Research Article
doi: 10.1590/01047760202330013381
vol(30), 2024

CERNE

Are tropical forests an *extremistan* environment?

Hassan C. David[1]✤

[1]Federal University of Parana, Agrarian Sciences Sector, Department of Forest Science, Curitiba, PR, Brazil.

FOREST MANAGEMENT

ABSTRACT

Background: This study reveals the surprising impact of large trees on biomass modeling and estimation in tropical forests. Findings emerged from viewing tropical forests as an *Extremistan* environment—a domain where a small number of extreme events disproportionately impact overall outcomes. The aims were to: (i) determine whether humid tropical forests can be characterized as an *Extremistan* environment, (ii) quantify the impact of large trees on the biomass quantification, and (iii) recommend better practices to mitigate the impact of large trees. The methods included forest simulation, biomass model calibrated with multi datasets and extensive examination of the impact of large trees on model performance and mean biomass estimation.

Results: The select group of the 1% heaviest trees account for 25–35% of the total biomass, a concentration analogous to the wealth concentration in developed countries. Additionally, a "tyranny" of the 5% heaviest trees (diameter >18–31 cm) was observed, in which 50-75% of the total biomass is retained, significantly affecting biomass modeling and mean biomass estimation regardless of the model used.

Conclusions: This study confirms that humid tropical forests behave as an *Extremistan* environment. For biomass and carbon inventories, installing 10,000-m^2 sample units is recommended to mitigate the "tyranny" effect of the 5% heaviest trees, with a minimum size threshold of 4000 m^2.

Keywords: Biomass and carbon inventory; pantropical biomass database; biomass retention rate; tree-level modeling.

HIGHLIGHTS

The *Extremistan* environment existent in humid tropical forests is demonstrated.
The 1% heaviest trees retain 25–35% of the total biomass in tropical forests.
A "tyranny" of the 5% heaviest trees massively affects the biomass estimation.
Sample units of 10,000 m^2 are recommended to reduce the "tyranny" effects.

DAVID, H. C. Are tropical forests an *extremistan* environment? CERNE, v.30, 2024, e-103381, doi: 10.1590/01047760202430013381.

✤Corresponding author: hassancamil@ufpr.br Received: January 9/2024 Accepted: October 31/2024

INTRODUCTION

An *Extremistan* environment refers to a domain where extreme events have a disproportionately significant impact on overall outcomes. The concept of *Extremistan* was introduced by Taleb (2007) and has influenced many authors across various fields of research, including Earth sciences (Rising et al., 2022; Fischer et al., 2021; Merz et al., 2021), economics and finance (Chowdhury et al., 2022; Yarovaya et al., 2022), risk management (Paté-Cornell, 2012), safety (Paltrinieri et al., 2019), and social sciences (Hofman et al., 2017). In an *Extremistan* environment, the distribution of certain phenomena is characterized by a high degree of inequality, where a small number of observations can drastically influence overall statistical properties.

One criticism from Taleb (2007) concerns the improper use of the Gaussian distribution in *Extremistan* environments, as this distribution fails to predict rare extreme events that exceed two or three standard deviations. These rare yet highly impactful events are defined as Black Swans. The Gaussian distribution is appropriate for what Taleb (2007) defines as Mediocristan; an environment free of Black Swans and inhabited by White Swans (regular events). Taleb (2007) classifies various human-induced phenomena in the real world, such as the frequency of words used daily, the number of book sales, and stock market fluctuations, as belonging to the *Extremistan* environment, where Black Swans are somewhat concealed. This paper demonstrates that the *Extremistan* environment extends beyond human-induced domains, reaching natural phenomena such as tree communities in humid tropical forests due to the presence of large trees.

Large trees differ from Black Swans for two main reasons. First, although rare, large trees are visible in the forest; second, the size of a tree is inherently limited by nature, as it cannot grow indefinitely due to biological constraints and external factors such as the environment. This implies that the impact of a large tree is restricted to its own size. In contrast, Black Swans are "invisible," rare, and possess scalable impacts because they can be as impactful as possible. Therefore, it is reasonable to associate large trees in an *Extremistan* forest with neither Black Swans nor White Swans (regular events). Fortunately, Taleb (2007) recognized that *Extremistan* environments might be inhabited by Gray Swans, which fall between the categories of White and Black Swans. As this study presents a novel way of visualizing tropical forests, two important definitions arise: *Extremistan* forest—a humid tropical forest where Gray Swans reside visibly (not "hidden," like Black Swans)—and tree-Gray Swan—a rare yet probable large tree that has a significant impact on a sample. The remaining issue is determining the size of a tree-Gray Swan.

To understand the narrative of this paper, consider the following process for quantifying forest carbon stock: (i) felling trees to directly obtain their weights (i.e., biomass), (ii) applying biomass models, (iii) using a factor to convert biomass into carbon, (iv) predicting the carbon (~50% of the dry biomass) stored in each sample unit (SU), and finally (v) estimating the mean and total forest carbon stock. Two remarks related to (ii) and (iv) are highlighted.

First, it is well known that individual-tree biomass models typically yield model residuals greater than ±100%, and sometimes ±500%, as shown in Trautenmüller et al. (2023) and Romero et al. (2022) (approximate percentages calculated in this study). Frequently, biomass models are calibrated with datasets containing numerous small trees and only a few large ones (when available), often with a sample size of fewer than 100 trees (Nogueira et al., 2008; Chambers et al., 2001; Brown, 1997; Overman et al., 1994; Brown and Iverson, 1992; Brown et al., 1989). However, the significant issue is not only the calibration dataset itself. Another core matter lies in the nature of the variable tree biomass being modeled at the individual tree level. As tree biomass becomes more variable with increasing diameter, biomass data generally exhibit heteroscedasticity (Parresol, 1999), and biomass models cannot be as accurate and precise for large trees as they can be for smaller trees. This means that most tree-level models carry a high level of uncertainty when predicting the biomass of large trees (i.e., those that hold most of the forest biomass).

Secondly, the variance of the mean biomass per unit area can be significantly inflated by the improper practice of installing sample units (SUs) smaller than necessary. For instance, consider a SU of 10,000 m^2 containing trees that collectively weigh 280 Mg, and another SU of 100 m^2 where the weight is 2.8 Mg. If — and this scenario is highly likely in any humid tropical forest — a large tree of 20 Mg is found in these SUs, its representativeness would be approximately 7% (=20/300) for the larger SU and 88% (=20/22.8) for the smaller one. That is, the same large tree might or might not pose an issue when calculating biomass variance among and within the SUs. The relationship between tree size and SU size drives the scalability of the error in carbon quantification. The larger the tree size in a SU, the greater the scalability of the error. Likewise, the smaller the SU size with a large tree, the larger the scalability of the error.

This paper explores the subject of biomass quantification under the hypothesis that large trees are responsible for transforming humid tropical forests into *Extremistan* forests, thereby impacting biomass quantification. The aim was threefold: (i) to verify whether tropical forests portray an *Extremistan* environment, (ii) to demonstrate the impact of large trees on tree-level models and mean biomass estimation, and (iii) to recommend better practices aimed at mitigating the impact of large trees.

MATERIAL AND METHODS

Data and study area

This study utilizes two data sources. The first is the pantropical biomass database of Chave et al. (2015) (hereafter referred to as the "pantropical database"), which is applied in the biomass modeling process described in Section 2.1.1. These authors compiled data from 4,004 trees collected across 26 tropical countries in Africa, Asia, Australia, and Central and South America, available at http://chave.ups-tlse.fr/pantropical_allometry.htm. The database comprises variables such as dry aboveground biomass

(AGB), diameter at breast height (Dbh), total height (H), and wood specific gravity (ρ) of the trees. Additional information regarding the database can be found in Chave et al. (2015). Figure 1 illustrates the relationship between tree AGB and Dbh, as well as the histogram of Dbh for the trees.

The pantropical database was selected primarily for three reasons. First, it is the world's largest tree biomass database compiled to date. Second, it has been extensively utilized for biomass quantification purposes in forests across tropical regions (the study has been cited more than 2000 times as of 2022). Third, and perhaps most significantly, the database encompasses large trees.

