This study provides an important step towards better and more reliable assessments of CO2 fluxes from LULUCF, as it presents a comprehensive dataset of LULUCF fluxes from country reports to UNFCCC, and it compares this dataset to LULUCF fluxes from other databases (such as the FAOSTAT emission database) with special emphasis on explaining the differences between the investigated datasets. The methods described in the manuscript are clear, the results and explanation of differences are convincing, and I think that the study will be an important contribution for future assessments of CO2 fluxes from LULUCF.

While reading the manuscript, I noted a few points and questions that might be of interest to the authors when revising the manuscript:

- To my knowledge, complete details about the applied methodology is not available for all countries, which further complicates the assessment of the reported fluxes (lines 96-98). It might be worth to note that as well.
- The third reason indicated for the differences between the investigated datasets (lines 99-101) sounds rather general. Would it be possible to explain a bit more in detail what this point encompasses?
- Line 107: Does “reporting” refer here to GHG fluxes? Or to other variables as well?
- I am wondering why gap-filling was also necessary for some AI countries (5%, line 167), given that they seem to report a complete time series (as indicated in line 163). What is the reason for the necessary gap-filling for AI countries?
- When reading the details about NC/BUR, REDD+ and NDC reports (lines 178-196) I was wondering how these reports were used to compile the data presented in this study. Although the information is given before (as I saw when checking the manuscript again) and Table 3 gives a nice and useful overview of the origin of the data, I think it might be worth to start the paragraphs in lines 178-196 by informing/reminding the reader again about why and how the NC/BUR, REDD+ and NDC reports are used here.
- Lines 203ff: Do the countries explicitly report the location of managed land or just indicate the total area of managed land?
I do not fully understand how this technical detail is related to the sentences before. I think it would be good to elaborate here a bit more about this mapping (but only general, as all details are listed in Table 2).

I like that the authors mention and highlight the importance of choosing the correct method for uncertainty estimations. However, from the text I did not really understand what the difference between both methods is. I think it would be helpful to explain this better, and maybe also give an example for each case, such that it becomes clear what the difference is (and why it matters).

Is the increasing number of available NAI countries due to the usage of NC/BUR, REDD+ and NDCs reports or is there another reason? I think it would be good to mention that here.

I think it would be nice to support the statements about the agreement between the different datasets by some statistical measures (e.g. correlation or trend analysis)

Why might the datasets fail to capture the recent increase in deforestation?

I think here it would be helpful to remind the reader again about the methods used to assess whether NGHGI DB or FAOSTAT is more complete (and to refer again to Supplementary Figure 2).

I did not fully understand how the percentages indicated here were obtained. Is this directly connected to the assessment of completeness for NGHGI DB and FAOSTAT? Or is it estimated based on a different approach? It might be worth to mention this more explicitly.

I am looking forward to seeing this study accepted and published in ESSD.