Referee #1 - N. Picard

General comments

This study develops allometric models for tree height and crown area that integrate species traits as co-variates to account for species differences in allometric scaling. Questions (ii) and (iii) addressed by this study (P2L32-34) are not new and have been addressed by other studies before. Related text in the manuscript could be shortened (e.g. P9L10-19). Question (i) is quite new, but a recently published paper by Loubota Panzou et al. (2018) also addressed it (with a slightly different approach, though). There are differences between the results reported by Loubota Panzou et al. (2018) and those shown here. For instance, Loubota Panzou et al. (2018) found that large-statured canopy species tended to be light-demanding and small-statured understory species tended to be shade-tolerant, whereas the current study found light-demanding species to have smaller stature than shade-tolerant species (Fig. 2d). Rüger et al. (2012) also reported positive (but weak) correlation between the light response of species and their maximum height. Hence, the relationship between light requirements and maximum height may be more complex than that described at P9L24-27. It has often been reported that there are many small tree species on BCI, while large lightdemanding species are common in central Africa. Could it be a confounding factor?

> AR: We appreciate the detailed comments by the reviewer. We agree that the work by Loubota et al (2018) is relevant for the current manuscript; we did not previously cite it because it had not yet been published at the time we submitted our manuscript. We now reference Loubota et al. (2018) in multiple places. Both the approach and the results differ, as noted by the reviewer; the differences are interesting and may reflect differences in life history strategies in different forests. We agree that the relationship between light requirements and maximum height may be more complex than described in our original discussion, both in our site and in other sites, and we modified our discussion to better address these complexities. We see potential differences among regions in life history strategy distributions not as a confounding factor, but as an interesting topic for future research, although clearly beyond the current contribution or Loubota et al. 2018.
>
> Regarding the novelty of this work, we have revised our text to more clearly emphasize the novel aspects of this study, which are to provide a sound methodological approach that allowed us to simultaneously examine species variation in tree allometry in relation to functional traits and compare alternative scaling models. We agree that other studies have addressed all three of our questions to some degree, and that some of our analyses confirm previous findings rather than presenting novel conclusions. Nonetheless, these questions continue to be of interest, as is clear from other recent publications on these topics, and from the comments of the other reviewers. We have revised the text to feature relevant prior work earlier in the introduction (the previous version treated such work mainly in the discussion.) We have also shortened some discussion text where the topics are well-treated in prior work.

The implication of species differences in height allometry for forest biomass estimation is not addressed by the study. Thus the statements at P11L6 and P12L28-31 is not fully supported by the current study. Rather than comparing the AGB estimates using different height models (power model vs. saturating model), it would be more interesting to compare the gain (both in terms of accuracy and precision of the estimate of forest biomass) of including species traits into the height model. The comparison should be done at plot level and not at tree level because practical needs are for estimates of forest biomass stocks. At plot level, random tree-level errors level off, so that reducing the residual error of the allometric model at the cost of greater uncertainty on the estimates of the parameters of the model is not necessarily the best option. Uncertainty on the species traits (in particular when it comes to forest inventory data where little information is available for most species) should also be considered when assessing the relevance of integrating species traits into allometric models.

> AR: Our manuscript did in fact address the implications of interspecific variation in height allometry for forest biomass estimates, and it reveals differences in AGB estimates for the BCI 50 ha plot based on height equations fitted for an average species with respect to those incorporating species differences in allometric parameters (Table 3, text section 3.3). We first compared individual tree AGB estimates based on actual height measurements with those obtained using allometric equations, in order to justify the use of the AGB estimates obtained using heights from the gMM model as a point of reference for the plot-level analyses (Figure 4). We then estimate plot-level biomass under alternative height models, with and without interspecific variation, and with the power or gMM models (Table 3).
>
> The two statements highlighted by the reviewer were part of the discussion and referred to potential avenues for future research rather than specific results presented in the manuscript. We adjusted our wording to avoid any confusion. The analyses suggested by the reviewer are interesting but beyond the scope of the current contribution; at the end, this manuscript does not include any actual measurement of biomass and our allometric models did not target specifically the prediction of forest biomass. The assessment of the advantages of trait based models and the impact of uncertainty on trait measurements deserve further attention and dedicated studies on their own.
>
> CHANGES: As detailed below, and in response to other reviewer comments, we edited the text extensively to include uncertainty measures for AGB estimates (90% credible intervals in Table 3), included figures in the text illustrating heterogeneity in species allometries, and provided further details about the comparison of AGB. As further detailed below, we now explicitly state in section 2.4.1 that our approach ignored measurement errors in trunk diameters and in species traits. Finally, we also included a new table in the supplementary material providing parameter estimates for all the models fitted; previously we included only parameter estimates for the best models (Table 2, S1).