The second data source utilized in this study pertains to stand-level information from forest sites located in the Brazilian Amazon (Higuchi et al., 1994) and the Atlantic Forest (Souza et al., 2011), which are necessary for the forest simulation described in Section 2.2.2. As the forest quantification and statistical analysis stem from these simulated forests, this study shares the same study area as Higuchi et al. (1994) and Souza et al. (2011).

Analytical procedure

This study focuses on the hypothesis that large trees cause humid tropical forests to behave as an *Extremistan* environment, thereby impacting biomass quantification. The analytical procedure delineated to verify that hypothesis and quantify the impact of large trees was divided into four sections: *forest simulation, biomass modeling, impact of large trees on the biomass estimation and impact of large trees on sample units*.

Forest simulation

Five 1-ha forests were simulated following a three-phase process: (1) characterization of tropical forests based on surveys, (2) utilization of the Tropical Biomass & Carbon application (TB&C App), and (3) selection of trees from the pantropical database. The phases are explained as follows.

Phase 1 – Characterization of tropical forests based on surveys. Surveys conducted on Brazilian tropical forests were consulted and selected when criteria were met, being availability of forest attributes such as basal area (BA), density of trees (N), minimum and maximum Dbhs, besides having satisfactory sample size and existence of large trees in the *su's*. Two surveys were selected. The first is Higuchi et al. (1994), where *n*=114 *su's* were installed in two remnants of Ombrophilous Forest (OF) in the biome Amazon (OF is also known as Rainforest). The second survey is Souza et al. (2011), where *n*=22 *su's* were installed in two remnants of Seasonal Forest (SF) in the biome Atlantic Forest. In addition, the Brazilian National Forest Inventory (NFI) database was consulted to obtain forest attributes in another remnant of OF, also located in the Amazon. This consultation comprised a set of *n*=212 *su's* (each *su* is a 0.8-ha cluster) installed in the southeast region of the State of Amazonas. These NFI data have not yet been published; however, they can be made available to the public upon request (data identification: AM-08). Details about the Brazilian NFI methodology can be seen in David et al. (2019b). The three remnants of OF are subsequently referred to as sites OF1, OF2 and OF3, while the two remnants of SF are designated as sites SF1 and SF2. Table 1 provides the values for the forest attributes BA, N, minimum and maximum Dbhs and sample size for each site. Further information about these sites (except OF3) can be accessed in the original articles (Higuchi et al., 1994; Souza et al., 2011).

Phase 2 – Utilization of the Tropical Biomass & Carbon application (TB&C App). The forest attributes BA, N, minimum and maximum Dbhs presented in Table 1 were utilized to simulate tree Dbhs in the TB&C App (David et al., 2022). To simulate tree Dbhs in this application, users need to follow these steps (i) selection of the vegetation class (in this study, OF and SF); (ii) input of N, minimum and maximum Dbhs as presented in Table 1; and (iii) input of parameters (α and β) of the beta distribution to ensure that the Dbh distribution of the forest sites reproduces a reverse J-curve naturally observed in tropical forests (David et al., 2019a). In the TB&C App, the reverse J-curve can be reproduced by setting $\alpha \cong 0.5$ and $\beta \cong 2.0$. However, to achieve the exact BA observed in each forest site of Table 1, α and β were slightly altered until reaching the exact BA. After executing the three-step procedure described above, the TB&C App generates a list of trees with their Dbh. The procedure was repeated for the five forest sites in Table 1, resulting in five sets of tree Dbhs.

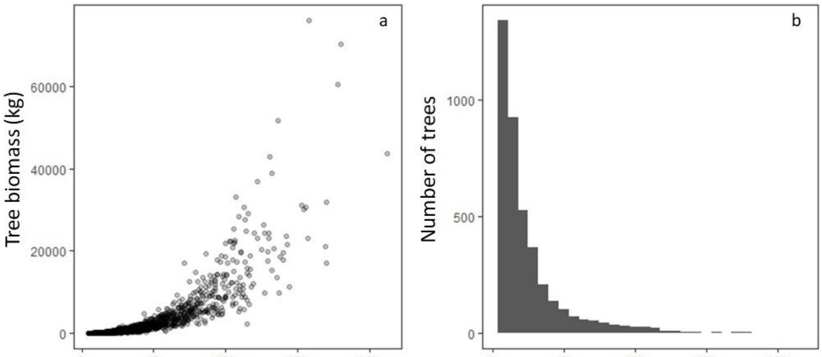

Figure 1: Allometric relationship (a) and histogram of tree Dbhs (b) of the pantropical database. *n* = 4004 trees.

Table 1: Characterization of inventoried forests of studies from literature.

Literature consulted	Forest Site	BA (m^2 ha^{-1})	N (trees ha^{-1})	Sample size$^{(2)}$	Tree Dbh (cm)	
					Minimum	Maximum
Higuchi et al. (1994)	OF1	23.92	1,158	57	5.0	117.2$^{(3)}$
	OF2	32.02	1,264	57	5.0	113.5$^{(3)}$
Unpublished data$^{(1)}$	OF3	32.93	1,135	212	5.0	212.0
Souza et al. (2011)	SF1	16.80	1,459	16	5.0	76.1$^{(3)}$
	SF2	25.70	1,897	6	5.0	76.0$^{(3)}$

OF: Ombrophilous Forest. SF: Seasonal Forest. (1) Data sourced from the Brazilian National Forest Inventory database. (2) Sample size refers to the number of sample units. (3) An approximated value was assumed because authors indicated only midpoint of the diameter class.

Phase 3 – Selection of trees from the pantropical database. The sets of simulated tree Dbhs from the previous phase were used to create the five 1-ha forests evaluated in this study. The simulated forests were composed of trees randomly selected with repetition from the pantropical database, adhering to the number of trees per 10-cm Dbh class. Since the trees were sourced from the pantropical database, the real biomass of the *ith* trees (AGB_i) of the simulated forests is known. The Dbh distribution of the five simulated forests is illustrated in Figure 1S (supplementary file).

Biomass modeling

Six biomass models (Equation 1, 4–8) were applied to predict biomass of each tree belonging to the five simulated forests. These models were applied under four contexts, as follows.

Context 1 – Pantropical biomass model of Chave et al. (2015). The widely recognized pantropical biomass model $\widehat{AGB}_i = e^{\left[0.0063+(\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,976}\right]}$ (Equation 4 from Chave et al., 2015) was applied in its original form. Chave et al. (2015) calibrated their model with a tree database encompassing 25 countries, including Brazil. One consideration is that, in biomass modeling, log-log models are preferred because the logarithmic form solves problems with data heteroscedasticity (Parresol, 1999). In this paper, tree biomass is analyzed in its untransformed original unit (i.e., in kg), and as such, the tree biomass predicted using Equation (4) of Chave et al. (2015) should be added to the correction factor $\sigma^2/2$ (Smith, 1993), resulting in the expression $\widehat{AGB}_i = e^{\left[\sigma^2/2+\hat{\alpha}+\hat{\beta}\cdot In(\rho_i \cdot Dbh_i^2 \cdot H_i)\right]}$. However, the effect of $\sigma^2/2$ was neglected here because the authors reported that it has a minimal impact on the model predictions.

Context 2 – Pantropical biomass model of Chave et al. (2015) calibrated with reduced datasets. In this context, Equation (1) was fitted using seven calibration sub-datasets derived from the pantropical database (n=4004 trees). These sub-datasets were created by successively removing the heaviest trees from the original database. In the first sub-dataset, the 0.1% heaviest trees were excluded from the original database, remaining n_{calib}=4000 trees; in the second sub-dataset, the 0.5% heaviest trees were excluded, remaining with n_{calib}=3984 trees; in the third, the 1% heaviest trees were excluded, and n_{calib}=3964 trees; in the fourth, the 2% heaviest trees were excluded, and n_{calib}=3924 trees; in the fifth, the 5% heaviest trees were excluded, and n_{calib}=3804 trees; in the sixth, the 10% heaviest trees were excluded, and n_{calib}=3602 trees; in the seventh, the 50% heaviest trees were excluded, and n_{calib}=2002 trees; the complete dataset comprising n_{calib}=4004 trees was also added to the graphs for comparison purposes.

$$In(\widehat{AGB}_i) = \hat{\alpha} + \hat{\beta} \cdot In(\rho_i \cdot Dbh_i^2 \cdot H_i) \tag{1}$$

Where, $\widehat{AGB}_i$: predicted dry aboveground biomass of the *ith* tree, in kg; ρ_i: wood specific gravity of the *ith* tree, in g cm^{-3}; D_{bhi}: diameter at breast height of the *ith* tree, in cm; H_i: total height of the *ith* tree, in m; $\hat{\alpha}$ and $\hat{\beta}$: model coefficients to be estimated; and ε_i: model residual.

