General Comments –

The manuscript submitted by Piccoli et al. presents the results of a 54 year long field experiment, with a focus on soil carbon accumulation in response to residue management treatments. Using mixed models, the authors attempt to understand how residue incorporation drives soil C and N stocks and concentrations, and then apply a series of empirical models for determining soil C saturation to explain their findings. Additionally, they examine the time-series of data stretching back to 1966 to understand the long-term impacts of the residue management treatments. The authors find that residue removal did significantly decrease SOC stocks relative to residue incorporation treatments, though the general trend of the SOC stocks since the experiments inception has been a loss. They find little evidence of saturation across the models they employ. Overall, the authors present an interesting and important dataset that should continue to be investigated, as long-term SOC datasets continue to be sparse. There are several issues that bar the manuscript from publication in its current form, however. The writing style is often overly casual, and does not adhere to typical paragraph form, making the reading of the MS somewhat confusing and choppy. In addition, the methods are not plainly described, leading to a difficulty on the part of the reader to evaluate the authors results. This is most relevant in the discussion of sampling, the calculations of the saturation values, and the rationale for the statistical tests that were employed. The results section is relatively brief, while some data, such as the time-series results, are not introduced at all until the discussion. Other data that are relevant the the conclusions drawn, such as the impact of N rate or the interaction effect between N rate and residue treatment, are not presented at all. The concluding statements of the manuscript introduce new concepts as well, not previously discussed throughout the paper. Finally, the MS is either incomplete or contains errors, as Table 1 is not included in the text, despite in-text references. I've detailed my specific comments below.

Specific Line Edits-

Lines 28 – 30: Do you think this assertion holds up across all ecosystems? I agree that the mineral-associated fraction has a finite upper limit on its accumulation, but see Cotrufo et al. 2019 for evidence of particulate organic matter accumulation that lacks a saturation limit. We thank you for suggesting such an interesting paper that will be promptly included in the revised MS version.

Lines 39-43: There's been a lot of recent work on saturation behavior in soil carbon (e.g., Georgiou et al., 2022; Heckman et al., 2023) that is relevant to the present study. I recommend the authors broaden the scope of their literature review in this section to include both the foundational and contemporary literature. We thank you for suggesting such interesting papers that will be promptly included in the revised MS version.

Line 53: There was a rapid shift in subject here that was hard to follow as a reader. I recommend the authors consider restructuring this paragraph to better communicate the ideas presented. We will revise the text, as suggested.

Lines 89-91: Please describe the overall tillage management for each plot here. It is currently unclear how tillage is handled between treatments – does the residue removal treatment receive the same level of cultivation as the incorporation and manure treatments? If not, please describe how this has been accounted for in the resulting analysis. Yes, the reviewer understood correctly as all plots are moldboard ploughed at the same time. We will specify better that all plots receive the same cultivation intensity.

Lines 99 – 103: I found the description of sampling here to be confusing. Two things in particular stand out: First, did the authors collect one sample per plot? Right now, that is not clear from the text, which implies that 60 samples were taken per plot. Second, were two sets of samples taken, one for bulk density and one for elemental analysis? Or was the original sample used for both analyses? We thank the reviewer for pointing out this issue. We will modify the entire sampling section clarifying the sampling procedure. In brief, the soil sampling involved two different samplings. 1) Undisturbed soil cores (7 cm diameter, 60 cm height) were collected from the middle of each plot using and hydraulic sampler, cut into distinct layers (0-30 and 30-60 cm) and oven-dried for bulk density determination. 2) A series (4/5) of remoulded soil sampling were collected from different parts of the plot (avoiding borders) to form a single bulk soil sample for each plot and sampling depth (0-30 cm and 30-60 cm) using a hand-push auger. Afterwards, the 120 (60 plots x 2 depths) soil samples were air-dried and subjected to chemical analysis.

Line 111: I believe that the experimental design the authors have described is a randomized complete block, split-plot design, which residue management the main plot effect, and N rate treatment as the split-plot effect. Did the authors consider this when designing the error structure of their mixed model? More directly, shouldn't there be an additional random effect that specifies the nested structure of the design and specifies random intercepts for the block/residue/N rate combination? We thank you for the comment, we performed an ANOVA for the split-plot design with the residue management as the main plot, N rate as the sub-plot and block effect as random. We will clarify better the text and check through the entire manuscript for consistency.

