# Answers to RC2

*RC2 original text is italicized and grey*. Our answers are in plain text and black

*I am happy to read this manuscrip from Schwartz et al. This manscript developed canopy height, wood volue density, and aboveground biomass density data products in France using GEDI, Sentinel-1 and Sentinel-2 datasets with a deep learning approach. The developed data products were assessed with multiple independent datasets and showed improvements over previous developed data products. Overall, this study is well organized and the data products are needed in time to support forest structure and carbon assessment in facing climate change. I only have minor comments.*

Thank you for this positive description of our work. We will do our best to answer your comments in the revised manuscript, as explained in the following.

*1. Abstract may also include the FORMS-V, which is one of the three data products develioed in this study.*

Thank you for this suggestion. FORMS-V will be added to the revised version of the manuscript.

*2. Table 1. "In this study" should be "This study".*

You are right. We will modify the revised version of the manuscript accordingly.

*3. Figure 1. The rasterization of 25-m GEDI footprints to 10m grid may introduce some uncertainties to the model and data products. Is there a way to reduce these uncertainties, for example, using more data quality control or GEDI footprints in pure landscape types?*

Thank you for this very relevant comment. In answer to reviewer #1, we acknowledged that this method could introduce some uncertainties. Still, we showed that rasterizing GEDI on a 10 m grid provided better results than rasterizing at a lower resolution, like 20 m. You can refer to it for additional details. As you suggested, several techniques could be used to control the quality of GEDI data and help reduce these uncertainties. We already applied basic filters to the GEDI footprints (e.g., quality flag =1), but other filters could have been used to refine the quality of the footprints, such as taking the night GEDI shots and the full power beams only. However, we assumed here that our deep learning model would be able to cope with these uncertainties, and we chose to keep as many footprints as possible.
We also thought of filtering GEDI data based on landscape types, as suggested. It would avoid the case where a GEDI footprint falls at the border between two landscape types (forest border for instance), thus possibly creating a high uncertainty in the rasterized value. However, our attempts did not yield significantly better results. Moreover, this type of landscape filtration would also remove all GEDI data on isolated trees and hedges, making it more difficult for the model to predict the height of these trees accurately. For these reasons, we finally decided not to use this type of filtration, but it is a relevant matter that should be addressed in further studies. This will be explained in the revised manuscript

*4. Figure 5. What is the reason that the R2 values are so different in Figure 5a and Figure 5b?*

We assume you are referring to the $R^2$ values of the figures 5.1.a and 5.2.a. They represent the validation scatterplots between FORMS-H and two validation datasets: the GEDI Test dataset for Fig. 5.1.a ($R^2$ = 0.33, MAE = 4.48 m) and the French NFI plots for Fig. 5.2.a ($R^2$ = 0.69, MAE = 2.94 m). We can indeed observe here a significant difference between the $R^2$ values and also between the MAE values, not totally expected because NFI plots are more independent and should be more difficult to predict than GED Test data. A visual interpretation of the scatterplots, associated with the histograms shown in Fig. 5.1.b and Fig. 5.2.b, reveals that low heights are poorly predicted in the GEDI Test dataset and thus greatly impact the $R^2$ value. This is rather due to GEDI label errors than an error from the model as long as it cannot be observed in other validation datasets. This issue was already addressed in the original manuscript in lines 228-233. The text in **bold** in the following has been added or modified to address your comment and bring additional information: "*Conversely, FORMS-H indicates higher heights than the labeled GEDI footprints for many areas categorized as low heights (Figure 5.1.a). This discrepancy can likely be explained by the quality of GEDI data, where the labels could be wrong due to atmospheric conditions or geolocation errors.* ***These geolocation errors should normally have a symmetric pattern, with as many points overestimated for lower heights as points underestimated for higher heights. However, as detailed in the figure caption, we plotted only the footprints geolocated in forest pixels of the Copernicus DLT map. Therefore the geolocation errors related to GEDI footprints located outside forests were excluded from this graph.*** *Our comparison with the completely independent French NFI data (Fig. 5.2)* ~~***excludes these types of outliers as it focuses only on forests measured in 2020***~~ ***does not reveal the same outlier pattern because these forest inventory measurements are more reliable and accurately geolocated.*** *It yields a smaller MAE of 2.94 m* ***and a higher $R^2$ of 0.69*** *(Fig. 5.2.a) with a distribution of predicted data very close to the NFI distribution of heights (Fig. 5.2.b).*"