events over that time. In my view, this shows two issues: (1) It could be possible that the treatment plots are not comparable to the control plots. This needs to be checked and, ideally, the initial SOC of each plot needs to be shown in the corresponding figure instead of using the mean SOC from the control plot for all treatment plots. And (2) surprisingly, there is almost no biochar-induced C increase detectable. In the discussion, the authors attribute this to the "randomness associated with manual soil sampling" (L. 405), but this seems to apply only to the biochar treatment but not the compost treatment and therefore, cannot tell the whole story. Could this also be a result of the outlier exclusion (L.135), where potentially those measurements with large local amounts of biochar-C were excluded? Do the authors have any recommendations how to improve future field experiments and sampling? The fact that this massive biochar application does not result in significant SOC increases is a surprising and interesting finding that needs more discussion (after thorough checks of potential errors) before the manuscript can be considered for publication.

Thank you for your comments and for highlighting this important issue. The behaviour of the biochar treatment in the field measurements is indeed one of the most interesting outcomes of this study, and we agree that it deserves clarification.

We used the average SOC value of the control plots at the first sampling date as the baseline because we considered it to be the most consistent reference, as it was based on a larger number of observations and represented soil conditions before any treatment effect could occur. Based on the design of the field experiment, we assumed that the plots were comparable at the beginning of the trial.

Regarding the amount of biochar-C added, we agree that the measured response appears unusual, as the biochar treatment does not show a clear increase throughout the experiment despite repeated applications. Concerning the filtering process, the excluded samples were identified according to the MAD-based procedure described in the manuscript. It is true that including these excluded values would increase the average SOC values of the biochar plots. However, we considered that these extreme values were more likely to reflect local hotspots with a high concentration of biochar, rather than representative plot-scale SOC values. In that sense, retaining them would probably overestimate the average SOC content of the sampled soil.

We agree that this point should be discussed more thoroughly. We will therefore expand the discussion to better address this unexpected result, including the possible role of spatial heterogeneity, sampling limitations, problems related to appropriate soil incorporation, and the challenges of adequately capturing biochar-derived carbon in field conditions. One important point is the limited association of biochar with mineral soil fractions, as most of the biochar remains as particulate organic carbon (POC) rather than mineral-associated organic carbon (MAOC), making it more susceptible to losses through wind or rainfall events.

### **Specific comments**

L92: Please clarify for which treatment and time this organic C content applies. Or is it the average from all measurements?

We will clarify this point to improve the description of the site characteristics. We assumed the same initial organic C content for all plots, based on the average of the control samples collected at the beginning of the experiment.

L135: Please add how many samples were classified as outliers (number and % of all samples). Could this lead to exclusion of samples with large local amounts of biochar-C? Please consider showing all data in the supplement.

Thank you for this constructive comment. We agree that the excluded samples are an important aspect to evaluate. To clarify this issue, we will include a table in the Supplementary reporting the number and percentage of samples classified as outliers.

We also agree that this point is particularly relevant because the excluded values mainly correspond to samples with very high C contents. These extreme values were more frequent in the biochar treatment, likely because the particulate and spatially heterogeneous nature of biochar increase the probability of collecting local hotspots compared with the compost treatment.

L147: How large was the short-term variability? Did this also affect the earlier measurements?

Thank you for this comment. Lines 145–147 refer to the procedure used to define the final SOC stock for the calculation of SCS, SOC increase, and the other parameters reported in Table 2. As shown in Figure 2, the last three samples were collected between November 2023 and May 2024. Therefore, the term “short-term variability” refers to the use of three consecutive sampling campaigns, instead of relying on a single final sampling date.

This approach was used to reduce the influence of temporary fluctuations and to obtain a more robust estimate of the final SOC stock than would be obtained from a single sampling event. In this context, earlier measurements were not included in the calculation of the final SOC stock and therefore did not affect this specific estimate. We will rewrite this paragraph to make the procedure and the meaning of “short-term variability” clearer.

L202: Please specify whether the amount of irrigated water was considered in the simulations, e.g. as precipitation. In case it was not considered, it needs to be added in the simulations.

