Response to reviewer 1 (R1): 'Simulation of soil carbon dynamics in Australia under a framework that better connects spatially explicit data with ROTH C'

Authors: We thank R1 for taking the time to review our manuscript. Below we respond to each of the comments made.

R1: This manuscript presents a simulation work on soil C dynamics using the RothC model over Australian croplands and grasslands. This topic is within the scope of the journal. The manuscript has a strong potential, as it uses a large and continental-scale set of plant and soil data for model parametrisation, simulation and prediction.

Authors: We thank the reviewer for the comment and for acknowledging that our manuscript is suitable for Biogeosciences.

R1: However, the manuscript suffers from important issues of orientation of study objective, modelling and redaction, rendering the nice dataset not well valorised.

Authors: We do not entirely understand this comment. The 'orientation...study objective and redaction' of our manuscript is to use the ROTH C model initialised with measured C fractions and using a framework to simulate soil C change under different land uses across Australia. There is no clear indication of how the reviewer thought we should better 'valorise' our dataset. Below, we summarise our intentions and hope that the reviewer might now understand the value and novelty of our research.

- We demonstrate the simulation soil organic C across Australia with the ROTH C initialised with measurements of the particulate, humus, and resistant organic C (POC, HOC and ROC, respectively), under a framework that enables the synthesis, processing and standardisation of measurements and data, and predictions.

- We initialised the model site-specifically (across Australia) with the measured C fractions and optimised the DPM/RPM ratio (also site-specifically) because we believe that these are essential to accurately represent the baseline soil organic C stocks and composition across different land uses. This is crucial for the model to be used with confidence and for predicting changes in C stocks and the potential of soils for C sequestration. We have not seen this approach reported in the literature.

- We showed that our simulations, with the model initialised as above, and using a 'standardised' modelling framework, accurately predicted the baseline soil C stocks and composition in the 0–0.3 m layer cross cropping, modified grazing and native grazing sites across Australia. Our predictions across natural environments were less accurate, but as we say in the manuscript, that was to be expected.
- We then used the model to perform a 100-year simulation and showed that with an annual increase of 1 Mg C ha^{-1}, the potential to increase organic C stocks, as well as the potential vulnerability to C loss, in Australian soils is smallest in soils under natural environments, larger under cropping and modified grazing, and the greatest in the soils under native grazing.
- Finally, we identified the soil and environmental controls on **the predicted changes.**

As stated, perhaps our aims didn't capture well our intent. We can improve them.

R1: In the manuscript, the proposed framework that allows bridging dataset and the model plays a central role in driving the study's storyline (see LN1-3 as the beginning of Abstract, LN54-64 as the key sentences for knowledge gap identification in Introduction and a whole Section 4.2 related to the framework). Too much emphasizing the framework makes the manuscript very technical, rather than scientific.

Authors: We disagree with this comment. From our perspective, the flow or our argument is clear: we researched C dynamics across Australia using the ROTH C model, described the framework under which the research was performed, and the experiments and simulations. Finally we describe and then discuss our findings around the potential C increases in Australian soils under the different land uses.

We find the comment that the manuscript is 'very technical, rather than scientific' quite 'unfair'. Technologies/methodologies and 'science questions' are intimately linked. Yes, we agree that the latter are important, but we argue that how we arrive at the new science is equally important. It is critically important to explicitly describe the 'technical' aspects if we are to improve the quality of the science. This might be specially so in our case because of the complexity and scale of our research. We ran a model and simulated C dynamics over all of Australia, with thousands of measurements, with many continental-scale and different datasets and testing different aspects of the simulations.

In the literature, we often see studies that quite generally describe the model used, the input data and the simulation. This might be fine if the modelling is being performed at a single, or few sites where the datasets used are relatively small and originate from one or a few sources. However, for large-scale, complex simulations one needs to know exactly how the various data are connected with the model. Additionally, and in support of our argument, we note that the development of simulation frameworks for reliable C predictions are recognised as a current topic of research (for example, see Smith et al., 2020). Therefore, we believe that the description of the framework is necessary.

However, if the general perception is that the technical–science balance is not quite right, we would be happy to improve the title and aspects of the manuscript to emphasise the 'science questions'.

R1: First, there are no alternative frameworks presented as a control for comparison, so the advantages and drawbacks of the framework cannot really be validated. Then, the novelty of such a framework is unconvincing. There are numerous studies on soil carbon modelling performed at regional, national or bigger levels in literature. In such a kind of study, gathering, synthesizing, processing and standardizing large-scale climate, plant and soil datasets from diverse sources are usually common and necessary steps in the modelling process. The flow chart in Figure 1, as well as the associated discussions does not seem to be particularly special or innovative to what has been routinely done in literature. Therefore, selling a framework makes the manuscript scientifically weak and structurally unbalanced.

