

Referee comment 2

We are very grateful for the time you took to review our manuscript and for the valuable comments that will help improve the work. In this document, we have included each of your suggestions and have attempted to address them as best as possible (text in blue following each comment). All these comments will be taken into account and incorporated into the revised version of the manuscript.

This manuscript analyses soil samples from a long-term agricultural experiment in Uruguay to assess how organic carbon and radiocarbon persistence is shaped by management. The study uses these sources of empirical data to construct two models from which a 3-pool model explains these data well. This model is then used to evidence different SOC compartments and specifically an intermediate pool which was quickly lost after a management shift in 2008. This is a valuable set of long-term measurements and a well defined analysis. I have two main comments about (1) the model and (2) the sampling approach that should be addressed.

We thank the anonymous reviewer for all the constructive comments. We will try to answer to all the questions raised, and any changes that arise during this process will be incorporated into the revised version of the manuscript.

For the model. Because the $\delta^{14}\text{C}$ is flat from 1984-2008 and only drops after 2008, I think there needs to be additional evidence of allocation to the intermediate pool before such a strong claim can be made about its increase and persistence. For instance, there looks to be data from the CC system in figure 3 with a continuous decline in bulk SOC. A counterfactual using a system that did not have an increase in SOC would lend much stronger evidence that this intermediate, recently acquired C is what is lost. Without that, I think there at least should be additional evaluation on the 3-pool model. Specifically, how sensitive the intermediate pool's trajectory is to its initial starting point; I do not fully follow how this was chosen. I would suggest sensitivity analyses for this and how robust the intermediate pool's trajectory is to its initial values. Relatedly, more clarification is needed to clarify which pools are tied to conceptual vs. empirical constructs/measurements. For instance, if the ^{14}C in the slow pool was measured empirically as MAOM (L207), the implications of this should be discussed in more detail.

We will begin our response to this comment from the end, as this may help clarify the rest of the questions. In these models, all compartments are conceptual, and none have a direct connection to measured fractions. These pools are kinetic compartments, conceptual fractions of the total stock that cycle at different rates. Therefore, they should be viewed as a hypothesis, a conceptual and mathematical representation that adequately represents the measured data. Regarding the measured data used to constrain the model's parameterization, we used the time series for bulk SOC and the time series for bulk radiocarbon, there is no measured data on fractions.

Another important point is the setting of the model's initial conditions, which are assumptions upon which the results are based. As outlined in L205, the way in which the initial conditions for the three-compartment model were defined derives from the way they were defined in the two-compartment model. The total initial SOC stock was set to adequately represent the average for the initial year of the LTE (1964). Regarding the radiocarbon initial condition, the value of the fast pool was set to 0 (an approximate pre-bomb signature; since it is the input compartment, it immediately aligns with the atmospheric signature). Furthermore, an initial manual adjustment was made, setting the signature of the slow pool to -70 so that the model would approximately represent the radiocarbon signature of the initial years of the LTE. Additionally, the initial partition

of the SOC stock between the slow and fast pools was adjusted as a calibrated parameter in the model (in the optimization procedure).

The initial conditions for the three-compartment model are derived from the previous one. The initial radiocarbon signature of the intermediate pool was defined as the mean value between the fast and slow pools. The parameter that defines how much of the total SOC stock at t_0 (1964) is stored in the slow pool was adjusted in the optimization procedure, while the remaining stock was divided equally between the fast and intermediate pools.

Regarding the initial conditions, we consider that, within the range of plausible assumptions, they provide a setup that reasonably represents the system at the beginning of the experiment. The model cannot adequately reproduce the initial radiocarbon conditions without a predominantly slow pool that is strongly depleted in radiocarbon.

Independently of this initial setup, the data showing a decline in the radiocarbon signature after 2008 requires a temporary storage of bomb-C in a fraction of SOC whose residence time is longer than that of the fast pool but shorter than that of the slow pool. This result, which arises from comparing models with different structures (2-box vs. 3-box), indicates the need for an intermediate kinetics compartment that stores the C derived from the nuclear weapons testing peak and from which it can subsequently be lost.

Nevertheless, it is possible to test the model's behavior under varying initial radiocarbon signature conditions in the intermediate pool, provided that these modifications do not substantially compromise the representation of the data at the start of the experiment (around 1964). We will therefore develop an analysis of this type, testing how the three-compartment model behaves at different initial values (in 1964) of radiocarbon in the intermediate compartment.

Regarding the comparison with the CC system, this was included to show that the Period 1 of the "Recovery system" behaves the same way as the CC system, previously evaluated by González-Sosa et al. (2024), with both showing a similar decline in SOC over time. We considered this demonstration relevant because, in Period 1 (1964–1984) of the "Recovery system" modeling, we used the C input rate reported by González-Sosa et al. (2024) for that system. Regarding the counterfactual analysis proposed by the reviewer, in González-Sosa et al. (2024), this CC system was adequately represented using a two-compartment model, without the need to include an intermediate kinetic compartment, as is the case with the "Recovery system" analyzed here. We understand that this comparison with the modeling already performed for CC in González-Sosa et al. (2024) can strengthen the logic of this work if included as a point in the discussion. We will include this discussion in the revised version of the manuscript.

