The authors have conducted an experiment in Gulu, Uganda in which the effects of various conventional and conservation agricultural treatments on crop productivity, soil N dynamics, and N2O emissions are analyzed. The study is conducted over two cropping seasons within a single year – April – October 2023 and October-Jan 2024. Plots that had been fallow for the previous 3 years were cleared and prepared in an RCB design with four treatments: conventional maize monocrop (ConMM), conventional pigeon pea/maize (Con) rotation, conservation agriculture (CA), and conservation agriculture with biochar ammendments (CA+bc). Some differences were observed in NH4 but not in NO3. Maize grain yield in season 1 was lower in conMM than pigeon pea grain yield in the other treatments; there was no difference in grain yields in the second season. Yield-scaled N2O emissions were higher from the conventional than the CA treatments in the first season, but there were no differences in the second season.

This study adds much-needed measurements of N cycling and N2O emissions from low/no-input cropping systems in Uganda, and I look forward to seeing them in the peer-reviewed literature. I do have several general comments.

Major comments:

A major limitation of this study is that data are only presented from a single year (which encompasses a long-rain and a short-rain cropping season). Interannual variability can play a big role in crop productivity and N cycling (highlighted by the statement in lines 548-550 that drought drastically reduced maize grain yield in the first season), and this type of study would benefit from more than one year of measurements. It is true that two cropping seasons are included, but these are different cropping seasons – long rains and short rains—and cannot be said to represent the same thing as multiple years of measurements. The fact that there are only one year of data presented here and, as a result, it is not really possible to generalize really needs to be emphasized front-and-center in the text. (There are so many examples in ecological studies of a second year of data drastically changing results that it's become cliché to joke that "I should have quite collecting data after one year!") That said, 1 year of data can still be quite useful, and I think it's particularly important to have these data published because there are so few data from the region, and particularly few data from pigeon pea systems. And I am especially sympathetic to the tremendous amount of effort required to establish these experimental plots and to collect and analyze these data. I leave it to the editor to decide whether a revised manuscript would meet biogeosciences' editorial threshold for publication.

On a somewhat related note, the authors' suggestion that treatment differences were not expected in this first year (e.g., Line 540-542) would seem to raise a fundamental question: if treatment differences were not expected, then why were measurements conducted in the first year, and why is it interesting to publish them? A way to address this concern would be to cite examples from the literature where treatment differences (when similar treatments were compared) *were* found in the first year of establishment.

Other general comments:

I don't understand why chamber position is included as a treatment in statistical analyses. If there were some targeted analyses into mechanism then it might make sense, but in this manuscript the central questions are how the agricultural management treatments differ, and how soils change over time within an agricultural management treatment. To answer this question, the weighted mean of the within row and interrow chambers would be the appropriate response variable—i.e., an estimate of the flux from each plot.

I think there's a missed opportunity for more discussion of what makes this study unique--the focus on unfertilized systems. Currently this is not discussed much at all. There are actually quite a lot of N2O data from unfertilized maize in tropical African systems that could provide context for understanding the emissions in this study– the control plots from fertilizer trials. Off the top of my head, there are studies in East Africa with lead authors that include Baggs, Hickman, Millar, Pelster, Rosenstock, Tully, Zheng, and studies in Zimbabwe from Mapanda. Typically there are multiple studies from each author, and at least some of them include explicit calculations of yield-scaled emissions.

Initial soil sampling for NO3- and NH4 analyses was conducted after treatment application. Since there were no NO3- and NH4+ measurements before treatment, these measurements are able to answer one question: how treatment affects changes in NO3- and NH4+ snap shots when comparing the start of the first crop season to the end of the second crop season of a given year. I think it would be very useful to include some discussion/contextualization as to why this is an important question—it's not immediately clear to me.