Thank you for this comment. The amount of added water was not considered in the simulation and, following the suggestion of the reviewer we will rerun the simulation considering also the average water provided with irrigation every year.

Irrigation water was not included in the simulations. As stated in the manuscript, the olive grove is equipped with a drip irrigation system, but irrigation is applied only occasionally and mainly during periods of high water demand. In addition, the amount of water applied was not controlled or recorded by us during the experiment. For this reason, we did not include irrigation as an additional water input in the model simulations.

In practice, the simulations were performed assuming rainfed conditions. We agree that this point should be clearly stated in the manuscript, and we will revise the text accordingly.

We acknowledge that excluding irrigation may introduce additional uncertainty in soil moisture and decomposition dynamics; we will mention this limitation in the discussion.

L206: Please clarify how the plant residues were handled prior to the experiment and during the experiment: Is this annual C input only belowground C input (roots and exudates) due to removal of aboveground biomass (branches, leaves are removed), or is it also woody biomass input (branches and leaves decompose directly in the soil)? Did the residue management change with the beginning of the experiment or during the experiment (e.g., branches were left on the ground and are now removed in order to convert them as compost/biochar)?

In the revised version, we will consider total plant C inputs in the simulations. Starting from the plant C input estimated through the model initialization procedure, we will adjust the yearly plant C input using annual indirect estimates of aboveground biomass derived from satellite images (see also our response to the comment on line 192). We will use the remote sensing estimates of aboveground biomass as a proxy for biomass availability, and therefore as an indicator related to plant-derived C inputs.

Regarding residue management, before the experiment shredded prunings were applied to the soil. pruning residues were chipped and at least partially left on the ground between the tree rows, while another significant fraction was transported to the composting facility. Compost had also been applied to the field since 1997, although at a different application rate. More details on the farm management practices will be included in the revised manuscript. We will clarify this point in the revised manuscript, as it is relevant to understanding whether pruning-derived C may have contributed to soil inputs. If possible, we will also explore whether an approximate quantitative estimate of these proportions can be provided. Importantly, this management did not change during the experimental period or in the period immediately preceding the start of the trial.

In the drip-line zone, where soil samples were collected, the main C inputs during the experiment were the applied amendments and the naturally fallen aboveground biomass, mainly leaves and olive fruits. Plant C inputs were not measured directly in the field, but were initially estimated from the 1000-year equilibrium run of the model, which provided the annual plant C input used in the

forward simulations. Due to the lack of direct measurements of aboveground biomass inputs, we assumed that this equilibrium-derived input represented the overall plant contribution to the soil.

L210: Please consider to add the distance of the weather station to the experimental field.

We agree that this information would improve the site description, and we will add the distance between the weather station and the experimental field in the revised manuscript.

L240: Please consider to add the amount of irrigated water to this figure, to cover the whole effect on soil moisture.

Thank you for this suggestion. In the previous version of the manuscript, irrigation water was not included in the RothC simulations, and we did not have reliable records of the amount of water applied during the experiment. For this reason, we were not able to add this information to Figure 2.

However, in the revised manuscript, we will include irrigation water in the simulations, and therefore this information will also be incorporated into Figure 2. Because irrigation was not directly monitored, the added water will be incorporated as an approximate average input.

L241: “Bars represent standard deviation.” This is incorrect for the upper panel: Bars show the precipitation; standard deviation is not shown in upper panel.

Thank you for this comment. The sentence “Bars represent standard deviation” was intended to refer to panel (b), not to the upper panel. We will correct this mistake in the figure.

Figure 3: The resolution is not sufficient; Please consider to submit as vector graphic if possible.

Yes, we will include a better version in the revised manuscript.

L274: Are there also measurements of the initial SOC of the treatment plots?

No, SOC was not measured in the treatment plots before the first amendment application. For this reason, the initial SOC baseline used in the analysis was the mean value of the control plots at the first sampling date.

L292: Which means that the assumption of the SOC being in equilibrium was not valid. Please clarify in the Material and Methods section how the land has been used before the implementation of the experiment and prior to the establishment of the orchard). Land-use changes can have a long legacy effect on SOC of more than 100 years (e.g., Emde et al. 2024).

