Review of: "Modelling and Mapping of Aboveground Carbon of Oluwa Forest Reserve Using LandSat 8 TM and Forest Inventory Data"

Sandeep Pulla[1]

1 National Centre for Biological Sciences

Potential competing interests: No potential competing interests to declare.

The authors estimated aboveground carbon (AGC) stocks in Oluwa Forest Reserve in Nigeria by (1) estimating AGC allometrically in twenty 30 m x 30 m sampling plots and (2) using those estimates to train a linear model having remote-sensing-based multi-spectral vegetation indices as predictors. Both 1 and 2 are useful contributions, and I commend the authors for their effort. The methods used in the paper are generally sound. However, there is room for improvement, and I have provided several suggestions below to help improve the manuscript. In addition, I would suggest a round of proofreading to correct numerous minor errors.

Unfortunately, the manuscript doesn't have line numbers, but I hope the context being referred to in the comments below is clear.

Abstract:

- The abstract can be condensed. For example, the sentence 'To estimate its forest ...' can be condensed, as the justification for and use of RS to estimate AGC is not particularly new. Also, the first four sentences have a fair bit of repetition that is not necessary. The sentence 'The chosen model ...' is not necessary since goodness-of-fit measures have been provided, and the reader can decide if that is adequate.
- The number of sampling plots and their sizes should be mentioned.
- At least the spectral indices used in the final model should be mentioned.
- It should be mentioned that the presented AGC estimates represent the mean. Standard deviations should be provided.

Introduction:

- When referring to global tropical forests, it is more appropriate to cite studies with a wider (preferably global) scope, e.g., Pan et al. (2013).
- Para 2: Please explain what you mean by 'direct and indirect approaches'.
- Para 2: 'Allometric equations' are typically part of field-based (non-destructive) methods, so they aren't a separate method.

Study Area:

- I recommend converting all method-related sections into sub-sections under a 'Methods' section.
- A map showing the study area in the regional context is needed.
- I strongly recommend including a figure with photographs of some of the sampling locations so the reader has a better sense of the overall forest type and the variability within it.
- Rainfall and annual temperature: During what time period? Please also mention either the range or standard deviation of annual temperatures.
- Species: The species description should be more detailed. What are the dominants in the canopy, understory, and ground cover? Species names should mention authorship and family.

Ground-based Biomass and Carbon Assessment:

- Instead of listing the equipment, which is not necessary, please describe how this equipment was used, i.e., the inventory protocol.

Data Collection:

- The methods used to measure Db, Dm, and Dt need to be described clearly as they will not be familiar to many readers. Dbh is a standard measure, but it's still worth mentioning at what height it was measured (since it varies from protocol to protocol).

Wood Density:

- A more standard method is to use genus- or family-level rather than forest-level average wood densities (when unavailable at species-level). The reason is that there can be substantial WD variation within a forest (e.g., pioneers vs. late-successional species). On the other hand, wood density is phylogenetically conserved, i.e., closely-related species tend to have similar WDs (see, for example, Chave et al., 2006). I recommend that the authors consider this alternative (perhaps the results from both approaches could be presented and contrasted).

Volume Estimation:

- Please provide a citation for the Newton method and a justification for its use (e.g., is it the most accurate for tropical rainforest trees in the region?).

Estimation of Carbon:

- 0.5 is almost certainly an overestimate for tropical angiosperms (see Thomas & Martin, 2012).

Spectral Indices Extractions:

- I recommend providing justification for choosing these specific indices (citations to studies that show their superiority compared to other indices, for example).
- What was the spatial resolution of these spectral indices? For what time period were they obtained?
- The spectral index datasets should be properly cited.

Table 1:

- Please check the formulae carefully. For example, the SAVI formula is incorrect.

Regression and Evaluation:

- Please provide a justification for the linearity assumption. Please provide scatterplots of biomass vs. each of the predictors, at least in a supplementary.
- It is well known that evaluating the performance of a model on the data it was fitted to (which is what R2, RMSE, etc. do) can result in overly-optimistic models, i.e., models that perform well on training data but poorly on novel data. Therefore, it is recommended that model performance be evaluated on a subset of the data ('testing data') that was not used to fit the model ('training data'). Cross-validation is a good approach that could be used here (see James et al., 2021; for a detailed treatment, see Yates et al., 2023). I recommend the authors consider such approaches to obtain models that have greater predictive accuracy and are more widely applicable.

Results:

- What species were found in the plots? What were their abundances? What was the AGC in the dominant species? This information is immensely useful and should be provided at least in a supplementary.
- For how many species was wood density available at the species level, and for how many at the forest level? What was the relative aboveground volume proportion for the two categories? This has implications for the accuracy of the AGB estimates. For transparency, I would recommend providing species-level WD estimates in a supplementary.
- The plot numbers mean little to the reader. Could the authors speculate on the causes of AGC variation between the plots with highest and lowest AGC? E.g., proximity to disturbance, topography, soil, species composition?

Table 2:

- The last row should also report standard deviations (this is also a general observation -- measures of uncertainty should always accompany averages).

Modelling Using Geographic Information System and Remote Sensing:

- The second sentence appears to contradict the first. I assume the first one refers to Pearson's correlations between AGC and each spectral index and the second one refers to multiple regression? If so, the sentences should be reworded to clarify this.
- 0.5 appears to be an arbitrary threshold. Please provide a citation if it is not.
- What motivated the log transformation? Was it based on exploratory data analysis or model diagnostics? For example, a log transformation is often used to linearize a non-linear relationship or if the residuals have a skewed distribution or are heteroskedastic. Or was it based on other studies? Please clarify.
- For transparency, the full list of models that were tried out, and not just the final models (Table 3), should be listed. Please also see the comment above regarding cross-validation -- such safeguards are recommended so the final

model is not overly influenced by the peculiarities of the specific random sample and generalizes well to locations not included in the model training dataset.

Spatial Distribution of Carbon with Spectral Indices Model:

- To help clarify why the mean AGC values reported from other studies differ from the value estimated in the present study, I suggest mentioning the methods used and biophysical characteristics of the forests in each of those studies.

Conclusion and Recommendation:

- For the reasons mentioned above, I would suggest utilizing a method such as cross-validation to select a model that can be adopted in forests other than the one studied here (see also Wenger & Olden, 2012).

REFERENCES

Chave J, Muller-Landau HC, Baker TR, Easdale TA, Steege H, Webb CO (2006) Regional and phylogenetic variation of wood density across 2456 neotropical tree species. Ecological Applications 16:2356–2367

James G, Witten D, Hastie T, Tibshirani R (2021) An introduction to statistical learning: with applications in R. Springer US, New York, NY

Pan Y, Birdsey RA, Phillips OL, Jackson RB (2013) The structure, distribution, and biomass of the world's forests. Annual Review of Ecology, Evolution, and Systematics 44:593–622. https://doi.org/10.1146/annurev-ecolsys-110512-135914

Thomas SC, Martin AR (2012) Carbon content of tree tissues: a synthesis. Forests 3:332–352. https://doi.org/10.3390/f3020332

Wenger, S.J. and Olden, J.D. (2012), Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution, 3: 260-267. https://doi.org/10.1111/j.2041-210X.2011.00170.x