Specific comments:

Line 40: delete “tropospheric”

Line 42: N2O is actually the largest anthropogenic driver of stratospheric ozone depletion (see Ravishankara 2009 in Science DOI: 10.1126/science.1176985)

Line 44: replace ‘atmospheric’ with ‘anthropogenic’ and change to “60% to global **anthropogenic** N2O emissions”

Line 60: you can delete ‘or by-product’

Line 64: also labile carbon availability (‘substrate availability’ would most likely be interpreted to refer to nitrogen compounds)

Line 137: indicate in the text when specifically the clearing was conducted and when glyphosate was applied.

Line 138: I’m not an expert in herbicides, but I think glyphosate can potentially have a range of impacts on crop production (via uptake by crops either from glyphosate that has persisted in soil or is released from decomposing weed biomass; it also has potential interactions with mineral nutrition: see https://pmc.ncbi.nlm.nih.gov/articles/PMC6918143/), which makes the timing of its use useful to know, and which might have implications for season 1 production.

Line 139: indicate in the text specifically when the plots were tilled

Line 144: ConventMM has not been defined at this point. The CA+BC abbreviation has also not been introduced—it was introduced as CA+biochar.

Line 157-158: I’m a little confused by plant spacing. In line 141, it sounded like the spacing was 70cm x 35 cm, which would yield a planting density of 40,898. Here it says 10cm; if that’s 10cm x 70cm, that would be a planting density of 143,143. And were the same planting densities really used for maize and pigeon pea?

Line 175-176: two different biochar treatments are described here, including one that isn’t defined or mentioned anywhere else in the manuscript (CA+BC+BC).

Line 213: Be specific about when the collars were installed (you can say something like ‘at least xx days before the first measurement).

Line 210-211: I don't think I quite have a complete description of the chamber top—it sounds like it should be a pvc pipe – is it a pipe that was manufactured with a top, or did you have to add and seal a top to the pipe?

Line 218: citing a reference here for using petroleum jelly would be great. It sounds like it would work, but a reference would be helpful.

Line 220: were the initial measurements really taken at 1 minute? (i.e. sampling was delayed until the chamber had been sealed for a full 60 seconds)

Line 220: 60 minutes is a fairly long period. Can you establish that N2O did not reach or approach equilibrium in the chamber?

Line 228: Did you have any checks to make sure there wasn't any sample contamination/leakage during shipping? (e.g., having vials filled with a standard gas in Uganda and then measured in Norway.)

Line 235: how was it determined when a linear or when a polynomial fit should be used? (i.e., it would be good to make sure overfitting was avoided; the additional parameter in the polynomial model can provide a closer fit to the data even if it is inferior to the linear model)

Line 242: a quadratic fit does not seem like an appropriate way to calculate a flux here, and makes me a bit worried about the polynomial fits in general. If concentrations are convex downward, that suggests to me that concentrations in the chamber have equilibrated with soil over the 60-minute period, and the resulting flux calculated would not be a good estimate of soil fluxes in the absence of a sealed chamber. If polynomials are fit to data in for which the slope between 30 and 60 minutes is smaller than the slope between 1 and 15 minutes, again it sounds like you're seeing a chamber artifact where concentrations in the chamber are approaching equilibrium with concentrations in the soil, and you might be better off using data from 1 to 30 minutes.

Line 249: Instead of 'scaling up' I think it would be more precise to say that you're estimating a flux that is representative of the entire plot, and it's being done by calculating a weighted mean of fluxes from the basin and interrow chambers. You could write something like "We estimated a flux representative flux for each plot (N_2O_{plot}) by calculating a weighted mean of fluxes from the basin and interrow chambers. We used weighting factors of 0.12 for basin . . .".

Line 259: indicate when extractions were conducted relative to sampling (e.g., x hours, the next day, etc)

Line 280: I would change "dividing the scaled cumulative N2O emissions . . ." to something like "dividing the weighted mean N2O emissions" or "dividing N_2O_{plot} emissions. . ."