We agree that the increase in SOC observed in the control treatment suggests that the assumption of strict SOC equilibrium was not fully valid.

We will try to address this concern in two ways:

1) We will apply a historically driven spin-up procedure to initialize the model, as suggested by Wiltshire et al. (2024), in order to account for the effect of land-use history prior to the establishment of the experiment. For this purpose, we can use SOC data from a neighbouring plot located approximately 300 m from the experimental site, where organic amendments have never been applied and irrigation has never been used. We consider these conditions to be the most similar to those of the experimental plot before the land-use change in 1997. This soil can reasonably be considered close to equilibrium, and therefore suitable for applying the “historical” spin-up approach proposed by Wiltshire et al. (2024). This method may improve the estimation of the initial

pool sizes, assuming that at the beginning of the trial the system was not at equilibrium because of the land-use change in 1997.

2) We will vary the annual amount of plant C inputs entering in the soil on the basis of annual variation in Above Ground Biomass as derived from remote sensing data.

Reference:

Wiltshire S., Grobe S., Beckage B. 2023. A historically driven spinup procedure for soil carbon modelling. *Soil systems*, 7.35. DOI: 10.3390/soilsystems7020035

We believe that this behaviour may be related to the growth of the olive trees during the experimental period, which likely increased plant-derived C inputs over time, particularly through aboveground biomass. This point is already discussed in the manuscript, and we will make it clearer in the revised version.

Regarding land-use history, the site has been managed as an olive grove since 1997. Before the establishment of the orchard, the land was also used for agriculture. We will clarify this information in the Materials and Methods section.

L318: Please specify for which model the statistics are given: RothC or the regression line. Please include your approach for the regression line in the Material & Methods section.

The regression line applies to the measurement data. It will be explained clearly in the revised manuscript.

L322: It is not clear to me how the control SOC change was included in this estimate. If we want to know the treatment effect, the control SOC trend needs to be subtracted from the treatment SOC trend (baseline== control trend), but I think this was not done here. Please clarify this accordingly.

As you say, the control treatment reflects the SOC dynamics in the absence of amendment application and therefore serves as the reference condition. In the upper part of Table 2, however, we show the SOC evolution of each treatment throughout the experiment, including the control as one of the treatments.

For example, the final SOC stock for the compost treatment corresponds to the mean SOC value of the last three sampling dates, collected after the final amendment application. This procedure is described in the Materials and Methods section.

The treatment effect relative to the control is considered in the lower part of Table 2, where the remaining EOM-C and the SCS rate are calculated with respect to the control treatment. So the control SOC values were subtracted from the treatment SOC for the calculation of the yearly sequestration rate of the amendments, as correctly suggested by the reviewer. Example for Compost  $(61.49-43.09)/11 = 1.67 \text{ Mg ha}^{-1} \text{ y}^{-1}$

The table 2 also reports SCS normalized for an EOM addition of  $1 \text{ t C ha}^{-1} \text{ y}^{-1}$ .

We will clarify this distinction more explicitly in the revised manuscript.

L325: Does this mean that the compost was more stable than the biochar?

This is the result obtained from the measured field data after normalizing for the amount of C added with each amendment. However, this result should be interpreted with caution, because the field measurements involve a high degree of uncertainty.

In Section 4.3, we discussed the apparent similarity between compost and biochar in the measured data does not necessarily indicate that compost was intrinsically more stable than biochar, but rather that biochar-derived C may be more difficult to quantify reliably under field conditions.

We think that these data highlight the difficulties to effectively measure variations in SOC stocks in biochar amended soils as outlined by Chiaramonti et al. (2026)

L355: There are also studies which claim that RothC does not represent the soil moisture conditions in semi-arid conditions well - please refer to Farina et al. 2013 for more information.

Thank you for this interesting reference, we will review it carefully. The main scope of this paper is to compare field-measured SOC data with RothC simulations, as RothC is a widely used biogeochemical model in many applications, including MRV assessments.

As the olive grove was irrigated, we assume that the soil never reached the condition (very low water content) for which the modified model proposed by Farina et al. 2013 would be more appropriate.

L364: The assumption of both the SOC and C input from the olive grove being in equilibrium is not valid.