Authors: We agree that the framework that we used isn't particularly special. However, we argue that explicit description of complex and large-scale simulations need to be undertaken

under one such framework, and that it is critical that it is explicitly described: the datasets, how they were prepared and used, the experiments, the simulations, etc. Hence, we do not see why we would need to compare our framework to other frameworks.
We do not know of any other literature where a C model was initialised with measured C fractions with a dataset like ours and over a large scale. If there is other similar published work, we regret to have overlooked them, and will appreciate it if the reviewer made us aware of those. We believe that the quality and validity of our science is enhanced by explicitly describing the framework (including Figure 1). Our work addresses important scientific questions pertinent to Australian soils: e.g. how to represent the current soil C stocks and composition in different environments under different land uses; what is the quality and amount of C inputs required to maintain the current C levels; what is the potential to change soil organic C stocks with different C inputs.

R1: In parallel, the nice dataset over the continental scale could have been used to address very appealing ecological/agricultural questions, such as impacts of land-use and grazing on the long-term soil C fates. The manuscript indeed presents some figures (Figures. 3-5) including these treatments, but there are no scientific questions driven behind and no associated knowledge gap could be found in Introduction. Nor were these effects fully discussed in the current version of Section 4.1 of Discussion, which is, again, fairly technical and, in most of time, centred to the model.
Authors: We thank the reviewer for the comment about our data. We found the comment a little unclear. It seems that the reviewer acknowledges that our work contains important scientific results, but suggests that we clarify our intent and better describe the science questions and knowledge gaps (?). If this is the case, as we stated in our previous responses, yes, we can do that and also improve our discussion in section 4.1. However, we believe that there is scientific value in what we did (both in terms of the framework and the understanding gained), and we regret that our intent was misinterpreted.

R1: As one of the most famous pool-based models, RothC itself and the associated modelling skills are well-documented. In my opinion, given the nice dataset, it would have been more original to focus on specific questions about land management than on the model or a "framework".

Authors: Yes, the ROTH C model is well-known. We acknowledged this in the introduction. We also acknowledged previous research on its use in Australia. Our work is original, novel and unique because of how we used the model with the measured C fractions, other large-scale environmental and management datasets, and under an explicit framework to answer important questions about the potential of soil C capture under the main land uses in Australia.

R1: The model's initialization procedure (LN160-163 and LN169-171) needs to be clarified too. It is well-known that the settings of initial relative sizes of soil C pools have a huge impact on the final outcomes. If I understand, at time 0, the authors used site-dependent (presumably observed?) carbon quantity with relative pools sizes corresponding to those at their theoretical equilibrium condition provided by the model, right? It is not very clearly said in the text. Have the simulated C dynamics or changes ever been compared with those (presumably?) measured at the 73 sites from 1991 to 2000 (LN161)? This may be a good manner to check and validate the "equilibrium condition" hypothesis, which has been considered strong and untrue by increasing studies.

Authors: We thank the reviewer for the comment. We can now see that section 2.3.4 may be difficult to follow. We can work to improve this. At each of the 4431 sites, we initialised the model's C pools with the measured C fractions, POC and HOC. Because our data represent a single time period, we made the equilibrium condition assumption using an independently collected dataset from Astralia's National Carbon Accounting System (NCAS). This dataset comprises C measurements at 73 sites across Australia mostly made twice, at intervals ranging from 2–20 years. Based on this assumption, we ran the model to optimise the unknown C inputs over the 100-year period. We thought that this was clear, but we can certainly improve the description in a revision. As we wrote in the manuscript, this equilibrium assumption may be untrue at some sites, but we do not actually know—there is no other data to confirm this.

From the data that we have, and our understanding of our study sites, we believe that our assumption isn't unreasonable for sites under continuous, long-term cropping and grazing and those under natural environments. We are not aware of any study indicating transient conditions of temporal changes in soil C stock over large scales. So, we assumed equilibrium condition based on the NCAS data and validated the baseline soil organic C changes simulated by the model.

R1: It is a very good idea to carry out an uncertainty analysis to test the impact of biomass DPM/RPM ratios on model results (LN166). However, the choice of biomass DPM/RPM ratio (LN184) is disputable. Despite some plasticity, a species' DPM/RPM ratio shall be quite stable depending on its taxonomical and functional features. For example, legumes which are richer in N and lower in C:N ratio shall have generally higher DPM/RPM ratios than grasses (e.g., rice, wheat…). But the authors' manner of choosing DPM/RPM ratio ("… selected the DPM/RPM ratio based on the minimum deviation of TOC"; see LN 184) may risk picking unrealistic values for species. This is because the model's fit/bias is not only dependent on biomass DPM/RPM ratio, but also on the settings of relative soil C pool sizes at time 0 (whose influence on model fit is even much more important). Therefore, it would be more reasonable to choose species' DPM/RPM ratios according to the literature data on plant decomposition traits (even though they were not published in Australian contexts)

Authors: We tested six different values of the DPM/RPM ratio (0.67, 0.96, 1.17, 1.44, 1.78 and 2.23) to assess the sensitivity of the simulated TOC, POC and HOC to this parameter (section 2.3.5). Those values are not unrealistic. They are within the range of values used by Janik et al. (2002) in a sensitivity analysis (of NCAS data) performed under Australian conditions. Yes, the DPM/RPM ratio represents the potential decomposability of the incoming plant material and may be a function of its taxonomical and functional features at a species level. However, this ratio may differ within a species, for instance when representing different crop cultivars. Unfortunately, there aren't many studies that report direct species specific measurements of this ratio or how we could use plant biochemical properties (N and lignin) to link them to this ratio. It is realistic to account for possible parameter values and to check model performance. The performance of the model depends on the quality and quantity of C inputs, as well as the current (initial) soil organic C composition.