As this context involves sub-datasets composed of increasingly fewer large trees, it was possible to quantify the impact of large trees on the biomass model. The impact was quantified through the mean error (Equation 2) and the root mean square error (Equation 3) of the model predictions. The first goodness-of-fit statistic measures model accuracy, whereas the second is a measure of model precision.

$$ME = \sum_{i=1}^{n_{calib}} (\widehat{AGB}_i - AGB_i) / n_{calib} \tag{2}$$

$$RMSE = \sqrt{\sum_{i=1}^{n_{calib}} (\widehat{AGB}_i - AGB_i)^2 / n_{calib} - p} \tag{3}$$

Where, AGB_i: observed AGB of the *ith* tree, in kg; $\widehat{AGB}_i$: predicted AGB of the *ith* tree, in kg; n_{calib}: number of trees in the calibration dataset; and p: number of model coefficients.

Context 3 – The 'perfect' models. Equation 1 was refitted using the trees selected to simulate the forests described in phase 3 of section 2.1.1 as the calibration dataset. This context, therefore, involves applying site-specific models, representing a utopian situation of maximum correlation, as the tree datasets used to calibrate the model are the same as those applied in the forest simulation. As a perfect situation, it is reasonable to expect that the models in this context will perform better than any other model.

Context 4 – Biomass models from literature. Five biomass models (Equations 4, 5, 6, 7 and 8) from the literature were tested. Based on the dataset calibration and following the authors' findings, Equations (4 and 5) can be generically applied to both forest types examined in this study; Ombrophilous Forest (OF) and Seasonal Forest (SF). In contrast, Equations (6, 7 and 8) are locally specific to the Amazonian region, where OF is abundant. Although Equations (6, 7 and 8) may not be entirely appropriate for SF, the idea was to determine whether the impact of large trees is substantial enough to negate the differences between forest type-specific and non-specific models. Key information about the authors' model calibration dataset is presented in Table 2, and further details regarding these models can be found in the original articles.

Impact of large trees on biomass estimation

Recall that the tree biomass of every simulated forest is known, allowing for the derivation of population estimates. Thus, the sum of real biomass of the *ith* trees (AGB_i) in a site corresponds to the total biomass, $TAGB_{site} = \sum_{i=1}^{N}(AGB_i)$, where *N* is defined in Table 1. The estimate of the total biomass is accordingly given by $T\widehat{AGB}_{site} = \sum_{i=1}^{N}(\widehat{AGB}_i)$, in which $\widehat{AGB}_i$ is obtained as in contexts 1–4 from the previous section. $T\widehat{AGB}_{site}$ was estimated for the four contexts and forest sites presented in Table 1. Note that, since the forests were simulated to have a 1-ha area size, the total biomass (in Mg) and the mean biomass (in Mg ha^{-1}) are equivalent, i.e., $T\widehat{AGB}_{site} = \widehat{AGB}_{site}$.

The impact of large trees was quantified by assessing the errors in mean biomass estimation resulting from the removal of the top *x%* heaviest trees from the simulated forests. The value of *x%* ranged from 0.1% to 50% of removal. This procedure required reducing the tree density of the simulated forests from *N* (see Table 1) to *N'*, where *N'* is *N* minus the top 0.1% to 50% heaviest trees). The estimated mean error in kg tree^{-1} is expressed in Equation (9) and the estimated total error in Mg ha^{-1} is provided in Equation (10).

$$MET_{site} = \sum_{i=1}^{N'}(T\widehat{AGB}_{site} - TAGB_{site}) / N' \tag{9}$$

$$TE_{site} = \sum_{i=1}^{N'}(T\widehat{AGB}_{site} - TAGB_{site}) \tag{10}$$

Where, $\widehat{AGB}_i$ and AGB_i were defined in Equation (1); *N'*: *N* minus the 0.1–50% heaviest trees.

Impact of large trees on sample units

Finally, the representativeness of the heaviest trees (the top one) belonging to each site was examined as if they were observed in a sampling process with varying *su* sizes. The area sizes of 5000 m^2, 2500 m^2, 1250 m^2, and 125 m^2 were considered, from which the proportionalities *p* are respectively derived as follows: .5 (=5000/10000), .25 (=2500/10000), .125 (=1250/10000), and .0125 (=125/10000). If the same large tree is observed in *su's* of different area sizes, its impact will be greater as the area size of the *su* decreases. In biomass quantification, this impact refers to the representativeness of the heaviest tree (RHT_{site}), which can be calculated for each forest site using Equation (11).

$$RHT_{site} = \frac{AGB^{+}}{AGB^{+} + p \cdot \sum_{i=1}^{n-1}(AGB_i^{-})} 100 \tag{11}$$

Where, AGB^{+}: observed *AGB* of the heaviest tree, in kg; AGB_i^{-}: observed *AGB* of the *ith* non-heaviest tree, in kg; *p* was defined. *n* is the number of trees in the site (see Table 1), subtracted 1 due to the heaviest tree.

Table 2: Models from the literature applied for tree biomass prediction.

Source	Calibration dataset		Model
	Dbh range (sample size)	Data collection site	
Brown (1997)	5–148 cm (170 trees)	Tropical region	$\ln(\widehat{AGB}_i) = -2.134 + 2.530\ln(Dbh_i)$ (4)
Brown and Iverson (1992)	4–112 cm (169 trees)	Tropical region	$\widehat{AGB}_i = -21.297 - 6.953(Dbh_i) + 0.740(Dbh_i^2)$ (5)
Chambers et al. (2001)	5–100 cm (315 trees)	Brazilian Amazon Forest	$\ln(\widehat{AGB}_i) = -0.370 + 0.333\ln(Dbh_i) + 0.933(Dbh_i)^2 - 0.122\ln(Dbh_i)^3$ (6)
Nogueira et al. (2008)	5–124 cm (262 trees)	Brazilian Amazon Forest	$\ln(\widehat{AGB}_i) = -1.716 + 2.413\ln(Dbh_i)$ (7)
Overman et al. (1994)	8–100 cm (54 trees)	Colombian Amazon Forest	$\ln(\widehat{AGB}_i) = -1.966 + 1.242\ln(Dbh_i)$ (8)

Variables were defined in Equation (1).

RESULTS

How much biomass is retained in large trees?

Table 3 presents the biomass retained by the heaviest trees in relation to the total biomass of the forest. More than half of the total biomass is retained in the 5% heaviest trees, which have a Dbh greater than 26 cm in the OF and greater than 16 cm in the SF. Generally, the 1% heaviest trees account for 23–38% of the total biomass, while the 50% heaviest trees retain 87–95% of the total forest biomass. The 0.1% heaviest trees (approximately 1–2 trees ha^{-1}) contribute 3–14% of the total AGB.

As additional information, Table 2S (supplementary file) presents the exponents related to the fractal phenomena of biomass stored in the forests simulated in this study. These values were proposed based on the exponents tabulated in Taleb (2007).

Expected impact of large trees on sample units according to area size

The resulting values of RHT_{site} are illustrated in Figure 2 and presented in Table 1S (supplementary file). In the smallest *su's* of 125 m^2, RHT_{site} exceeded 90% of the total biomass of the *su*, which means a retention over 90% (of the total biomass in the *su*) in a single tree – the heaviest one. That extreme retention occurred at site OF3.

Figure 2 shows that the impact of the top one heaviest trees increases sharply in *su's* < 2500 m^2, and more abruptly in *su's* < 1250 m^2. The heaviest tress at sites OF1, OF2 and OF3 had the largest impact due to their larger size, while sites SF1 and SF2 exhibited less impact. Nonetheless, representativeness in the SF remained notable (>60%) for the 125 m^2 *su*. The maximum representativeness was observed in site OF3, where the heaviest tree (with a Dbh of 212 cm) accounted for 13.9% of the total biomass of the site. Conversely, the smallest representativeness was recorded at site SF2, where the heaviest tree (with a Dbh of 58.9 cm) retained only 4.4% of the total biomass.