Thank you for these important remarks. As explained previously (L292) we will try to address both these concerns by considering an historically driven initialization procedure and by varying plant C during the experiment input according to AGB data

L379: Please consider to add the biochar content in parentheses, e.g. (5%).

Thank you for the comment. We will add the biochar content to the mixture

L390: Please clarify what other "routine agricultural practices" could increase the "background carbon inputs".

By this, we refer to natural and management-related processes that could affect SOC over the course of the field experiment. Using the control treatment as a reference allows us to account for these effects and makes the comparison among treatments more consistent.

L401-409: Please elaborate a bit more on how biochar field experiments should be sampled to fully capture the SOC increase. How could the study design and sampling design be improved in the future? Now it is not clear enough how to reduce the error, and everything between 100% model error and 100% sampling error seems possible. Can you give a tendency, what needs to be done to find out whether the biochar was really not as stable as we thought, or whether the sampling approach needs to be changed? How should biochar research proceed in the future?

Thank you for the comment. The main focus of this study is to quantify the discrepancy between SOC measurements and RothC simulations. A detailed methodological assessment of how to reduce uncertainty in biochar field sampling goes beyond the scope of the present work and cannot be fully addressed with the methodology used here. However, we believe that this study helps to identify the main sources of difficulty, particularly for biochar, which is a material that is difficult to quantify reliably under field conditions.

We agree that this point needs further discussion, and we can modify this section to better highlight the challenges associated with biochar sampling and field-based verification.

Furthermore, it is worth noting that FAO recommends using less data-demanding models, such as RothC, which parameterizes soil processes primarily through empirical and conceptual functions. Due to its low input data requirements and its proven suitability for simulating SOC in various LULC areas with limited data availability, the RothC model was selected for use in the FAO procedure for the determination of the Global SOC Sequestration Potential (GSOCseq) Map.

We think that an integration of well-designed sampling strategies with accurate modelling could be the way forward for reliable SOC measuring fit/adapted to MRV systems

L463: Do you think these output pathways could be of relevance in this specific study? Please elaborate a bit more on potential losses via erosion.

Given the topography of the experimental field, we consider fluvial transport of the amendments to be insignificant in this specific study. However, we agree that fluvial and wind transport could represent a potential source of uncertainty, particularly for biochar, and we will mention this as a possible source of error.

Technical corrections

L92: Remove the % sign after the pH value.

We will remove the % sign

L219: Abbreviation not yet defined.

Root Mean Square Error (RMSE) will be written as it.

L283: Move the definition to the first mentioning of the abbreviation.

We will move the RMSE for the previous mention

## Reviewer 2

This manuscript presents an 11-year field experiment in an organic olive grove combined with RothC simulations to evaluate soil organic carbon (SOC) dynamics under different organic amendments (compost, biochar, and their mixture). This manuscript involves a considerable amount of work on field experiments and model simulations. The availability of long-term experimental data is a notable strength and provides valuable data for model validation, yet these uncertainties in model validation and model prediction accuracy are not adequately discussed. I have the following suggestions for improving the research.

### Major

1. This manuscript compares the model-simulated SOC with observed values, and a key result of the study is the poor agreement between modelled and measured SOC, particularly for the biochar treatment (e.g.,  $R^2=0.10$ , high RMSE). However, the manuscript does not adequately explain the reasons for this discrepancy. More discussion is needed to explore potential sources of inconsistency, such as model structural limitations, the absence of potential mechanism processes (e.g., priming effects), and parameterization choices.

Thank you for your constructive comment. As you say, the agreement between simulated and measured SOC was poor, especially for the biochar treatment. We acknowledge that SOC monitoring under field conditions is affected by substantial uncertainty related to soil sampling and spatial heterogeneity, as discussed in Section 4.1. In addition, model performance depends strongly on calibration and initialization procedures. In the first version of the manuscript, the forward simulations were initialized assuming equilibrium pool sizes and constant plant C inputs, but as the reviewer pointed out (Point n. 12) this seems not to be the case.

As explained above, we will try to address and solve these two important issues.