Line 281: I would change "scaling factor" to "weight"

Line 283: I would change "scaling factor" to "weight"

line 293, you indicate that random effects were introduced to account for repeated measurements. I need more justification for this: typically, one assumes that repeated measures of the same plot will be correlated to some extent. And since statistics are only conducted on cumulative variables (which is perfectly reasonable, and a good way to avoid issues of autocorrelation in repeated measures), I'm not sure why time is discussed as an effect at all.

Line 296-298: I find it rather remarkable that no data transformations were required (I've never seen field data like this that didn't).

Line 333: change "were" to "when"

Line 349: change "second respectively" to "second season, respectively"

Figure 2 & 3: I think you want to combine these into a single figure—I don't think it's necessary to compare Conv to convMM in one figure and then compare Conv to the CA treatments in a separate figure, especially since you're presenting post-hoc contrasts.

The higher nH4+ in convMM than all other treatments in season 1 seems unexpected, and (maybe) higher or equivalent to the CA pigeon pea treatments in both years? Something that may be worth discussing.

Line 388: Delete the text "Fig. 4" -- it does not present cumulative emissions

Line 392: Interpreting the treatment main effect in the presence of a significant treatment x position interaction (which says that the treatment effect depends on the chamber position) is complicated—another reason to use the weighted mean flux for each plot as the response variable. But I do believe interpretation of a significant main effect in the presence of an interaction in an RCB design does indicate a main effect that is over and above the treatment x block interaction.

Figures 5 and 6: I think you want to combine these into a single figure—I don't think it's necessary to compare Conv to convMM in one figure and then compare Conv to the CA treatments in a separate figure, especially since you're presenting post-hoc contrasts.

Table 3:

- It would be helpful to have a reminder in the Table caption that the weighted mean fluxes were used to calculate yield-scaled emissions.
- I would add something to the discussion explaining why we would care about N-yield scaled emissions.
- Why are the season comparisons (capital letters) only included for ConventMM and not the other treatments?
- I would explain in the table caption why ConventMM is not included in the treatment tests (lower case letters) in Season 1 (presumably because comparing maize yields to pigeon pea yield isn't a useful comparison).

- I think it would be useful to include the P values rather than “ns.” It provides the reader with more information and context for interpreting the results. A P value of 0.08 is very different from a P value of 0.5

Line 434: Again, and throughout the manuscript, I would report the actual P value rather than P>0.05 or P<0.05

Line 434-443: I would just make it explicit here that you are not including ConvMM in treatment comparisons for season 1 since maize was grown in that treatment rather than pigeon pea, as in all the other treatments.

Line 442: “significantly higher” needs to be changed: Maize yield was higher, but yield-scaled emissions were lower

Line 461: I think it would be useful to compare emissions to conventional agricultural settings as well—control plots from other experiments in East Africa. It may also be useful to provide emissions—including yield-scaled emissions--from conventional management using fertilizer as context for the conventional and CA plots in this study, while also discussing issues associated with no-input agriculture such as nutrient depletion—how long could these different practices remain sustainable?

Line 472: https://doi.org/10.1029/2020JG005742 could be helpful for the rewetting discussion.

Line 487: much of the fixed N may be in the harvested biomass, limiting the amount returned to soil.

Line 494: I don’t think NH4 is higher in CA systems than conventional systems—isn’t the NH4 in ConvMM no different than the CA treatments? And if ConvMM is higher than conv, it doesn’t follow that the mineralization of pigeon pea residues is responsible for the difference. It also conflicts with the statement in line 486.

Line 514-516: this sentence is missing something, or ‘nutrient cycling’ should be deleted.

Line 540-542: You may want to rephrase this. Ff no treatment effects were expected, it would seem to undermine a justification for the entire experiment (i.e., readers may ask why measurements were conducted in a year when no treatment effect was expected, instead of at a time when the treatments would be expected to have an effect).

Line 566: delete “We established that”