Line 120: As mentioned above, several recent papers have provided updated means for understanding soil C saturation and saturation deficit (Georgie et al., 2022 in particular). Please provide some rationale for the use of the Hassink and Dexter models in lieu of the more contemporaneous approaches, either in response or in the text. We arbitrarily decide to use the most used (cited) models. The number of citations does of course not necessarily means the paper is better than another one but this is a standard practice literature review. We thank you for the paper suggestion, unfortunately with only "Georgie et al., 2022" and any other identifier (e.g., doi) we cannot find any relevant paper with that name. If the reviewer refers to Georgiou et al., 2022 10.1038/s41467-022-31540-9 please see the reply to Reviewer #1 comments.

Lines 123 – 125: Please show the equations for the saturation models that you employ here such that they can be referenced during the reading of the results.

Ok, we can add the equations during text revision.

Line 136: Is there an established agronomic optimum N rate for this site? If so, does it inform at all the relationship between SOC and N rate? The studied N range (0-240 kg N/ha/y) covers the commonly used N fertilization level in the area which depends on crop type.

Lines 170 – 175: Is there a description of how the authors either isolated or estimated the different soil particle size class distributions used in these models? Please describe this process and include in the methods section. The soil particle class were analysed with laser diffraction and converted into pipette using the algorithm reported in Bittelli et al., 2022 https://doi.org/10.1016/j.geoderma.2021.115627. We will add this information to the text.

Line 184: None of the time-series data is presented in the Results section, which makes its introduction here somewhat surprising and confusing. I encourage the authors to add these data and the results that they glean from it to the results section prior to discussing it here. We understood the Reviewer's point of view and we will add those data descriptions in the results section.

Line 195: I'm confused as to why the x-axis on this plot extends to 2030. If the authors aren't projecting out to that date, I suggest that reformat the axis to represent the data that they have available and are presenting.

Ok, we can modify the x-axis.

Lines 198 – 199: Please provide further explanation as to your assertion here, that a lower CN ratio in 1966 is evidence that the system was out of equilibrium. Generally, very low C/N value refers to high mineralization activity, however, we can remove this half sentence for more clarity.

Lines 209 – 211: I'm a little confused here – are the authors suggesting there is a biophysical limitation on the ability to maintain carbon stores over time? Unless I'm misinterpreting, does this imply that agricultural systems are restricted from reaching a steady state, even over the course of decades? Agricultural systems usually reach a steady state only after decades (e.g., more than 30 yr). What we would stress here is that, as reported by Berthelin et al., 2022, the practice of adding C input with agricultural biomasses is not so efficient because requiring about 10 times the weight of what is expected to be sequestered as SOC. We will re-phrase the sentence for more clarity.

Lines 224 – 228: This is an interesting point, but I'm not sure how it's relevant to the question at hand regarding the potential of reaching the 4 per mille goal via residue incorporation. I recommend the authors make the linkage more explicit here. Further, I'm not sure this accounting is fully inclusive. Given the results from above, that over the course of the experiment residue removal has lost significantly more carbon that residue incorporated treatments, would you still arrive at this results if you factored in the lost soil C in addition? I could be misunderstanding the math the authors use here, but I would appreciate their clarification. We got the point of the reviewer. The use of crop residue for bioenergy production should have saved 0.74 t C ha/ y x 50 years = 37 t C/ha while their incorporation led to a delta of ca. 4 t SOC /ha, 10-fold more CO_2-equivalents than SOC accumulation by incorporation. A correct comparison between different crop residue usage would be necessary, anyway, we wish to stress that different types of use of residue can be conceivable.

Lines 232-233: Please clarify this statement, as in lines 135 and 154 the authors state that N rate does affect both C concentrations and stocks. We will clarify better this paragraph.

Line 244: Where is Table 1? The authors have not included it in the present manuscript, limiting the interpretation of the results they discuss here. The reviewer is right, we will add the table in the revised version.

Line 264 – 267: The authors here are presenting the discussion of carbonates as though this is a central finding from the manuscript, but this is the first time that it is discussed. Additional text is needed in the discussion section for the reader to have the context necessary to interpret this. We will add additional text, as suggested.