The discrepancy between measured and modelled SOC stocks dynamics in the biochar amended plot rely more on the problems related to the field measurement as depicted by Chiamonti et al 2024.

The model does not take in consideration priming effect but due to the quite low H/C of applied biochar, indicative of extremely stable biochar, possible positive priming effect due to biochar are likely negligible. In addition, if any priming effect would have occurred this would have been too be higher in compost amended soil.

We adopted this approach because, given the lack of direct data to better constrain plant-derived inputs and initial pool sizes, it was the most practical initialization procedure for our study.

The biochar treatment showed the lowest agreement between measured and modelled SOC. We believe that this discrepancy is influenced by several factors, mainly:

- The high spatial heterogeneity of field measurements, which is increased by the random component inherent to manual soil sampling.
- The particulate component of biochar, which makes its distribution in soil more heterogeneous and may affect the representativeness of normal soil sampling for monitoring SOC changes.

From a modelling perspective, several aspects related to model parameterization and application may also contribute to the observed discrepancies, and we agree that the discussion can be expanded to better explain these aspects, in particular the role of parameterization choices. In particular, biochar was parameterized using the H/C molar ratio following Keel et al. (2023), whereas compost was parameterized by model inversion using the approach proposed by Mondini et al. (2017). This last approach is more specific to the material used in our case and likely provides a more precise parameterization than the literature values adopted for biochar.

In the new version of the manuscript, we will adopt for biochar the parameterization proposed in the Deliverable WP3.1-3.5 of the EJPSoil project CarboSeq (Leifeld et al., 2024) based on the H/C molar ratio of biochar. In this procedure the more relevant parameter (i.e. decomposition rate of biochar resistant pool) is obtained by fitting a double exponential model to Fperm i.e. the amount of biochar C remaining in the soil after 100 years. Fperm was derived by several laboratories and field biochar decomposition study of at less one year of duration. As such, biochar parameters estimation is also based on experimental data and we assume to be sufficiently robust as the approach we have followed for compost

Reference: Leifeld J., Hardy B., Budai A., Elsgaard A., Keel S.G., Levvasseur F., Liang Z., Mondini C., Plaza C., Rodrigues L. 202. WP3.1-3.5. Final report work package 3 – Biochar and other organic amendments. CarboSeq project. DOI: 10.5281/zenodo.14046237

In summary, we agree that the manuscript already addressed several sources of discrepancy between measured and modelled SOC, but that the discussion was too dispersed and not explicit enough. We will revise this section in detail to present these reasons more clearly in the revised manuscript.

Keel, S. G., Bretscher, D., Leifeld, J., von Ow, A., and Wüst-Galley, C.: Soil carbon sequestration potential bounded by population growth, land availability, food production, and climate change, *Carbon Manag.*, 14, <https://doi.org/10.1080/17583004.2023.2244456>, 2023.

Mondini, C., Cayuela, M. L., Sinicco, T., Fornasier, F., Galvez, A., and Sánchez-Monedero, M. A.: Modification of the RothC model to simulate soil C mineralization of exogenous organic matter, *Biogeosciences*, 14, 3253–3274, <https://doi.org/10.5194/bg-14-3253-2017>, 2017.

2. Has the model been sufficiently validated with consideration of soil heterogeneity? What is the scalability of the single-point validation? That is, can the model be reasonably applied to other regions? It is recommended that the authors further validate the RothC model using literature data or by testing alternative validation data from other soil carbon models (e.g., MIMICS-BC, EPIC, APSIM). At the very least, this issue should be discussed.

References:

Han M, Zhao Q, Wang X, et al. Modeling biochar effects on soil organic carbon on croplands in a microbial decomposition model (MIMICS-BC\_v1.0)[J]. *Geoscientific Model Development*, 2024, 17(12): 4871-4890.

Lychuk T E, Izaurrealde R C, Hill R L, et al. Biochar as a global change adaptation: predicting biochar impacts on crop productivity and soil quality for a tropical soil with the Environmental Policy Integrated Climate (EPIC) model[J]. *Mitigation and Adaptation Strategies for Global Change*, 2015, 20(8): 1437-1458.