Impact of the large trees on model performance

This section focuses on the pantropical biomass model developed by Chave et al. (2015) and does not involve the simulated sites presented in Table 1. The impact of the heaviest trees on the model's performance was assessed by calibrating Equation (1) with progressively reduced datasets. Initially, the model was calibrated using the complete pantropical database, which includes 4004 trees. Subsequently, the 0.1% to 50% heaviest trees were successively removed from the dataset. Table 4 presents the estimated model coefficients and their standard errors for each reduced dataset. Notably, the coefficients and model specifications are reported in their original scale, rather than the log-log scale used in Equation (1) as described by Chave et al. (2015).

Table 3: Representativeness of the heaviest trees across the studied forest sites.

	x% heaviest trees	Number of trees	% of the population biomass	Tree Dbh (cm)	
				Minimum	Maximum
Site: OF1 – Ombrophilous Forest 1					
x =	0.1	1	6.2	79.1	79.1
	0.5	6	21.7	58.9	117.8
	1.0	12	31.3	56.0	117.8
	2.0	23	44.6	44.0	117.8
	5.0	58	66.9	27.6	117.8
	10.0	116	75.6	14.9	117.8
	50.0	579	92.9	7.3	117.8
	100.0	1,158	100.0	5.1	117.8
Site: OF2 – Ombrophilous Forest 2					
x =	0.1	1	4.0	79.1	79.1
	0.5	6	20.3	75.1	116.0
	1.0	13	34.4	74.8	116.0
	2.0	25	50.2	57.3	116.0
	5.0	63	73.7	36.0	116.0
	10.0	126	83.1	15.9	116.0
	50.0	632	95.4	6.9	116.0
	100.0	1,264	100.0	5.1	116.0
Site: OF3 – Ombrophilous Forest 3					
x =	0.1	1	13.9	212.0	212.0
	0.5	6	28.7	74.8	212.0
	1.0	11	38.2	58.9	212.0
	2.0	23	50.8	44.0	212.0
	5.0	57	67.5	26.0	212.0
	10.0	114	76.3	15.9	212.0
	50.0	568	94.2	9.0	212.0
	100.0	1,135	100.0	5.1	212.0
Site: SF1 – Seasonal Forest 1					
x =	0.1	1	3.4	59.7	59.7
	0.5	7	16.2	44.0	76.0
	1.0	15	25.8	36.0	76.0
	2.0	29	37.0	26.0	76.0
	5.0	73	50.9	15.9	76.0
	10.0	146	58.8	13.0	76.0
	50.0	730	87.6	6.9	76.0
	100.0	1,459	100.0	5.1	76.0
Site: SF2 – Seasonal Forest 2					
x =	0.1	2	4.4	58.9	59.7
	0.5	9	14.1	45.3	79.9
	1.0	19	23.1	36.0	79.9
	2.0	38	35.8	26.0	79.9
	5.0	95	52.7	19.7	79.9
	10.0	190	62.0	14.0	79.9
	50.0	948	89.0	6.9	79.9
	100.0	1,897	100.0	5.1	79.9

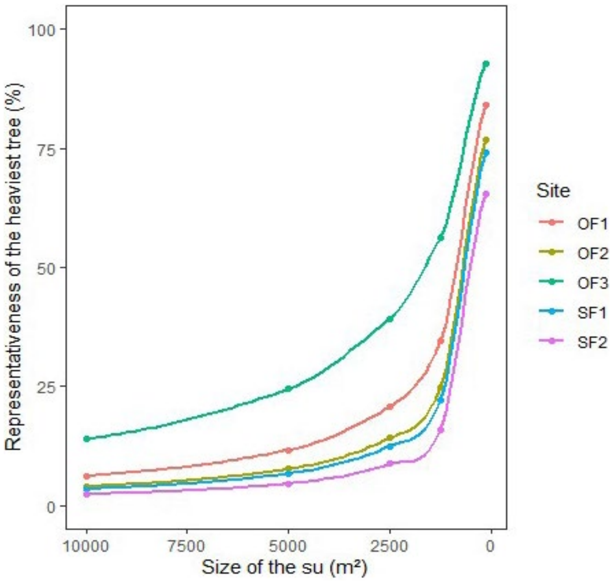

Figure 2: Representativeness in the plot-level biomass of the heaviest tree (RHT_{site}) in sample units (*su*) with different sizes and sites (The Dbh of the heaviest tree is shown in the top line of each site in Table 3).

The greatest variations in the model coefficients and standard errors were observed among the models calibrated when the 50% heaviest trees were excluded. The results suggest that the exclusion of up to the 10% heaviest trees from the dataset exerts little impact on the model coefficients when compared with the complete dataset that includes all 4004 trees. The goodness-of-fit statistics for the models listed in Table 4 are presented in Table 5.

The statistics *ME* (Equation 2) and *RMSE* (Equation 3) in Table 5 were calculated by considering the trees belonging to the reduced datasets, whereas ME_{10} and $RMSE_{10}$ were computed using only the 10 heaviest trees. Both *ME* and *RMSE* successively decreased as much as the heaviest trees were excluded from the database. This might sound contradictory at first sight, having in mind that larger-size datasets should produce better statistics. But this result makes sense if we consider that the heaviest trees are those ones capable to inflate *ME* and *RMSE*, since they retain the most biomass and variability, as Table 3 shows. An expected result in Table 5 is that ME_{10} and $RMSE_{10}$ (not *ME* and *RMSE*) increased as the heaviest trees were excluded from the calibration dataset. This is an intuitive result because, if heavier trees are excluded from the calibration dataset, the biomass model would lose accuracy in the biomass prediction of the heaviest trees. For the 10 heaviest trees – which is 0.25% (=10/4004×100) of the trees in the pantropical database –, the largest differences in accuracy (ME_{10}) and precision ($RMSE_{10}$) were noted when the 50% heaviest trees were excluded. The ME_{10} and $RMSE_{10}$ estimated when the 50% heaviest trees were excluded were about twice worse than the other ones. For all % of removal, *ME* and *RMSE* were much more sensible to the reduction of the calibration dataset than ME_{10} and $RMSE_{10}$.

Impact of the large trees on the total biomass estimation

This section presents the variation of MET_{site} (Equation 9) and TE_{site} (Equation 10) when the heaviest trees are excluded from the simulated population. The idea was to examine the impact of the large trees on the mean biomass estimation using those models of Table 4 and those available in the literature. In all figures of this section, on the *x*-axis, 0% means that 0% of the *N* trees (see *N* in Table 1) was excluded, *x*-axis=1% means that the 1% heaviest trees were excluded from the *N* trees, and so on.

The first model was produced following *Context 1 – Pantropical biomass model of Chave et al. (2015)* calibrated with the full dataset comprising 4004 trees. Figure 3 shows a notable impact of the 5% heaviest trees on the mean biomass error, reaching a maximum underestimation of -35 kg tree^{-1} or -40 Mg ha^{-1}.

Table 4: Variation in model coefficients calibrated with datasets in which x% heaviest trees are excluded. Goodness-of-fit statistics are shown in Table 5. The resulting model when x = 0.0% corresponds to Equation (4) of Chave et al. (2015).

	x% heaviest trees excluded	Number of trees	Model	$\sigma_{\bar{\alpha}}$	$\sigma_{\bar{\beta}}$
x =	50.0	2,002	$\widehat{AGB}i = 0.093 + (\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,902}$	± 0.050	± 0.008
	10.0	3,602	$\widehat{AGB}i = 0.065 + (\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,971}$	± 0.026	± 0.003
	5.0	3,804	$\widehat{AGB}i = 0.066 + (\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,970}$	± 0.024	± 0.003
	2.0	3,924	$\widehat{AGB}i = 0.066 + (\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,970}$	± 0.022	± 0.003
	1.0	3,964	$\widehat{AGB}i = 0.065 + (\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,972}$	± 0.022	± 0.003
	0.5	3,984	$\widehat{AGB}i = 0.064 + (\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,974}$	± 0.021	± 0.003
	0.1	4,000	$\widehat{AGB}i = 0.063 + (\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,975}$	± 0.021	± 0.003
	0.0	4,004	$\widehat{AGB}i = 0.063 + (\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,976}$	± 0.021	± 0.003

σ: standard error of the model coefficient.