The systematic bias of AGB prediction using the power model for tree height (Figures 4 and 5) is a bit surprising. Yellow dots in Figure 5 show the ratio AGBpow/AGBobs in log axes, where AGBpow is the biomass estimated using the power model for tree height. In other words, Fig. 5 shows log(AGBpow) − log(AGBobs). Because log(AGB) = log(0.0559) + log() + 2log(D) + log(H), all terms except the one depending on height cancel out, so that Fig. 5 is actually showing log(Hpow) − log(Hobs), where Hpow is the height predicted by the power model. In other words, yellow dots in Fig. 5 are showing the residuals of the fitted power model for log-transformed height. If the power model for tree height had been fitted by linear regression on log-transformed data, then, by definition, these residuals would have a zero sum, which does not seem to be the case in Fig. 5 (but the x-axis is truncated to dbh 30 cm, so maybe we do not have the right picture). The fact that the residuals of the power model for height (Eq. 1) are not centered on zero is a bit concerning and seems to contradict what is written at P6L16. Is it a consequence of the hierarchical approach where species-level models are averaged at community-level? Could you clarify how the community-level model is obtained from the species-level models? (Do all species have the same weight, or do all trees have the same weight? Etc.)

AR: Figures 4 and 5 were included to highlight the better performance of the gMM model for large trees with respect to the biases found for the power function model. The reviewer is correct that the residuals are the same as those for tree height and that residuals in log(AGB) sum to zero across the fitted data when the full range of DBHs and random effects are considered. However, plot-level estimates of AGB are concerned with the sum of residuals in AGB (not log(AGB)) over trees with the observed plot-level DBH distribution (not the DBH distribution of the trees included in the allometric model fitting), and these do not sum to zero since the contribution of large trees outweigh the underestimates in many small trees. Our aim with Figure 5 was to illustrate this bias in large trees, so we plotted a small subset of DBH ranges to avoid the clutter at very small sizes. We have now included another version of this figure featuring the full range of DBH in the supplementary materials (Fig S3).

It is important to note that the AGB predictions in figures 4 and 5 do not include species-specific information; they are based on the community level model and use an average wood density. The community level height model features expected parameter values for an average species. The hierarchical structure accounts for potential differences in sampled DBH ranges and in sample size across species. The bias in the prediction of the power function reflects the inability of this functional form to accommodate the saturation and convergence in tree height at large diameter sizes; indeed, the misfit at larger sizes is compensated by deviations at small tree sizes. The statement in P6L16 about the unbiased community-level models refer in the case of tree height to the gMM model, which provides a better fit and much better residual patterns than the power function (at the end, we rejected power function models for tree height).

CHANGES: As detailed below, we revised the presentation of the AGB comparisons in the methods and in the results. We also revised the captions of figures 4 and 5 and combined both plots in a single figure. We included a new figure in the supplementary material featuring the full range of DBHs for the comparison of AGB.

Specific comments

P2L19-20 and P10L28-29: these sentences are a bit misleading. The use of wood density as a way to account for specifies differences in multispecies biomass equation is an old idea (e.g. Brown et al. 1989) that is now commonplace. It is accordingly less common for height- and crown-diameter allometries, but see Loubota Panzou et al. (2018).

AR: Good point; we edited the text to account for the common usage of wood density. In the first case, we were referring to interspecific differences in allometric scaling for different tree dimensions, so we edited the text to make clear the distinction.

CHANGE: [P2L19] *Approaches pooling data across species inherently fail to recognize species heterogeneity <u>in allometric scaling</u> and limit the potential to identify and define plant functional groups.*

[P10L28] *Whereas many models incorporate only <u>individual</u> trunk diameter (Brown, 1997) <u>and species wood density (Brown et al. 1989)</u>, current state-of-the-art models typically include estimates of ~~wood density and~~ tree height as well (e.g., Chave et al., 2014), and crown dimensions have also been incorporated in some models (Goodman et al., 2014; Ploton et al., 2016; <u>although Fayolle et al. 2018 found a minor role of either crown or height dimensions on biomass estimates</u>).*

P5 Eq.(2)-(4): these expressions do not correspond to the f function in Eq.(1) but rather to its exponential transform, right? In fact, the confusion comes from P4L29: Should not it be "f predicts expected log tree height or crown area" rather than "f predicts expected tree height or crown area"? Please clarify.

AR: Good point; we corrected the text as suggested.

CHANGE: [eq (1)] *where the process model, f, predicts expected <u>natural log</u> tree height or crown area* [...]

P5L18: why a univariate linear function? Why not combining several traits?

P5L29: why using the same trait for all models? Why not using different traits for the different parameters of the allometric equations?