R1: An additional uncertainty analysis would always be appreciated to test the amplitude of impact of chosen DPM/RPM ratios with variance for a given setting of relative soil C pool sizes at time 0.

Authors: In this case, we think that a sensitivity assessment, like the one we performed, is more suitable for this particular analysis. The reason is that we used this to optimise the model. See Table S1 in the supplement for the range of C inputs and soil organic C stocks by the chosen ratios.

R1: The removal of the 388 sites may need some more justifications. Why did these sites (not the others) yield such unrealistic values? Why 10 Mg C/ha as the threshold?

Authors: Based on the NCAS data, the range of annual change was found to be up to 10 Mg C and we selected this as the threshold. We described the conditions that need to be satisfied for the modelled soil organic C to reach equilibrium. For these 388 sites, one of the dynamic pools, either POC or HOC, failed to be constant with time. To clarify, we could add the following sentence: "We considered 10 Mg C ha^{-1} as the threshold based on the range of measured annual changes in TOC." As described in the manuscript, the 388 sites were characterised by large TOC stocks (median 75.04 Mg C/ha). However, we do not know why these sites had such large changes in the dynamic pools—we cannot say whether or not these are unrealistic. One possibility is that the pool composition of large organic C stocks is not fully constrained by the decay rates and environmental factors calibrated in the model (see Figure 6).

R1: LN 209: Which model type did the authors choose for the test of the environmental factors? How do these factors cross or nest among each other? And how did they treat in the model? Additional information about this may be represented in Materials and Methods and Figure 6. When looking at the Figure 6, it is not surprising to see that Clay, MAT, MAP and PET stand out, as they are all directly or indirectly involved in the model as key inputs/parameters. What would be much more meaningful is to do the same test over the residuals (i.e., measured C changes minus modelled C changes for the 73 sites 1991 to 2000). This helps see which environmental factors should be further taken into consideration by the model.

Authors: We used the regression tree model, Cubist, which we described in section 2.3.6, lines 200–210. In a revision, we could provide more details on Cubist, but we note that this algorithm is well described elsewhere and we provided citations for the interested reader. We used Cubist to determine which environmental factors affected the **changes** in C stock of each pool (induced by different C inputs). This should be clear from the subheading and text in section 2.3.6 'Empirical assessment of controls on the simulated C change'. Here, we are not interested in looking at the factors that affect the residuals from RothC. Therefore, in our case, because we modelled the changes, the analysis suggested by the reviewer isn't relevant.

R1: Overall, the manuscript tackles a timely and important question in the current research context. However, due to the unconvincing scientific orientation, unbalanced structure and ambiguous modelling procedures, I don't think the manuscript is mature enough to reach a reviewable condition. I sincerely suggest the authors make good use of such a nice dataset and rework on the question and modelling processes.

Authors: We thank the reviewer for acknowledging that our research is timely, but we disagree that our work has an 'unconvincing scientific orientation'. We hope that our responses above, have helped to clarify the intent and novelty of our research. We also disagree that our modelling and simulation is 'ambiguous'—on the contrary, the thorough and explicit description of the simulations under the framework that we described, reports our approach in a transparent manner. We also disagree that the manuscript is too technical—it is essential to explicitly (and technically) describe the modelling approach for the simulations of soil C dynamics at such large scale to have real meaning. We do agree, however, that we can clarify and improve aspects of our writing.

The reviewer comments on the 'maturity' of our manuscript but does not specifically suggest why he/she thinks it so. We hope to have addressed his/her main concerns regarding the scientific value of our work. We also hope to have clarified the relatively minor comments around the simulation procedure. Certainly, we can improve clarity in a revision.

The reviewer also suggests that we ‘make good use’ of our dataset, but he/she does not say what a better use might be. Thus, the purpose of the comments isn’t’ clear. Our manuscript describes, for the first time, simulations of soil organic C across Australia with the ROTH C model site-specifically calibrated with measured organic C fractions and sensible DPM/RPM values across different land uses. Using the well calibrated model, we demonstrated the potential effects of changing C inputs on the changes in soil organic C stocks and its pools. We found that with an annual increase of 1 Mg C ha^{-1}, the potential for C sequestration, as well as the potential vulnerability to C loss, in Australian soils is smallest in soils under natural environments, larger under cropping and modified grazing, and the greatest in the soils under native grazing. The simulations across Australia were performed under a framework that establishes a much-needed connection between measurements, datasets and the model. It enabled consistent processing of measurements and datasets from different sources, and standardisation and configuration of the model for calibration, verification, and prediction.

Finally, we thank the reviewer for reading our manuscript. Although we do not agree with many of the comments in the assessment, as we say above, we see that we could improve the clarity of our intent and also emphasise the significance of our findings.