Table 5: Goodness-of-fit statistics of Equation (1) when calibrated with datasets in which x% heaviest trees are excluded. Model coefficients are shown in Table 4. The resulting model performance when x = 0.0% corresponds to the performance of Equation (4) of Chave et al. (2015).

	x% heaviest trees excluded	Number of trees	ME (kg)	RMSE (kg)	ME_{10} (kg)	$RMSE_{10}$ (kg)
			For the x% heaviest trees		For the ten heaviest trees	
x =	50.0	2,002	-1	14	-30,024	38,142
	10.0	3,602	-5	157	-15,783	21,718
	5.0	3,804	3	346	-15,871	21,812
	2.0	3,924	7	555	-15,815	21,753
	1.0	3,964	-11	701	-15,310	21,219
	0.5	3,984	-25	807	-15,048	20,944
	0.1	4,000	-55	1,046	-14,645	20,525
	0.0	4,004	-74	1,273	-14,521	20,397

ME: mean error, Equation (2). RMSE: root mean square error, Equation (3). Biomass estimates.

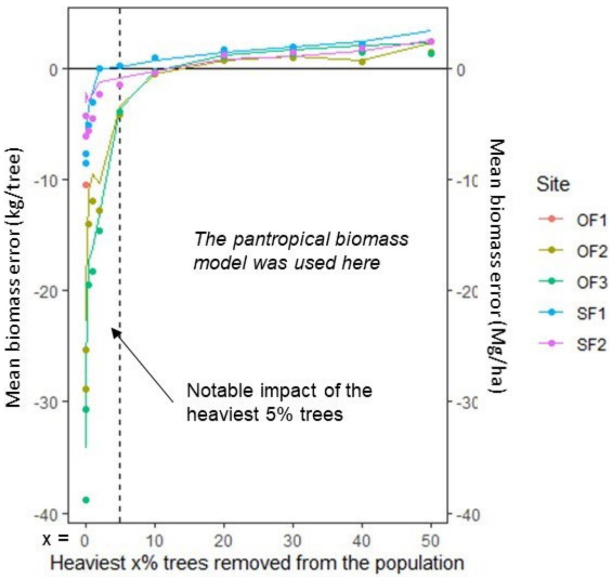

Figure 3: Error variation caused when the x% heaviest trees are excluded from the forest. The mean biomass (Mg ha^{-1}) is estimated using the tree biomass predicted from Equation (4) of Chave et al. (2015): ($\widehat{AGB}_i = e^{\left[0.0063+(\rho_i \cdot Dbh_i^2 \cdot H_i)^{0,976}\right]}$). OF1, OF2, OF3, SF1 and SF2 are tropical forest sites (see Table 1). Colored lines indicate error in kg $tree^{-1}$. Colored circles indicate error in Mg ha^{-1}. Dashed black line indicates x=5%.

The next analysis stems from *Context 2 – Pantropical biomass model of Chave et al. (2015) calibrated with reduced datasets*. Results are shown in Figure 4. The main difference between contexts 1 and 2 is that this time, the models shown in Table 4 were used rather than the original model calibrated with the full database with n = 4004 trees. Each graph in Figure 4 is specific to every forest site presented in Table 1. Figure 4 reveals an expressive underestimation of the mean biomass due to the presence of the 5% heaviest trees. Results suggest that, in general, the mean biomass error is relatively similar when 0–10% of the heaviest trees are removed from the pantropical biomass.

Notably, the model calibrated with the dataset reduced by the removal of the 50% heaviest trees (pink lines, Figure 4) produced sharper underestimations of the mean biomass than the other models. This result was expected because of the expressive differences in the coefficients (shown in Table 4) among such models. Graphs showing the relationship of observed and predicted tree biomass through the models of Table 4 are provided in Figure 2S (supplementary file).

The next results are from *Context 3 – The 'perfect' models*, where the same trees belonging to the simulated sites were used to calibrate Equation (1). Surprising results were seen because these models produced mean biomass errors slightly larger than the pantropical biomass model.

Results regarding *Context 4 – Biomass models from literature* show that, excepting Brown and Iverson (1992)'s model (Equation 5), the performance of the models yielded smaller errors when compared to the pantropical biomass model. Illustrative results are given in Figure 6.

DISCUSSION

The tyranny of the 5%

The select group of the 1% heaviest trees retained, on average, 35% (31–38%) of the total biomass in the sites of OF, and 25% (23–26%) in the sites of SF. These results suggest that the biomass concentration in humid tropical forests is comparable to the wealth concentration in developed countries. For example, in 2010, European wealthiest 1% held ~25% of the total wealth. In the United States of America, the wealthiest 1% held ~35% (Piketty, 2014). Results align to studies as Bastin et al. (2018), Sist et al. (2014) and Slik et al. (2013). These authors also noted that large trees explain much of the total AGB of tropical forests. The existence of a "tyranny of the large trees" is here revealed and notable, pushing humid tropical forests into the concept of *Extremistan* environment – this confirms the hypothesis initially raised. However, it is not only the 1% that compromises the biomass estimates.

