Tropical forest in general are still largely understudied and to understand possible future pathways of these extensive forests, first the current state and drivers need to be understood, which is where studies like this come in play. Lont term data sets of ecosystem and soil fluxes especially of non-co2-greenhouse gases are rare. This study therefore delivers an important and useful contribution to the study of tropical forests. The combination of Eddy covariance (EC) and automated soil chamber measurements for N2O and CH4 in a tropical forest is novel and certainly not trivial to accomplish. The highly variable fluxes with a significant seasonal effect is as expected from n2o and ch4 fluxes in tropical forest.

I agree with referee #1 that the removal of flux values outside the 5th - 95th percentile flux range is a concern. I think it at least needs a strong reasoning and explanation in the text as why the authors chose to do this and which values were removed. Also the sporadic high ch4 emissions and the many n2o uptake events are not discussed, even though they are of interest.

In the attachment I add some comments, a few questions and some suggestions to improve readability and/ or interpretation.

Line 186: automated static non-steady through flow chambers

- ➔ automated non-steady-state flow-through chambers

<u>Chamber measurements</u>

> There are two closure times (2 min and 25 min) and 16 chambers. Could you describe in short how one cycle of all chambers happens? Do the 2 min and 25 min closure time happen shortly after each other for the same chamber, or only specific chambers that close 2 min and other 25 mins? How many measurements per chamber per day? Did you use both closure times in the calculations or did you select one of the two (in caption of Figure S4 is stated that 5 min closure time is used for ch4)? You then link them to the according half hour to match up with the EC measurements. How many measurements per half hour?

Line 212-213: Flux values were selected based on the model that provided the best fit and highest determination coefficient (R2)

> How was the best fit determined?

Line 219-221 : As an improvement over Courtois et al. (2019), all CH4 fluxes with R2 < 0.80 were excluded regardless of the measurement length (i.e. 2-min and 25-min). For N2O, all short measurements (i.e. 2-min) with R2 < 0.80 were discarded.

> If the CO2 measurement during a period is correct, why then still throw out the low R2 measurements for ch4 and n2o? If co2 flux is correct, the low r2 can not be due to analyzer failure or bad closure of the chamber, so then the flux must be correct and might be just small instead of false? Is it not better to put them to 0 instead of remove them out of the data set?

Flux data analysis

Line 248-249 For eddy covariance and chamber data, the 30 min observations were filtered and flux values outside the 5th - 95th percentile flux range were discarded.

> What do you mean with filtered here and why are the values outside the 5^{th}-95^{th} percentile thrown out of the data set? Frequently fast increases (high spikes) after rainfall events are seen in N2O soil fluxes (Daelman et al. 2025 https://doi.org/10.5194/bg-22-1529-2025, Werner et al. 2007 **https://doi.org/10.1029/2006JD007388**), these peaks might be thrown out here.
>
> Also sporadic high emissions for CH4 are measured, which also might be partly removed (again Daelman et al. 2025 https://doi.org/10.5194/bg-22-1529-2025 or Barthel et al. 2022 https://doi.org/10.1038/s41467-022-27978-6)
>
> Also for the EC data: if the data QA/QC is carefully carried out, it seems that there is no need to remove values outside the 5^{th} – 95^{th} percentile. However you could opt for a despiking algorithm like the median absolute deviation algorithm (Mauder et al., 2013; Papale et al., 2006).

Line 249 – 257 To calculate daily averages for greenhouse gas fluxes, we first estimated the optimal number of observations per day necessary to obtain representative daily averages. To do this, we selected a data pool with at least 42 observations per day in the eddy covariance dataset. In the soil chamber dataset, we calculated daily means for each of the thirteen chambers and retained only the data when at least five observations per chamber per day were recorded. Subsets of values from 1 to 42 for the eddy covariance data and from 1 to 13 for the soil chamber data were then created for each day based on 100 bootstrap iterations. Representative daily means were found for thresholds of 12 minimum observations per day for eddy covariance and 10 for chamber data.

> How did you quantify the representativeness?

I think this part is interesting but a bit confusing in the way it is formulated. I suggest some small alterations and I have some questions linked to the way I understand the paragraph. If I just misunderstood the analyses, please ignore the questions that are not relevant.

Alteration: To calculate daily averages for greenhouse gas fluxes, we first estimated the optimal number of observations per day necessary to obtain representative daily averages for EC measurements **and the optimal number of chambers to obtain spatial representative daily averages for chamber measurement**.

- ➔ You differ in amount of chambers in the bootstrap (1 to 13), not the amount of measurements per chamber, which is always at least 5, right? So you want a daily coverage of at least 5 measurements and you are mainly looking for the correct spatial coverage with the bootstrap?

Alteration: **Subsets of each day from the selected data pool were created of sizes 1 to 42 for the eddy covariance data and subsets of daily averages of size1 to 13 for the soil chamber data were then created for each day based on 100 bootstrap iterations.**

Could you include anything about the diel cycle of the EC measurements. Many more night time measurements will be missing compared to day time measurements, so when taking only 12 half hourly measurements, a large percentage of this will be daytime measurements. As only these 12 half hourly measurements are enough to represent a daily average, I suspect that the diel cycle is not that pronounced.

Could you also include something about the spatial variability of the chambers? As you need 10 out of 13 chambers to achieve representative values, does this mean that there is a high spatial variation between the chambers? Or does this mean that you need a lot of chambers because they are al measured in different half hours and you need to cover many half hours and thus the temporal coverage rather than the spatial coverage is important?

Line 331 – 335 In contrast to the ecosystem-level fluxes, soil CH4 fluxes in some of the upland forest were mainly negative, indicating net soil CH4 uptake throughout the year (Fig. 2b), even under varying environmental conditions. Soil CH4 uptake did decreased significantly in the wettest season compared to the driest season, although the fluxes remained negative overall (i.e. CH4 uptake, Table 1; Fig. 3b).

There are however many high emissions periods for several chambers at different times for ch4. These are not discussed in the article.

Line 342- 345 In contrast to the ecosystem-level fluxes, soil N2O fluxes in upland areas not only had a more pronounced seasonal pattern, the upland soils also emitted more N2O during the wettest season than during the driest season, when the average flux was near-zero N2O (Table 1; Fig. 3d).

There are however many uptake events for N2O, these are also not discussed in the article even though soil uptake of N2O is not that frequently measured in other tropical forest.

Line 559 - Measurements at the ecosystem and soil levels showed divergent fluxes, probably because soil fluxes represent only one compartment in the whole ecosystem. Furthermore, upland soils (52% of the footprint area) are only one type of soil within the large range of soils found inside the Guyaflux tower footprint. In addition, soil chambers provide integrated fluxes for a much smaller area than does the eddy covariance technique.

Just a very small suggestion, not really a must, might be something to think about: This is, I think, the only section where the two kind of measurements (soil chambers and EC) are mentioned together. As stated, the soil chambers are only one compartment of the ecosystem and have a small area so I realize a real comparison or upscaling can not be made, but is there some way both methods could be linked more? I am thinking of a small overview or repetition of what is comparable, what is not and which elements may cause the divergence. Or in someway how these soil chambers fit in the results of the EC. I think there are many elements in the text that point to connection or divergence (like uptake of stem/canopy, different soils, SWC, ….) but somehow the connection between the measurements is lost for me because these elements are spread out in the text. Anyway, not easy to do, just a thought.