(2) There is clear description of the management changes that occur after 2008, however the implications of these are only evaluated through the sensitivity of C inputs. Going from tillage to no-till seems to also be a major change here. There should be more discussion on this and how this could influence the measurements themselves in addition to carbon dynamics. Specifically, on how the tillage change could produce similar SOC and $\delta^{14}\text{C}$ trajectories through alternative explanations.

We appreciate the reviewer's comment, we believe it is highly relevant and can substantially improve the discussion. From a logical point of view, we attempted to structure the study as follows:

Modeling using the 3-box system and comparing it with the 2-box system leads to the conclusion that it is necessary to consider an intermediate kinetics pool that stores most of the C inputs during the 1984–2008 period and is destabilized after 2008, which provides a mathematical explanation for the combined decline in SOC stock and radiocarbon signature. Because multiple management changes occurred after 2008 (including the 100% transition to no-till), we did not provide a definitive mechanistic interpretation of the causes of that decline; rather, we focused on analyzing the fate (in terms of allocation to different kinetic compartments) of the increases in carbon stock during the 1984–2008 period.

Since the changes that occurred in 2008 (including the 100% transition to no-till) may have led to a decrease in the effective C inputs to the mineral soil, we conducted a sensitivity analysis to assess whether decreases in C inputs alone can explain the data. Since these changes could not explain the data, we reaffirmed the conclusion that the temporal dynamics of SOC and radiocarbon responded to the system's internal dynamics (decomposition rates and transfers), and we reaffirmed the interpretation of storage in the intermediate-kinetic pool.

However, the reviewer's point regarding the potential alternative effects of the 100% transition to no-till, which may not have been adequately discussed, is clear and pertinent. First, there are possible direct effects on the measurements. We will expand the review and improve the discussion about this point. Another possible hypothesis, which may not have been sufficiently elaborated in the discussion, is the increased stratification of C inputs into the system. We hypothesize that this may lead to a greater accumulation of material on the surface compared to the situation with sporadic tillage. However, this process, although not explicitly represented in our model, would be accounted for by an effective decrease in C inputs to the mineral soil, which has already been considered in the sensitivity analysis.

In summary, and as the main takeaway from this response, we find the reviewer's point very interesting and see it as a potential opportunity to strengthen the Discussion. We will expand the review of these issues where appropriate and incorporate them into the Discussion.

Using a single bulk density value measured in 2021 for the previous 60 years is a strong assumption, especially when there is a change in tillage intensity. If BD was lower before 2008, stocks would be overestimated and part of the observed decline could reflect rising BD rather than C loss. I would suggest either keeping the manuscript to SOC concentrations (using the 2021 BD as context on soil condition), or running simulations with plausible BD changes to show how sensitive the stock conclusions are. I expect the results to hold, but it should be addressed directly.

We appreciate the reviewer's comment because it provides a good opportunity to clarify and provide further information on this point. We have data on bulk density measured in 2008 (immediately before the transition to 100% no-till). The average bulk density of the treatment (average of the three corresponding plots) was virtually the same between the two years evaluated (2008 and 2021), being 1.268 Mg m^{-3} in 2008 and 1.253 Mg m^{-3} in 2021, representing a decrease of only 1.18%. Because the difference in bulk density was so small between the two years evaluated, and because this study is based on a long-term time series rather than a comparison between two specific sampling years, we decided to use only the 2021 bulk density data.

Overall, the data support that SOC declined after 2008, that bulk d14C became more depleted, and that among the two model structures tested the three-pool model reproduces these trajectories by increasing loss from an intermediate-timescale pool. It is less clear if this intermediate pool represents a distinct physical SOC compartment that was selectively destabilized. The input sensitivity analysis shows that reduced C inputs alone cannot explain the observed changes, but it does not test how sensitive the inferred intermediate-pool loss is to initial conditions, alternative loss pathways, or changes in sampling representation following no-till. I think the authors should either provide additional sensitivity analyses addressing these assumptions or state conclusions such as “selective destabilization” and “can only be explained” more explicitly as interpretations conditional on the fitted three-pool model.

Thank you for this comment, which will help improve the clarity of the paper’s conclusions and their scope. The reviewer’s point in the first statement of this comment is an accurate summary of the message we aim to convey in this study. The data show that it is possible to recover carbon at a previously degraded site at high rates. However, the only way we were able to mathematically represent the measured data was by storing most of those gains from the 1984–2008 period in a pool that was later destabilized by changes in management practices. This representation allowed the model to reproduce the decline in both SOC stock and radiocarbon after 2008. We agree with the reviewer that the study’s conclusions are derived not only from the data but are also dependent on the structure, initial conditions, and assumptions that define the model. Therefore, we agree that it is appropriate to include the clarification suggested by the reviewer regarding the dependence of the conclusions on the elements mentioned above. This will be done in the revised version of the paper.