Archontoulis S V, Huber I, Miguez F E, et al. A model for mechanistic and system assessments of biochar effects on soils and crops and trade-offs[J]. *Gcb Bioenergy*, 2016, 8(6): 1028-1045.

Thank you for this comment. This study was designed as a site-specific comparative analysis under different amendment applications. Therefore, the main objective was not to perform a multi-model validation, but rather to compare field-based SOC measurements with the outputs of a widely used model such as RothC, identify the main limitations of both approaches, and evaluate how different amendment types affect the agreement between measured and modelled SOC.

For this reason, we did not intend to validate RothC against other modelling frameworks. We agree that such a comparison would be interesting and valuable, particularly for assessing the scalability of the results and the representation of biochar-specific processes, but it is beyond the scope of the present article.

RothC has been extensively tested in different regions and under different managements. Moreover it has been validated at different scales from single field data to regional scale

We also agree that this issue should be discussed in more detail. In the revised manuscript, we will clarify that our conclusions are based on a single long-term experimental site and that the transferability of the results to other regions or systems should be considered with caution.

3. Given that the compost contains sheep manure, it is likely a source of nutrients. How were these nutrients accounted for in the model? Was the nutrient content of the added compost measured experimentally? Because the nutrient may stimulate soil microbial activity, thereby influencing soil carbon decomposition. This is a critical issue that the manuscript does not address. For example, on L175, the mixture of compost and biochar is treated as independent components without considering interactions. However, nutrients from compost could stimulate microbial activity and potentially affect biochar degradation. This will introduce uncertainty that should at least be discussed.

Thank you for this important comment. We did not explicitly account for nutrient interactions in the simulations, mainly because RothC does not represent nutrient microbial interactions. Our simulations were performed from the perspective of SOC dynamics and C sequestration. Therefore, the mixture treatment was modelled as the sum of two independently decomposing components, which is a simplifying assumption of the present study and a potential source of uncertainty.

We agree that nutrients supplied by compost could influence microbial activity and potentially affect the decomposition dynamics of both native SOC and biochar. This effect was not explicitly represented in our modelling approach, and we will discuss this limitation more clearly in the revised manuscript.

Biochar is acknowledged to be extremely resistant to degradation, and we do not think that adding sheep manure would increase biochar degradation. In contrast there are evidence that adding biochar to compost increases the stability of compost

Regarding the interaction of biochar and composting, although this interaction is relatively well understood in composting, the long-term behavior in biochar-enriched compost after soil application remains uncertain. While biochar addition is known to enhance microbial activity and influence carbon stabilization during composting, it is unclear whether this interaction persists once the material is stabilised and incorporated into soil organic matter.

In this study, the similar field behavior and comparable model performance for compost-only and biochar-enriched treatments indicate that representing compost and biochar as two separate components does not introduce a significant source of uncertainty in modelling long-term SOC dynamics. Both amendments exhibit similar experimental trends in the field, which are consistently reproduced by the model, suggesting that any potential interaction between compost and biochar during soil mineralisation does not substantially alter SOC dynamics in a way that is not captured by the current modelling approach. Therefore, the use of two independent EOMs (compost and biochar) rather than a single combined representation does not appear to bias the simulation of biochar-enriched compost behavior under the conditions of this study. Nevertheless, this aspect warrants further investigation under different soil and climatic conditions.

4. The irrigation was conducted in the field experiment. However, how was this considered in the model? Irrigation practice can affect soil moisture, which further influences soil carbon decomposition (e.g., soil moisture process in Millennial, MEND model).

#### References:

Abramoff R Z, Guenet B, Zhang H, et al. Improved global-scale predictions of soil carbon stocks with Millennial Version 2[J]. Soil Biology and Biochemistry, 2022, 164: 108466.

Wang G, Huang W, Zhou G, et al. Modeling the processes of soil moisture in regulating microbial and carbon-nitrogen cycling[J]. Journal of Hydrology, 2020, 585: 124777.

Thank you for your suggestion. We will rerun the simulations including an average proxy for the irrigation water applied throughout the trial. Because irrigation was not directly monitored during the experiment, this input will be incorporated as an approximate average value rather than as an exact annual record.

