Reviewer 2 Audit: Implementation Check from Original Submission to Version 3

Scope: Audit of Referee 2 comments in the response-to-reviewers document against the original submission (version 1) and the revised manuscript (version 3).

c1 = original submission line(s)	c2 = original comment	c3 = remarks about implementation in version 3
L40, L43, L64, L66, L69, L85, L88	Citations in Introduction need balancing: some are excessive/repetitive; others are missing. Reviewer specifically flagged support for pantropical allometries, H–D forest-type dependence, and the claim that wood density is often assigned from databases.	Partly implemented; mostly sufficient. The Introduction is cleaner and better supported in version 3. The statement on pantropical models is streamlined at v3 L60–L66; H–D variation among species/forest types is supported at v3 L79–L82; and the wood-density database statement now includes Fischer et al. (2026) at v3 L83–L86. The most obvious gaps were addressed, though some citation density is still fairly high in places.
L98–L106	Incorporate more miombo destructive data (e.g., Mugasha et al. 2013), or explain clearly why these datasets were not used.	Implemented sufficiently. Version 3 explicitly adds a paragraph stating that other miombo harvest datasets exist but were not suitable because they lack individual-tree wood-density measurements, which are required for this analysis, at v3 L189–L195. This directly answers the reviewer's concern. The broader defense about spatial range is less explicit than promised, but the key methodological justification is present.
L138–L146; L227–L232	Possible circular reasoning in the local H–D model: are the same trees used both to fit the H–D model and evaluate AGB?	Implemented well. Version 3 now states that the H–D dataset contained 4410 trees, excluding the 155 harvested trees, at v3 L200–L208. This directly resolves the concern and is clearer than in version 1.
L75	L76. What is causing this bias? Is H usually over- or underestimated?	Implemented sufficiently. Version 3 now states that field height estimates can be systematically biased, with the height of tall trees often underestimated, at v3 L72–L74. This is a direct and sufficient fix.
L77–L79	L79. Is the issue bias in D, or bias in the underlying H data used to build H–D models?	Implemented sufficiently. Version 3 clarifies that H–D allometries can propagate underlying biases in the H measurements used to parameterize them, at v3 L76–L79. The ambiguity present in version 1 has been removed.
L97–L98	L98. Quantify "substantial" carbon storage.	Implemented sufficiently. Version 3 now quantifies miombo vegetation carbon stocks as around 20–30 Mg C ha^{-1} at v3 L95–L97. This directly addresses the comment.
L163	L163. Explain the bias correction factor in the equation.	Implemented. Version 3 explicitly explains the Baskerville correction factor, defines it, and states that it corrects the bias from back-transforming a log-log model, at v3 L239–L241. This is a strong implementation.
L172–L177	RMSE and bias do not need citations; simplify wording.	Implemented sufficiently. Version 3 now states the comparison more directly at v3 L260–L265, without the unnecessary citations that were present in version 1. This is a straightforward improvement.
L187; Fig. 1a caption L188–L196	Figure 1a is hard to interpret; why is the 14% bar larger than the 29% bar?	Implemented sufficiently. Version 3 adds an explicit note in the figure caption that the bars are not plotted to scale, at v3 L181–L182. This directly resolves the confusion.
L239–L242	Add a figure or table showing the 3×3 matrix of the nine scenarios.	Implemented well. Version 3 now includes Table 3, a clear 3×3 summary of all nine combinations, at v3 L325–L333. Headers could specify similarities between rows/columns though.
L239	Use "data set mean" instead of "population".	Not implemented. Version 3 still says "population-level fit (or fixed effects)" at v3 L313–L315 rather than "data set mean." This is a minor wording issue, but the specific reviewer suggestion was not adopted.
Fig. 3 / original L321–L334	Improve the scenario figure to a clearer 3×3 matrix; panel (a) is repeated and space could be used better.	Implemented; much improved figure re. speed of understanding. Version 3's corresponding figure presentation and caption at v3 L389–L401 clearly describe the nine scenarios as a matrix, and Table 3 now helps interpret them.

L322 / results around WD attribution	Would coarser wood-density attribution make predictions worse?	Implemented sufficiently. Version 3 states explicitly that replacing measured ρ with a regional mean slightly increased underestimation of AGB at v3 L379–L383, whereas species-specific GWDD values had little effect at v3 L369–L372. The point is now clearly shown in the Results.
Fig. 5 / original L368–L374	There seems to be an artefact/banding in the height data; what were the methods used to measure height in the 4321-tree dataset?	Partly implemented; borderline sufficient. Version 3 adds much more detail on how H was measured in the larger H–D dataset, including the use of a Suunto hypsometer and measurement precision to 0.1 m, at v3 L200–L205. However, the specific explanation for the apparent 'banding' in the plot is not clearly added in the manuscript text, even though it was promised in the response letter. The methods part is fixed; the interpretation of the figure artifact remains only partly addressed.
L388 / discussion of ILUA2 applicability at original L387–L398	Are the ILUA2 calibration forests sufficiently similar in structure/species composition to apply the model here?	Partly implemented. Version 3 now explains that ILUA2 was built from 1319 harvested trees across Acacia, mopane and miombo, with the majority coming from dry and wet miombo, at v3 L249–L258. That helps, and the Discussion also notes forest-type/species effects at v3 L437–L445. Still, the manuscript does not state as directly as the response letter promised that the training data represented the same species/forest types plus other woodland types. Good improvement, but not fully explicit.
Discussion order (original Sections 4.1–4.3)	Would it make sense to discuss WD and H effects before Chave vs ILUA2?	No change, and acceptable. The authors explicitly said they preferred not to restructure the discussion, and version 3 retains the same order: AGB model choice first, then height, then wood density (v3 Sections 4.1–4.3). Since they did not promise a change, this should not count as a missed implementation.

Overall assessment

- Implemented sufficiently / well: most reviewer points, including the non-circular H–D clarification, quantified carbon stocks, bias-correction explanation, simplified RMSE/PE wording, the bar-chart clarification, and the 3×3 scenario table.
- Partly implemented / still somewhat weak: explicit explanation of the apparent height-data banding, the full justification of ILUA2 comparability .
- Not implemented: the wording change from "population" to "data set mean".