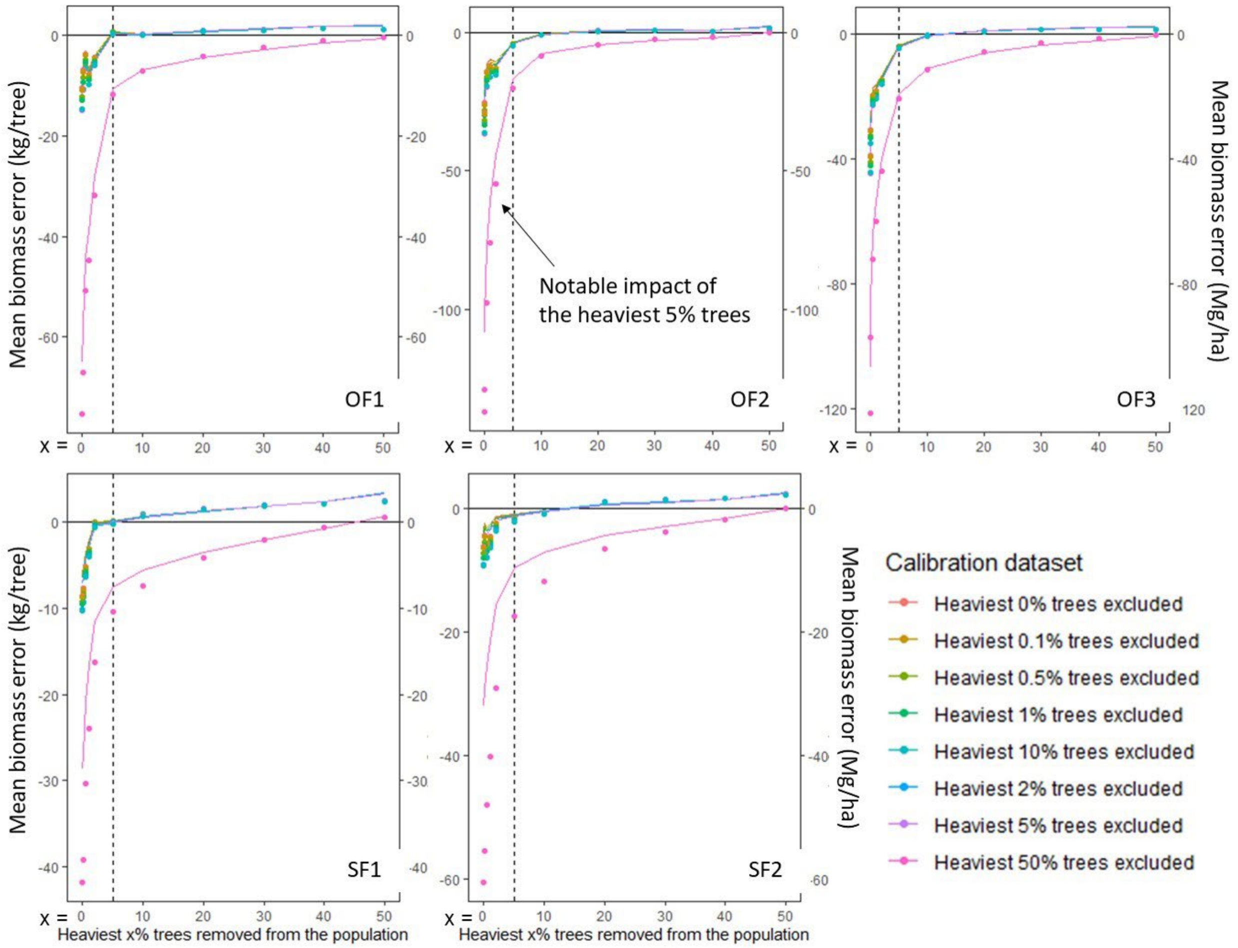

Figure 4: Error variation caused when the x% heaviest trees are excluded from the forest. The mean biomass (Mg ha^{-1}) is estimated through tree biomass predicted from Equation (1) calibrated with datasets reduced by the exclusion of the heaviest trees (see Table 3). OF1, OF2, OF3, SF1 and SF2 are tropical forest sites (see Table 1). Colored lines indicate error in kg tree^{-1}. Colored circles indicate error in Mg ha^{-1}. Dashed black line indicates x=5%.

The tyranny of the 5% heaviest trees is evident. These trees exert such a substantial impact that, regardless of the model used, they generally induce a systematic negative error in the mean biomass estimation. The impact of the 5% heaviest trees makes sense when considered that they retain 50–74% of the total biomass (Table 3). In several cases, the mean biomass per unit area would not be underestimated if the 5% heaviest trees did not exist, in which the mean biomass errors would get close to 0 kg tree^{-1}. Table 5 reveals that the tree-level biomass models tested can predict biomass of the small trees with reasonable accuracy, but they severely fail for the larger ones. For example, for the 10 heaviest trees of the pantropical database, the pantropical model of Chave et al. (2015) predicts tree biomass with an average error of 14.5 Mg/tree (last row of Table 5), which is equivalent to ~30%, corresponding to studies as Romero et al. (2022). The error on the biomass predictions of large trees is potentialized at the plot level if small *su's* are used. Results of Figure 2 and Table 1S (supplementary file) reveal that a single large tree may represent up to 90% of the plot biomass of a small *su*.

The tyranny of the 5% was somehow observed in other studies such as Sist et al. (2014), in which biomass in 18 Amazonian forests was assessed and reported that trees with Dbh ≥ 60cm retain >50% of the total biomass, i.e., a retention rate even higher than the one found in this study. In another study, Silk et al. (2013) found that the density of trees with Dbh ≥ 70cm explain ~70% of the pan-tropical AGB, and that these trees retained 25% (America), 39% (Asia) and 44% (Africa) of the total AGB. Note in Table 3 that when the Dbh ≥ 60cm and 70cm, the highest biomass retention was respectively 30% and 19%, suggesting that the results of this study may be conservative.

The general learning with this tyranny is that finding the most suitable tree-level model for an *Extremistan* forest seems to be a "waste of time". Note that models from the literature (Figure 6) in general performed similarly or better than the 'perfect' models (Figure 5) and the pantropical biomass model (Figure 3), this last fitted with the current largest calibration dataset for the tropical zone. Recall that the 'perfect' models did not help the biomass estimation to be more accurate. I.e., the best tree-level biomass model will probably not mitigate the impact of the tyranny of the 5%. One reason for this result is that tree biomass varies so much for the large trees that no tree-level model could capture well their biomass variability based on only allometric relationship. As the large trees are those responsible for the most biomass retained (Table 3), a model fitted with and without the 10% heaviest trees would produce, in general, similar outputs for the tree population. Figure 4 shows that the calibration datasets with removal of 0.1–10% of the heaviest trees worked similarly from each other but worked greatly better than that removal of the 50% heaviest trees. It means that the original pan-tropical biomass model by Chave et al. (2015) works very similar to if it was calibrated without the 10% heaviest trees. On the other hand, if the model was calibrated without the 50% heaviest trees, then the effect of this exclusion would become significant, both on the model coefficients (Table 4), biomass prediction of the heaviest trees (Table 5), and mean biomass estimation (Figure 4). The findings clearly indicate that the lack of those 50% makes the model predict the biomass of large trees with less accuracy and precision. Another remark is that the biomass underestimation provoked by the 5% heaviest trees (Figures 3–5) seemed to be inherent to Equation (1). Results confirm that Equation (1) does not always underestimate biomass of those heaviest trees after redoing several times the simulation process described in section 2.2.1 and redrawing Figures 3–5. Figures 3–5. were chosen from several because they faithfully portray the overall performance of Equation (1).

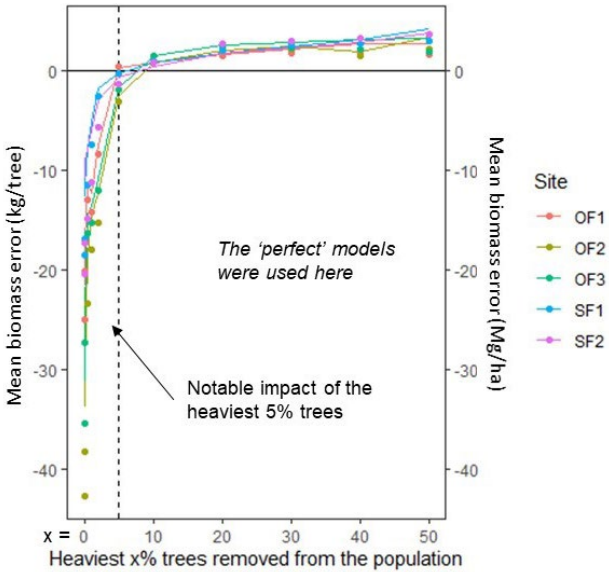

Figure 5: Error variation caused when the x% heaviest trees are excluded from the forest. The mean biomass (Mg ha^{-1}) is estimated through tree biomass predicted from Equation (1) calibrated with the own trees used to simulate the forest sites OF1, OF2, OF3, SF1 and SF2. Colored lines indicate error in kg tree^{-1}. Colored circles indicate error in Mg ha^{-1}. Dashed black line indicates x=5%.

Forest modelers typically assess their tree-level models using goodness-of-fit statistics such as RMSE and R^2, along with an analysis of model residuals. A low RMSE may create a misleading perception of "statistical security." For instance, if a model is calibrated with a dataset featuring many small trees and few large ones, the RMSE is likely to be lower than if calibrated with a dataset containing few small trees and many large ones. This discrepancy arises because RMSE is proportional to the magnitude of the data. While the relative form of RMSE can mitigate this issue, a high RMSE does not necessarily indicate a poor model for forest applications. Neither the absolute nor relative RMSE accurately reflects the error in stand-level or large-area estimations. For instance, the biomass models in Table 4 yielded RMSE values ranging from 14 to 1273 kg, as shown in Table 5. When these models were used to predict tree biomass and estimate the biomass error per hectare (illustrated in Figure 5), the model with the lowest RMSE resulted in the highest stand-level error. This outcome was due to the exclusion of the heaviest 50% of trees in the calibration dataset, leading to a low magnitude. Surprisingly, the other models produced similar stand-level errors despite a RMSE variation of 157 to 1273 kg. This indicates that the magnitude of the stand is more critical, and the presence of numerous trees can complicate analysis.

How large should a tree be to play the role of a Gray Swan?

Surprisingly, the tree-Gray Swan is smaller than one may expect. Though determining Gray Swans may be a bit subjective task, this study strongly indicates that the 5% heaviest trees are the Gray Swans. Results indicate that trees-Gray Swan were those with a Dbh >31 cm (Dbhs of 26–36 cm, see Table 3) for the OF, and with Dbh >18 cm (Dbhs of 16–20 cm) for the SF. The next section describes how to mitigate the impact of the Gray Swans in *Extremistan* forests.