#### Minor

5. Line 50: In addition to these first-order kinetic soil carbon models, there are also microbial models, such as MIMICS and MEND, which might offer greater utility for understanding the impact of biochar addition on soil carbon dynamics (e.g., microbial process, priming effect). The authors should provide a more comprehensive overview of existing modeling approaches and discuss how their model choice compares with these alternatives.

Thank you for this comment. We will consider mentioning these kinds of models to provide a broader overview of the existing modelling approaches. However, the purpose of this paragraph was to mention the main standard SOC models that have been widely used for soils receiving organic amendments.

There are indeed models more complex than RothC capable of a better description of the processes affected by biochar addition, but such models require a larger set of input data that are also difficult to measure/obtain. In our opinion the RothC model represents a good trade-off between reliable and robust results and simplicity and availability of input data.

6. Line 128-129: The determination method for SOC concentration should refer to the corresponding references.

The parentheses in this paragraph were intended to indicate the analytical instruments used to determine SOC concentration, not references. We will rewrite this paragraph to avoid misunderstanding and make the analytical method clearer.

7. Line 134: Why was the MAD threshold set to 4, and what percentage of the data was filtered out as a result?

Thank you for your suggestion. The threshold was selected after testing different MAD thresholds and evaluating their influence on data retention and on the exclusion of clearly extreme values.

There is no universal MAD threshold in the literature. Thresholds can vary depending on the application and the desired balance between sensitivity and data retention. In our case, because the dataset comes from field experiment with limited replication and substantial spatial variability, we intentionally adopted a relatively high threshold (MAD = 4) as a conservative filtering criterion. The threshold was selected after testing different thresholds and evaluating their effect on data retention and on the exclusion of only the most extreme values. We can add a table with the percentage of removed values per date and per plot in the supplementary.

8. Line 153: Is there evidence to support the assumption that complete mixing occurs within 5 months? Several methodological choices are not sufficiently justified and may bias the results.

We agree that this decision has limitations. The exclusion of the first five months after amendment application was intended as a conservative choice, in order to avoid using SOC measurements that could still be affected by incomplete incorporation or by short-term variability immediately after amendment addition.

As shown in Figure 2, some SOC measurements taken soon after amendment application displayed unusual behaviour. This was particularly evident after the last amendment event, when the observed values appeared anomalous. For this reason, we considered that excluding the first five months after each application provided a more reliable basis for comparing measured SOC with the modelled values.

We will revise this part of the manuscript to explain this methodological choice more clearly and acknowledge its limitations.

9. Line 155-159: The current description of the RothC model is too general. At least a structural diagram of the RothC model should be provided, illustrating the carbon pools and carbon flows, to facilitate a more intuitive understanding of the simulation processes for different amendment additions. In addition, the model time step of the RothC simulations is not explicitly specified.

We appreciate the reviewer's comment. While a detailed description of the RothC model (a widely used SOC turnover model) is beyond the main scope of this study, we agree that the current presentation can be clarified to improve readability.

In this work, the key focus is the parameterization of amendment-specific pools and their application to the field experiment, rather than a full methodological description of the model itself. Nevertheless, we acknowledge the importance of providing sufficient clarity for the reader.

10. Line 188-200: Compost parameters are derived via model inversion, but biochar parameters are adopted from the literature without calibration. This inconsistency may limit the comparability of treatments and should be explicitly acknowledged and discussed.

We thank the reviewer for this comment. The different approach used for compost and biochar reflects differences in data availability and suitability for parameter estimation rather than an inconsistency in the modelling framework.

For compost, extensive datasets generated within our research group across different soils and organic amendments (including olive mill residues) were available, allowing robust parameter optimisation through model inversion and ensuring that a wide range of system variability was captured.

For biochar, such datasets were not available. In addition, the very low CO<sub>2</sub> emissions associated with biochar mineralisation limited the applicability of model inversion, as the resulting signals were too small and insufficiently variable to reliably constrain parameters.

Therefore, biochar parameters were taken from previously published and widely validated literature sources (See our comment to the point n.1). This approach is commonly adopted for highly stable carbon fractions and provides consistent and comparable parameter values across studies.

This point will be explicitly clarified in the revised manuscript.

