**Reviewer 1**

This manuscript considers the issue of property scaling with respect to spatial scale. In particular, the authors examine how the aboveground biomass distributions in forests change depending on the scale they are measured at. Such information may have important implications for global models that incorporate site-level data and run with pixels covering hundreds of square kilometers.

My thoughts on the manuscript can be summarized in three categories, and I have the feeling they are strongly related to the intendend audience. A journal like Biogeosciences has a bit wider audience than some other journals, which means that folks will be approaching it from different backgrounds. Indeed, this article is meant to address folks from several different backgrounds. As someone with a modeling background, I may have missed what is evident in other fields. To whit:

1) I did not find the problem well-demonstrated

2) I felt that the discussion was not well-developed enough to convey the significance of the results and convince a general reader

3) I was not convinced that the method was successful compared to the case of not applying a scaling factor

Thank you for reviewing our manuscript and for the valuable feedback. In the following, we will address all your points.

For the last point (3), this seems easy to address by showing a figure like Figure 7 but replacing the green historgrams and the blue line with the results of the unscaled distribution. Would this just be the existing green histograms at 50m? If so, I would really appreciate somehow making this more clear (ideally in the figure, but also adding text would be good). I feel like Figure 7 is the figure showing the method was successful, but I do not see that immediately.

Thank you for this comment. The unscaled simulated distribution is indeed the green histogram at 50-m scale. We will add “Unscaled forest model” in the graphic and add text to make this clear.
With Fig. 7 we tried to show the best functioning case in the main manuscript, which was the one starting with fitted parameters at 50-m reference scale. It was impossible to include graphics for all different simulation cases. Thus, we provide their results in tables S1 and S2. To show some detailed graphics about each simulation and result aggregation scale, we provided S3 to S6 (which we will mention more prominently in the text), which are still only the best cases of each reference scale. This already shows, that even for the best cases and using the scaling relationship, the simulated distributions can diverge remarkably from the field distributions.
We did not present a simulation case without applying any scaling coefficient (i.e., applying scaling coefficient -0.5, according to your point 3), since rescaling with -0.5 already led to drifts in the input distributions (AGB gain and mortality), when we tested it against the field data. We will add a new graphic in the supplements to clarify this point.

Figure 7 amazed me. I was shocked to see how the distributions shifted. However, I don't think I should be on page 15 of an article before I'm intrigued by it. I feel like the right hand side shows why this issue is important, and relates to my point (1) above. The introduction of the problem seems to occur in lines 75--7 with a single citation (Wong, 2008). In my mind, this should be an entire paragraph to emphasize the point: "Models which fit biomass distributions at 10 m^2 spatial resolution and reproduce them perfectly at 100 km^2 are incorrect." However, this represents a Catch-22. Phrasing the problem this way makes it much more appropriate for Biogeosciences, but would also require more evidence in the case of land surface models. However, the authors could (and I believe, do) demonstrate this problem with two simple models. Therefore, the information seems to be already present and just needs some restructuring to be more evident and grab the eye of a non-specialist (which is the case with the vast majority of Biogeosciences readership). More citations to the last sentence of the paragraph on line 80 ("But it is often unclear how scale affects observed and simulated distributions"), in particular with regards to forest plot and larger area modeling related to the carbon cycle, would be very welcome for point (1).

Thank you. We will expand the respective paragraph in the Introduction, to state the problem more clearly. We agree that the model results in Fig. 7 appear late in the manuscript. However, the field derived distributions and the scaling relations derived from them already demonstrate the problem earlier. The field derived distributions were essential part of the analysis and necessary input to the model. Hence, we don't see how the models can appear earlier in the story. We will add citations to the mentioned sentence.

For point (2), it was not clear to me why the standard deviations are different. Figures 5-7 show that they are, but I don't understand why this happens. Section 4.2 mentions that different fitting approches had different levels of success, and explains what these fits where, but it does not explore why they had different levels of success. Is there something about the underlying data or problem which means this could have been foreseen from the beginning?

We were not sure whether this comment refers also to the differences between SDs at different scales per se, or to the differences between fitting methods. We will add text to explain the differences between scales and about the differences between fitting methods. Biomass gain was the variable which was the most difficult to fit with the parametric distribution functions, as its distribution apparently conforms the least to the tested lognormal and gamma distribution shapes. However, we do not see how this could have been foreseen.

Minor comments:

Line 80 and 81: Perhaps "extends" should be "extents"

Thank you. We will correct it.

Line 170: It seems that mortality modeling presents an issue with respect the scale. The simulation model chosen resets the area of a whole grid cell to zero. For a 10m x 10m pixel, this could be a single very large tree. For a 100m by 100m pixel, this seems like a larger event. Biomass gain, on the other hand, seems to be similar for every size of pixel (if a 100 m^2 plot grows by 100 g C m-2 yr-1, then either one trees grows like that on a small pixel or it's spread among many trees on a larger pixel). Does this difference in behavior have something to do with why the simulation results change depending on pixel size?

In the original model by Fisher et al. (2008) a mortality event was indeed modelled by setting the biomass of the pixel to zero. However, such a (stand replacing) approach is not applicable across multiple scales. Therefore, we changed the model by drawing mortality from a continuous distribution. We will add text to make this clearer. Thus, this is not the reason for the differences between scales.

Table 2: The number of signficiant figures used seems almost excessive. Is there truly rationale for mean OVL of 0.883 and 0.887? I guess if the error bars on the distribution are taken into account, the mean OVL will fluctuate by much more than that. Although the bins are big enough that the measurement errors are likely small. I would be happy if the authors could confirm this for me (a non-experimentalist).

Thank you. We think it is common to provide percentages with one additional decimal digit, which is equivalent to three digits if given as a fraction. From a practical stand point the third digit allowed us to better select the best fitting method for each case (which was most relevant in Table S2 with several very similar values). We agree that small differences in OVL are not necessarily meaningful or significant and might change if data slightly changes. For this reason, we show the whole table with all the different values to present also the "almost best" cases.

Figure 7: Please add, "On the left are the G and M distributions used as input" or something similar to clarify what the left side of the plot is for the reader.

Thank you. We will add it to the caption.

Line 326: The line begining with "Theory states that the SD" implies to me that there is rationale behind this. I would appreciate this rationale being expressed more in the introduction to introduce the reader to the fact that this is a well-known problem with both observational and theoretical background. Perhaps it is mentioned in the Wong, 2008 reference, but adding a couple sentences would be welcome. The same for the fact that the mean is stable across all scales (line 338), which indicates it really is just an issue with the standard deviation.

Thank you. We will add information about it in the introduction.