Dealing with *Extremistan* forests

Quantifying biomass in general requires that a forest inventory be conducted. This study indicates that using large size *su's* in tropical forest inventories mitigates troubles related to the tyranny of the 5% heaviest trees. Findings suggest that *su's* ≥4000m^2 are safe for mature humid tropical forest, and 10000m^2 as the ideal size. The point is that large size *su's* mitigate the impact of large trees when they are observed in a *su*. The larger the *su's*, the more homogeneous their stocks, and this leads the sample towards a normal distribution. Small *su's* (e.g., <1000 m^2) would inflate the sample variance. Figure 2 illustrates the plot-level impact of the largest tree (here treated as tree-Gray Swan) found in the forest. The five trees-Gray Swan (one per site) of Figure 2 had Dbh ranging from 59 to 212 cm. Although the difference in Dbh, their impacts followed a similar trend in relation to the decreasing of the *su* size.

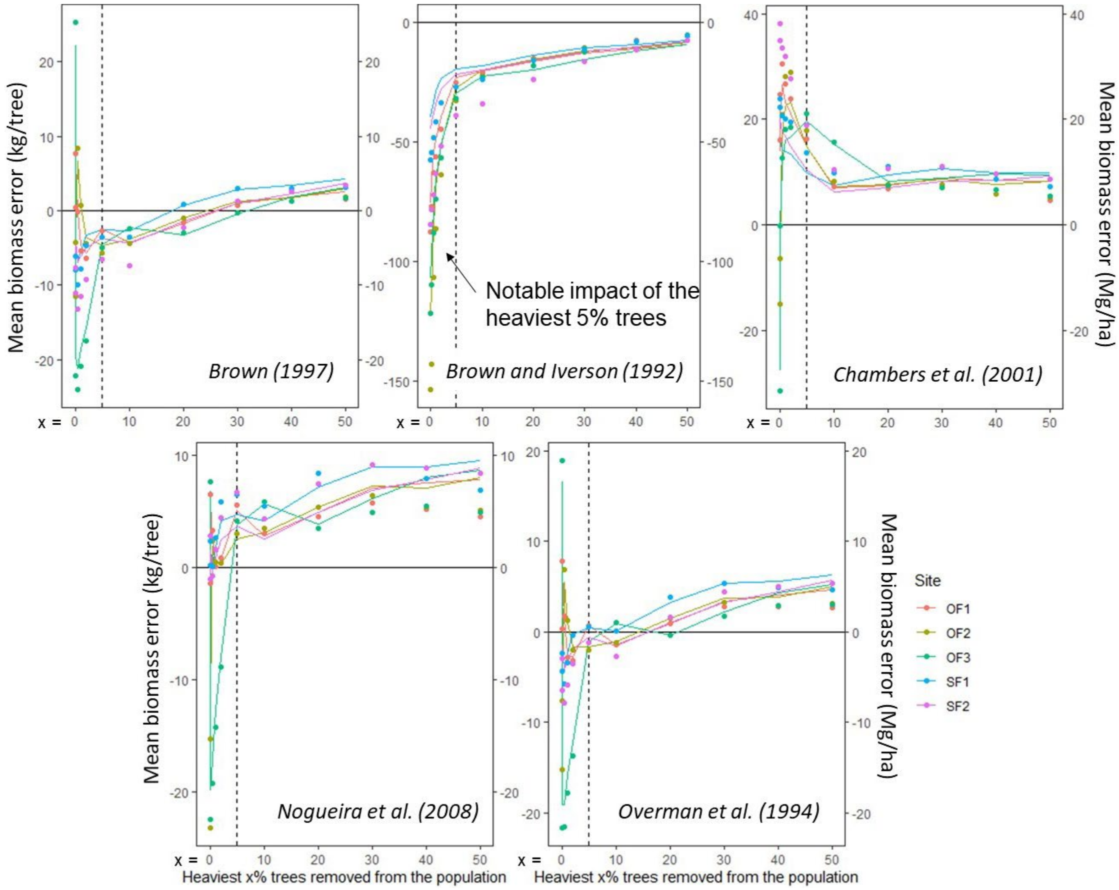

Figure 6: Error variation caused when the x% heaviest trees are excluded from the forest. The mean biomass (Mg ha^{-1}) is estimated through tree biomass predicted from Eqs. 4–8. OF1, OF2, OF3, SF1 and SF2 are tropical forest sites (see Table 1). Colored lines indicate error in kg tree^{-1}. Colored circles indicate error in Mg ha^{-1}. Dashed black line indicates x=5%.

Another practice that possibly potentialize an *Extremistan* forest is the way that tree biomass is modeled. Considering tropical tree species, the allometric relationship (biomass as a function of Dbh and h) works relatively well for small trees (Dbh <20 cm). The reason is because these trees, in general, present smaller stem volumes and crowns, and their branches tend to be less thick. Stems of small trees also tend to be more regular so that the tree variables Dbh and h capture well the stem volume. Based on empirical observations, this author indicates the crown (branches+foliage) as the main source of biomass variability in adult trees. Unfortunately, few forest scientists have published surveys about the biomass variability in compartments of tropical tree species. In Brazilian forests, biomass in the tree compartments was recently studied in Trautenmüller et al. (2021; 2023). Trautenmüller et al. (2023) tested models with and without stratification, in which strata corresponded to groups of Dbh size. For models fitted with a non-stratified calibration dataset, the biomass in branches and leaves was predicted with residuals of ±1000%, whereas the residuals for the trunks ranged between ±500%. While stratifying the calibration dataset, the authors obtained smaller residuals, but biomass errors in the branches and leaves remained higher than in the trunks. I.e, biomass in the crown is much more variable than in the trunks. Trautenmüller et al. (2021; 2023) also showed that biomass variation increases with the tree growth. Another expressive biomass variability in the large trees can be seen in Romero et al. (2022), while modeling biomass in only trees with Dbh ≥50cm. The trees with Dbh of 100 cm that composed their calibration dataset varied approximately between 2–13 MgC (this is carbon, not biomass); a 550% difference in weight for the same Dbh (see Romero et al. 2022's Figure 5). All these studies demonstrate to us that the biomass in the large trees (our Gray Swans) is inaccurately predicted at the tree level. Taking this study as example, if the 5% heaviest trees have their biomass inaccurately predicted, then 50–74% of the total biomass (Table 3) would consequently be inaccurately predicted. Lack of agreement between tree-level models was also examined in David et al. (2022). These authors simulated sites of Brazilian tropical forests and applied six tree biomass models for humid tropical forests (see David et al. 2022's Table 5), five of them are presented in Figure 6. In their first site, the six models produced estimates ranging from 262–469 Mg ha^{-1} (difference of 207 Mg ha^{-1}); in the second site, it was from 383–696 Mg ha^{-1} (difference of 313 Mg ha^{-1}); third site, 484–860 Mg ha^{-1} (difference of 376 Mg ha^{-1}); and in the fourth site, 488–851 Mg ha^{-1} (difference of 363 Mg ha^{-1}).

It is evident in the literature that tree-level models are much more commonly used than the stand-level models. Especially due to the large trees, this study suggests that, in tropical forests, the tree-level approach is likely impotent to reduce the mean biomass error. Further research exploring stand-level modeling is needed to verify if such an approach can mitigate the effects of the tyranny of the 5% heaviest trees.

CONCLUSION

Extremistan environments are exhibited in humid tropical forests, where large trees (referred to as tree-Gray Swans) are sheltered. In the Ombropilous Forest, tree-Gray Swans typically have a Dbh >31 cm, while in the Seasonal Forest, they have a Dbh >18 cm. The 1% heaviest trees, on average, retain 25–35% of the total biomass in *Extremistan* forests. There is a 'hidden tyranny' among the 5% heaviest trees, which has a significant impact on biomass modeling and estimation. In biomass and carbon inventories, installing sample units of 10,000 m² is effective in mitigating the effects of this tyranny, although a minimum area size of 4,000 m² can also help reduce these effects.

AUTHORS' CONTRIBUTION

This is a single-author paper, and all credit is attributed to the author.

ACKNOWLEDGEMENT

The author thanks the Brazilian agency CNPq by funding the project Tropical Biomass & Carbon through "Chamada CNPq/MCTI Nº 10/2023 – UNIVERSAL".

REFERENCES

BASTIN, J. F. et al. Pan-tropical prediction of forest structure from the largest trees. **Global Ecology and Biogeography**, v. 27, n. 11, p. 1366-1383, 2018.