11. Line 189: More details or a summary of the inversion procedure should be provided in the main text. Currently, the description is too dependent on supplementary material.

We understand your concern. Our intention was not to describe the inversion procedure in full detail in the main text, in order to keep the manuscript concise and avoid overloading the reader with methodological detail. For this reason, the full description was provided in the Supplementary.

12. Section 2.3.3: The model assumes constant plant carbon inputs derived from equilibrium runs. However, the field system is a perennial olive grove, where biomass accumulation is likely to increase over time. The authors should justify this assumption, discuss its implications, and consider uncertainty tests or sensitivity tests on the main model parameters to strengthen the credibility of the findings.

Thank you for your constructive comment. This issue is already addressed in the current Discussion, but we agree that it should be explained in more detail. In particular, we will clarify more explicitly that the assumption of constant plant C inputs is a simplification, and that it may not fully represent the progressive biomass accumulation expected in a perennial olive grove.

To solve this problem, we will perform simulations varying C plant inputs based on aboveground biomass availability, in order to better evaluate the sensitivity of the results to this assumption.

13. Section 2.3.3: The input variables information (e.g., source, spatial and temporal resolution) of the model, including those used for long-term equilibrium and forward simulations, should be described in detail. A summary table would greatly improve clarity and reproducibility.

Thank you for your comment. In the next manuscript version, we can include a table in the supplementary material specifying the model input variables, their source, and the main settings used for both the equilibrium and forward simulations.

14. Line 219: Performance metrics ( $R^2$ , RMSE, Bias) should be explicitly defined at first mention, including their calculation, unit, and interpretation.

In the revised manuscript, we will define the performance metrics in the Materials and Methods section, including their formulas and units.

Model performance was evaluated using the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and bias.

$R^2$  (unitless) was obtained from a linear regression between measured and simulated values and was used to describe the strength of the association between both datasets.

RMSE was calculated as the square root of the mean squared difference between measured and simulated values, and is expressed in the same units as the data ( $\text{Mg C ha}^{-1}$ ). Lower RMSE values indicate better agreement between observed and simulated values.

Bias was calculated as the mean difference between simulated and measured values (simulated – observed), also expressed in  $\text{Mg C ha}^{-1}$ . Positive bias indicates model overestimation, whereas negative bias indicates model underestimation.

15. Line 241: The note “Bars represent standard deviation” is suggested to be added to the end of the Figure 2 caption.

Yes, it was an error. We will correct this problem in the next manuscript version

16. Fig. 4: Statistical significance should be added.

We thank the reviewer for this suggestion. Figure 4 already includes regression lines fitted to the measured SOC values. To clarify their statistical support, we will add the significance level of each regression (p-value) in the corresponding Figure

17. Line 377-380: The authors should explain why compost treatments are more predictable than biochar treatments. One reason is likely due to the lack of mechanisms to capture biochar-induced changes in native soil carbon dynamics. These uncertainties should be discussed.

Thank you for this comment. This point was already addressed in the Discussion, although we agree that the explanation can be made more explicit. In the manuscript, we already indicated that compost showed the closest agreement between field measurements and RothC simulations, whereas biochar treatments exhibited substantially larger discrepancies.

We also discussed that this difference is likely related to the more homogeneous distribution and integration of compost within the soil matrix, in contrast to biochar, whose granular nature, high carbon concentration, and spatial heterogeneity make its field quantification more uncertain. In addition, the parameterization of biochar in RothC is more simplified and may not fully capture its behaviour under field conditions.

We will revise this part of the Discussion to make these sources of uncertainty clearer and more explicit.

18. Line 454-455: The authors mention that a two-pool decay model was tested, which seems that the test was conducted by the authors, yet no supporting results or references are provided. Please clarify this point.

These lines refer to the work of Sanei et al. (2025), not to a test conducted by the authors of the present study. We will rewrite this part of the manuscript to avoid misunderstanding.

Sanei, H., Petersen, H. I., Chiaramonti, D., and Masek, O.: Evaluating the two-pool decay model for biochar carbon permanence, *Biochar*, 7, 9, <https://doi.org/10.1007/s42773-024-00408-0>, 2025.