AR: We agree that it would in principle be interesting to evaluate additional different, more complex models, incorporating multiple traits, more alternative traits, more complex relationships with traits, etc. However, even our large dataset has power limitations. Fitting more parameter-rich models invariably results in greater uncertainty in individual parameter estimates, and a data mining approach of fitting large numbers of alternative models has known disadvantages with respect to interpretability and reproducibility. Our work fits more complex models than most previous studies, compares a limited number of models involving traits chosen based on previous work and first principles, and incorporates these traits in the simplest possible framework (univariate, linear, one trait at a time). Future research based on larger datasets could usefully evaluate more different and more complex models.

P5L29-30: the sentence is confusing. I guess "trait model" is referring to Eq.(5), but why call it a trait model? A model is usually called after its response variable, not after its predictors (e.g. AGB = aDb is called a biomass model, not a diameter model). Moreover, the fact that the "trait" models have twice the number of community-level parameters is not due to the fact that all models have the same predictor. It is due to the fact that each model has a single predictor.

AR: We are comparing multiple models for, e.g., height, and we need to distinguish them, so it would be confusing to refer to them all as “height models”. We could refer to them as “model 1” and “model 2”, etc., but preferred instead an informative name that provides information on what distinguishes the models, which in our case is in part whether they are based on traits or not. We agree that any model in which each parameter had a single predictor would have the same number of parameters, and we did not mean our wording to suggest otherwise.

CHANGE: We revised the text to avoid any confusion regarding the use of the “trait model” naming convention. Where necessary, we detailed whether we were referring to height or crown area models featuring the effect of a trait. With respect to the trait models having twice the number of parameters, we clarify the text as follows;

We refer to models incorporating relationships between allometric parameters and species traits as ‘trait models’. Because each allometric parameter was a linear function of a trait, ~~they~~ trait models had twice ~~the number of~~ as many community-level parameters as corresponding models lacking covariates. Our trait models each featured a single trait (all parameters in a trait model depended on the same trait). ~~Each trait model modeled all parameters as functions of the same focal trait; thus, the trait models had twice the number of community-level parameters as corresponding models lacking covariates.~~

P8L19 sqq., “ (...) based on the power model”: which power model? No power model to predict tree height has been presented so far (P8L2 only mentioned that power models did much worse than the other models). The power model for height is actually later given (P8L24) but without specifying on which species trait it is based (presumably the growth trait). Please clarify and present the power model before presenting the consequences for AGB estimates.

AR: We thank the reviewer for drawing our attention to this oversight. We now introduce the equation for the tree height, community level power model used in Figures 4 and 5 earlier in the text. We also edited the text to detail which type of model was used for each comparison. Figures 4 and 5 are based on community level, height-diameter models with traits, although the inclusion of traits does not affect the divergence in AGB predictions. Table 3 details the impact of taking into account species identity using the 50 ha BCI plot.

CHANGE: [First sentence Section 3.3] *Individual tree aboveground biomass (AGB) estimates based on the community-average power model ($H = 3.02D^{0.56}$, Table S2) were strongly upwardly biased for large trees relative to estimates based on measured heights, whereas AGB estimates based on the gMM height model were unbiased (Fig. 4).*

P8L24-27 and Table 3: comparing AGB estimates without specifying the estimation uncertainties does not make much sense. An AGB difference of 283 vs. 252 Mg ha−1 does not have the same meaning if its estimation error is 10 or 1 000 Mg ha−1. Therefore, please complement Table 3 with the estimation uncertainties. Referring to P6L14-15, please also clarifies whether you are considering only the uncertainty on the parameter estimates (confidence interval) or also the residual error (prediction interval), and possibly if you are also propagating other sources of errors (e.g. measurement error, or the error in the estimation of the species traits).

AR: We added uncertainty estimates for plot-level AGB using 90% credible intervals based on 5000 samples from the posterior distributions of all the parameters of our allometric models (paralleling those given for other parameters and derived quantities features in the manuscript. This addition does not alter the interpretation or conclusions of the analysis of differences in AGB estimates across models. We ignored measurement errors in trunk diameters and in species traits, as we now explicitly state in section 2.4.1. We considered these sources of error to be minor compared to measurement errors in height and crown area, although we agree their full consideration would be a useful addition to future work. We also noted to readers that we ignored uncertainty in the Chave et al (2014) AGB model.

CHANGE: [End of Section 2.4.1] *Finally, the model ignored measurement errors in trunk diameter and in ancillary trait data.*

[Section 2.4.3]: *We computed 90% credible intervals for each AGB estimate based on 5000 samples from the posterior distributions of all parameters of the corresponding allometric models.*

[Table 3]: *Posterior mean estimates (with their 90% credible intervals) of total aboveground biomass*

P10L7-8: unclear. What is emerging? Model fitting shows that differences among species are not strong enough for the model with a species effect to be better than the model without a species effect. It does not show the emergence of a pattern at community level from species-level patterns. Moreover the power model for crown allometry in Farrior et al. (2016) is an assumption of the model, not an emerging property.

