**Tree height uncertainty biases aboveground biomass estimation more than wood density in miombo woodlands**

**Review – Miro Demol**

1. **Overall Feedback**
    a. **What I liked**

The Authors present a clear and concise study decomposing the effects on AGB estimations of wood density and tree height data source alternatives when direct measurements are not available. The study uses data from two destructive validation experiments (previously published) in miombo woodlands and applies a Bayesian framework for parametrizing models. The results are readily applicable for carbon quantification of miombo woodlands, for instance for national accounting or REDD+ initiatives. The text is well written and clear, with a good selection of graphs (a few minor specific comments below).

    b. **What should be improved**

Citations in Introductions – For some statements, citations seem superfluous, for others they are lacking. E.g. do you need 4-5 citations for what now is well-established (L40, L43, …)? L88 Réjou-Mechain is I think not the right citation to claim 'almost always assigned from databases' – as it is a methodological paper introducing R Biomass package. L64, do you really need the two Chave citations (which are repeated on L66 again)? But it would be nice to get citations for e.g. L69 (pantropical allometries performing very well on destructive data) or L85 (an example of H-D models overlooking forest type dependence).

Incorporation of other miombo destructive data – The destructive data used here (Handavu et al. 2021, Kapinga et al 2018) is valuable bit fairly limited e.g. in D range (max value of 52 cm). The limited geographical range of the calibration data can cause overfitting and spatial autocorrelation. The leave-one-out cross-validation will help with the first issue, but not with the second. I suggest incorporating more destructive data from miombo woodlands (for instance the Mugasha et al. (2013) dataset that spans 150+ trees ranging up to 110 cm D). Since no data was collected for this study as far as I can tell (the data comes from previously published papers) I would have expected a broader search for reference data in literature.

Circular reasoning in local H-D - The "locally-calibrated species-specific allometry" (Eq. 7) is fit using data from the same sites where the 154 harvest trees come from, and then predicted heights are used in Chave 2014 to estimate AGB for those same 154 trees. The scenario in Fig 3b shows "no bias" but this is partly because the H-D model was calibrated on overlapping/same data. This should be acknowledged as a limitation - the "species-specific allometry" scenario is not truly independent.

2. **Specific Feedback**

| *Line* | *Comment* |
|---|---|

| | |
|---|---|
| L76 | What's causing this bias? Is H usually over or underestimated? |
| L 79 | Should be D instead of H – as in a bias in input D to infer H? Or do you mean underlying bias in the calibration H data to construct H-D models? |
| L98 | It would be good to quantify 'substantial' |
| L163 | Please explain the bias correction factor in the equation |
| L172-177 | I don't think RMSE and bias need a citation as they're well-established model performance metrics. Replacing L172-175 with "We then compared the predictive ability of these four models with the root mean square error (RMSE) and percentage bias (PE):" would be clear enough. |
| L187 | I've got difficulties interpreting Fig 1a. Why is the red 14% bar larger than the 29% bar? |
| L240 | Maybe a figure or table with the 3x3 matrix with the nine scenarios will help to understand these permutations more quickly? |
| L240 | Consider change 'population' to data set mean |
| L320 | This figure could be improved in my opinion. Again as a 3x3 matrix, with shared axis to make maximal use of the space, with on the top the H gradient and to the right the density gradient as titles. Panel (a) is repeated from Fig 2, so I would omit it in Fig 2. |
| L322 | I wonder if coarser attribution level of WD made predictions worse. |
| L367 | There is an artefact in the height data – values center around integer values – or is this intentional? What were the methods to measure tree height in the 4321 trees data? |
| L388 | Apart from tree size, is the forest structure, species composition, etc. sufficiently similar between the ILUA2 calibration data set to apply the ILUA2 model on the data in this study? |
| - | Would it make sense to swap the discussion of the WD and H effects with the Chave vs ILUA2? I.e. get to the point you're making in the title of the paper first? |