Minor comments:

Figure 4 and 6 in the caption says temporal evaluation of SOC stock at 0-20cm. This might be an error, but could you describe how the stock was estimated when changing from 0-20 to 0-15 cm and similarly for d14C?

The reviewer is correct in noting that information was missing to clarify this point; the methodological description was incomplete and did not adequately describe what was done in this regard. Sampling was conducted to a depth of 20 cm until 1996 and to a depth of 15 cm from that year onward. As noted in previous studies (González-Sosa et al., 2024; Grahmann et al., 2020), we expect this modification to have no significant effect on the results because tillage interventions were carried out until 2008, which homogenized the upper part of the soil profile (0–20 cm).

Taking this into consideration, we extrapolated the calculated SOC data from the period 1996–2021 down to a depth of 20 cm by multiplying them by a coefficient of 20/15, as we did previously in González-Sosa et al. (2024). Similarly, it is an assumption of this study that this change did not affect the radiocarbon signature. As we noted in the methodological description, if there were an effect, it would only be evident in the most recent years of the time series (after several years of accumulation under the no-till system). In addition, any potential effect on the radiocarbon signature arising from a progressive increase in the stratification of C inputs would be expected

to increase the radiocarbon signature, because more recent C would accumulate near the soil surface and there would be less dilution with older C from deeper soil layers. Therefore, any correction for this potential process would be expected to act in the same direction as the observed results and increase their magnitude.

In any case, we apologize for omitting a complete explanation of how SOC stock was calculated after 1996. We will include this in the Materials and Methods section, along with the considerations that should be taken into account when interpreting the results due to the assumptions arising from this aspect of the sampling.

Table 1 / Fig. 2: Please use consistent notation for defining the different variables (e.g. k value subscripts) . Also can you clarify output rate?

Thank you for your comment. We will correct all these inconsistencies throughout the manuscript by referring to the output rate constants of the fast, intermediate, and slow compartments as k_f , k_i , and k_s , respectively.

The output rate constants (k , yr^{-1}) of each compartment in a dynamic compartmental model are the first-order rate constants that determine the rate at which C leaves the corresponding compartment. Thus, at any given moment, the outflow rate from that compartment is equal to the product of the compartment's corresponding k and its stock. Moreover, that outflow can have two alternative destinations: it can move to another compartment (determined by the transfer coefficient α) or leave the system ($1 - \alpha$).

Study design: please expand the description rather than relying on the previous 2024 study. It would help to state here which of the seven treatments were used, the three replicates per treatment, and the composite sampling (20 points from the plot center). The staggered cropping rotation across replicates is worth clarifying as well.

Thank you for this comment. We will add this relevant information to the Materials and Methods section.

L265: “low intensive” I think is a typo.

Thank you for your comment. It will be changed to “low-intensity agricultural system” in the revised version of the manuscript.

L285: Less uncertainty in the three-pool model does not necessarily mean a more accurate estimate.

Agree with the comment. We will remove this sentence and focus on the three-compartment model's greater ability to jointly represent SOC and radiocarbon data.

L340: “The observed decline resulted from destabilization...” I think this should be more explicitly linked to the model-based evidence such as: “In the fitted three-pool model, the observed decline was represented by...”

We agree with the reviewer. We will modify the text wherever necessary to limit these results to interpretations derived from the model.

L435–440: “can only be explained” It is still under the specific model constructions that were tested.

Same response as in the previous case. We will revise the text as necessary to present the results as model-derived inferences that depend on the model structure, assumptions, and other defining characteristics.

L496 this study found increases in POC and MAOC at 0 - 10 cm but decreases at 10 - 20 cm following no-till. It would be useful to discuss more what the depth redistribution means for the observed measurements.

We appreciate the reviewer's comment; we will explore this point and include it in the discussion section.

509-510 “did not allocate that C into persistent compartments”. There should be more consistent language that specified this is from the model.

We will revise the text to make it clearer that this conclusion is an inference derived from the model and not a direct observation.

L534 we should "minimize tillage interventions and maximize C inputs to minimize the loss of previously stabilized C." But the paper shows that SOC and d14C declined after the shift to no-till in 2009, so this should be more clearly linked to the study's results and interpretation.

The reviewer is correct in noting that this text does not fully align with the study's results and contains logical inconsistencies with the rest of the study. Based on the study's results, it can be interpreted that high rates of C input are necessary to sustain the previous increases in SOC stocks during the recovery phase. However, the results do not demonstrate that the shift to no-till caused the decline in SOC after 2008, but neither do they support the conclusion that this management change alone was sufficient to prevent it.

We appreciate this comment and will modify the text in the revised version of the manuscript, as it is overly general and not sufficiently restricted to the results of this study.