P10L12, P11L6 and P12L26-27: interspecific variation in crown allometry is not that "high"/"considerable"/"extensive" since the gain in predictive accuracy brought by the species trait does not even compensate for the increase in the effective number of parameters of the model.

AR: We agree that the wording regarding "emergence" was unclear, and have revised it. The models with no traits also feature variation in allometric parameters among species, it is just that this variation is not linked to the traits included in the model. Individual effects differed markedly among species both for tree height and crown area (see Figs 2 and 3). The estimated crown area exponent b ranges from 1.09 to 1.77 across species, which is accurately described as "high", "considerable" or "extensive" variation. Following a comment by reviewer R#2, we provide a summary of the information available in Table S2 in section 4.2 to support the statements about the variability in crown area scaling across species.

CHANGE: *This large-scale consistency in community-level relationships emerges despite local variation among species (e.g. the exponent b ranged between 1.09 and 1.77 across species, Table S2)~~, and suggests the operation of a general mechanism in the emergence of community-level allometric scaling in crown geometry (Farrior et al., 2016)~~.*

P12L18-26: repetition of previous text.

AR: We tried to summarize the main points of the manuscript in the last, concluding paragraph. We believe that a certain degree of repetition is allowed (and recommended) in this context.

Figure 4: what are the lines representing? I understand that you are using for height the community-level across-species relationship (Eq. 7), so that height is a function of diameter only. However the biomass equation (Eq. 6) still depends on wood density that varies across species. Therefore, one would expect to have different lines for the different species rather than a single line for all species.

AR: The lines in the figure portray predictions based on community level allometric equations and community average wood density, for comparison with predictions using individual height and community average wood density (points). We revised the relevant text in the methods and the figure caption to clarify these details. We use the same wood density for all species to highlight the effects of the height allometry; as pointed by the reviewer, we note that the proportional error for each tree would be the same if we included species-specific wood densities in both cases instead. AGB predictions based on models taking into account species differences in wood density and allometric scaling are compared in Table 3. As noted in the text, taking into account species identity can improve the quality of AGB estimates based on a power function model for tree height, although they do not correct for biases at large DBHs.

CHANGE: [Section 2.4.3] *For those species for which species-specific wood densities were not available, we used the average over species for which values were available (ρ = 0.5304 g cm^{-3}, Wright et al., 2010).*

[Figure 4 caption] *Comparison of estimates of aboveground biomass (AGB, Kg dry matter) as a function of DBH based on observed tree heights (grey points) with those based on height predicted from community level power function (orange lines) or generalized Michaelis-Menten (blue lines) models. ~~for individual trees based either on measured heights (grey points) or on heights predicted from a power function fit (orange) or a generalized Michaelis-Menten fit (blue).~~ All AGB estimates were based on the biomass allometry of Chave et al. (2014) and used the average value of wood density across species (ρ = 0.5304 g cm^{-3}; data from Wright et al., 2010) to highlight variation related to the height allometry. Predictions from the allometric models are based on simulations of the posterior distribution (lines correspond to the median and 90% posterior central interval) of the community-level, across-species relationships.*

Figure 5: what are the lines representing? Smoothing functions? Because Hobs is an individual tree-level data and not a one-to-one function of diameter, neither is the ratio Hmod/Hobs.

Figure 5: why are dbh starting from 30 cm instead of 1 cm?

AR: Good point; we edited the figure caption to detail both aspects. The lines are LOESS smoothers and they were added to show the overall departure of each model from observations for different diameter ranges. We restricted the range to focus on the large tree sizes where the predictions of the power and gMM model differed systematically, and which drive differences in plot-level estimates. We have now added a figure showing the full range of diameters to the supplementary materials.

CHANGE: [Figure 5 caption] *Relative error for estimates of individual tree dry aboveground biomass (AGB, Kg dry matter) based on model predictions of tree height (AGBHmod) compared with estimates derived from height observations (AGBHobs), for trees with DBH > 30 cm (the full range is shown in figure S3). Modeled tree heights were from community-level models fitted with either the power function (orange dots) or generalized Michaelis-Menten function (blue dots). All AGB estimates were based on the biomass allometry equation 6 (from Chave et al. 2014) and used the average value of wood density across species, to highlight variation related to the height allometry. The lines are LOESS smoothers that illustrate the overall departures of each model from perfect prediction (i.e. AGBHmod/AGBHobs ratio equal to unity) as a function of DBH.*

We added a new figure S3 in the supplemental material featuring deviations over the full range of DBHs.