BROWN, S. Estimating biomass and biomass change of tropical forests. A primer. FAO Forestry Paper 134. **Food and Agriculture Organization of the United Nations**, Roma, Itália, 1997. p. 55. Disponível em: http://www.fao.org/docrep/W4095E/W4095E00.htm.

BROWN, S.; GILLESPIE, A. F. R.; LUGO, A. E. Biomass estimation methods for tropical forests with applications to forest inventory data. **Forest science,** v. 35, p. 881–902, 1989.

BROWN, S.; IVERSON, L. R. Biomass estimates for tropical forests. **World Resource Review**, v. 4, p. 366–384, 1992.

CHAMBERS, J. Q. et al. Tree damage, allometric relationships, and above-ground net primary production in central Amazon forest. **Forest Ecology and Management**, v. 152, p. 73–84, 2001. Disponível em: https://doi.org/10.1016/S0378-1127(00)00591-0.

CHAVE, J. et al. Improved allometric models to estimate the aboveground biomass of tropical trees. **Global Change Biology,** v. 20, p. 3177–3190, 2015. Disponível em: http://dx.doi.org/10.1111/gcb.12629.

CHOWDHURY, E. K.; DHAR, B. K.; STASI, A. Volatility of the US stock market and business strategy during COVID-19. **Business Strategy & Development**, v. 5, n. 4, p. 350-360, 2022.

DAVID, H. C. et al. The tropical biomass & carbon project–An application for forest biomass and carbon estimates. **Ecological Modelling**, v. 472, p. 110067, 2022.

DAVID, H. C. et al. A 20-year tree liberation experiment in the Amazon: highlights for diameter growth rates and species-specific management. **Forest Ecology and Management**, v. 453, p. 1–10, 2019a. Disponível em: https://doi.org/10.1016/j.foreco.2019.117584.

DAVID, H. C.; et al. Exploring coarse- to fine-scale approaches for mapping and estimating forest volume from Brazilian National Forest Inventory data. **Forestry**, v. 92, p. 1–14, 2019b. Disponível em: https://doi.org/doi:10.1093/forestry/cpz030.

FISCHER, E. M.; SIPPEL, S.; KNUTTI, R. Increasing probability of record-shattering climate extremes. **Nature Climate Change**, v. 11, n. 8, p. 689-695, 2021.

HIGUCHI, N. et al. **Aboveground biomass estimate for Amazonian dense tropical moist forests.** Memoirs of the Faculty of Agriculture, Kagoshima University, v. 30, p. 43–54, 1994.

HOFMAN, J. M.; SHARMA, A.; WATTS, D. J. Prediction and explanation in social systems. **Science**, v. 355, n. 6324, p. 486-488, 2017.

MERZ, B. et al. Causes, impacts and patterns of disastrous river floods. **Nature Reviews Earth & Environment**, v. 2, n. 9, p. 592-609, 2021.

NOGUEIRA, E. M. et al. Estimates of forest biomass in the Brazilian Amazon: new allometric models and adjustments to biomass from wood-volume inventories. **Forest Ecology and Management**, v. 256, p. 1853–1867, 2008. Disponível em: doi.org/10.1016/j.foreco.2008.07.022.

OVERMAN, J. P. M.; WITTE, H. J. L.; SALDARRIAGA, J. G. Evaluation of regression models for above-ground biomass determination in Amazon rainforest. **Journal of Tropical Ecology**, v. 10, p. 207–218, 1994. Disponível em: https://doi.org/10.1017/S0266467400007859.

PALTRINIERI, N.; COMFORT, L.; RENIERS, G. Learning about risk: Machine learning for risk assessment. **Safety Science**, v. 118, p. 475-486, 2019.

PARRESOL, B. R. Assessing tree and stand biomass: a review with examples and critical comparisons. **Forest science,** v. 45, n. 4, p. 573–593, 1999.

PATÉ-CORNELL, E. On "black swans" and "perfect storms": Risk analysis and management when statistics are not enough. **Risk Analysis: An International Journal**, v. 32, n. 11, p. 1823-1833, 2012.

PIKETTY, Thomas. **Capital in the Twenty-First Century**. Belknap Press, 2014.

RISING, J.; TEDESCO, M.; PIONTEK, F.; STAINFORTH, D. A. The missing risks of climate change. **Nature**, v. 610, n. 7933, p. 643-651, 2022.

ROMERO, F. M. B. et al. Aboveground biomass allometric models for large trees in southwestern Amazonia. **Trees, Forests and People**, v. 9, 2022.

SLIK, J. F. et al. Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics. **Global Ecology and Biogeography**, v. 22, n. 12, p. 1261-1271, 2013.

SIST, P. et al. Large trees as key elements of carbon storage and dynamics after selective logging in the Eastern Amazon. **Forest Ecology and Management**, v. 318, p. 103–109, 2014. Disponível em: doi:10.1016/j.foreco.2014.01.005.

SOUZA, A. L. et al. Estoque e crescimento em volume, biomassa, carbono e dióxido de carbono em floresta estacional semidecidual. **Árvore**, v. 35, p. 1277–1285, 2011. Disponível em: http://dx.doi.org/10.1590/S0100-67622011000700014.

SMITH, R. J. Logarithmic transformation bias in allometry. **American Journal of Physical Anthropology**, v. 90, n. 2, p. 215-228, 1993.

TALEB, Nassim Nicholas. **The Black Swan: The Impact of the Highly Improbable**. Random House, 2007.

TRAUTENMÜLLER, J. W. et al. Regression estimators for aboveground biomass and its constituent parts of trees in native southern Brazilian forests. **Ecological Indicators**, v. 130, p. 108025, 2021. Disponível em: https://doi.org/10.1016/j.ecolind.2021.108025.

TRAUTENMÜLLER, J. W. et al. Ratio estimators for aboveground biomass and its parts in subtropical forests of Brazil. **Ecological Indicators**, v. 154, p. 110530, 2023.

YAROVAYA, L. et al. Rethinking financial contagion: Information transmission mechanism during the COVID-19 pandemic. **Journal of International Financial Markets**, **Institutions and Money**, v. 79, p. 101589, 2022.

SUPPLEMENTARY FILE

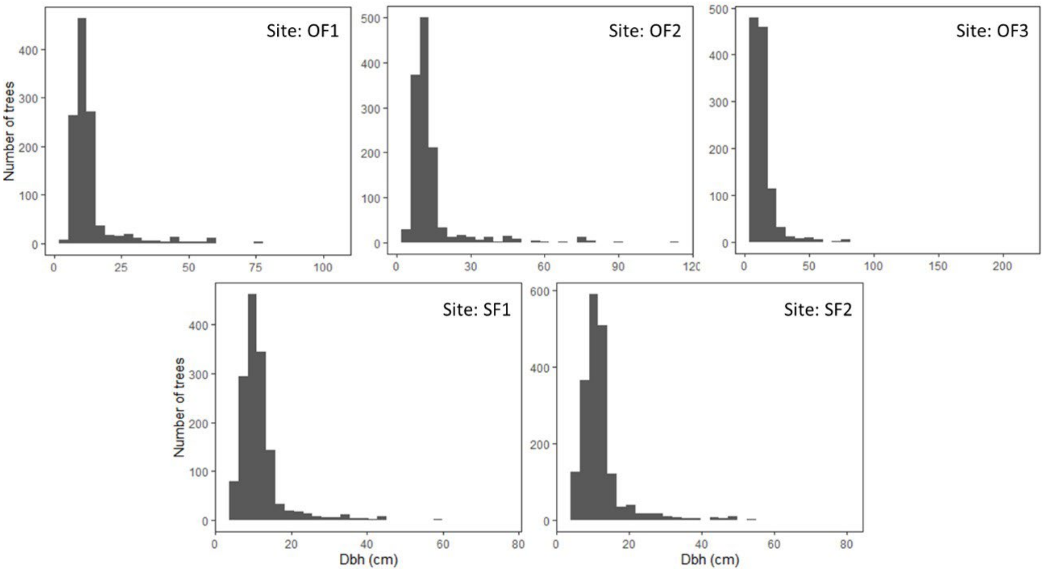

Figure 1S. Histogram of tree diameters of simulated populations based on the pantropical database. OF1, OF2, OF3, SF1 and SF2 are tropical forest sites (see Table 1).

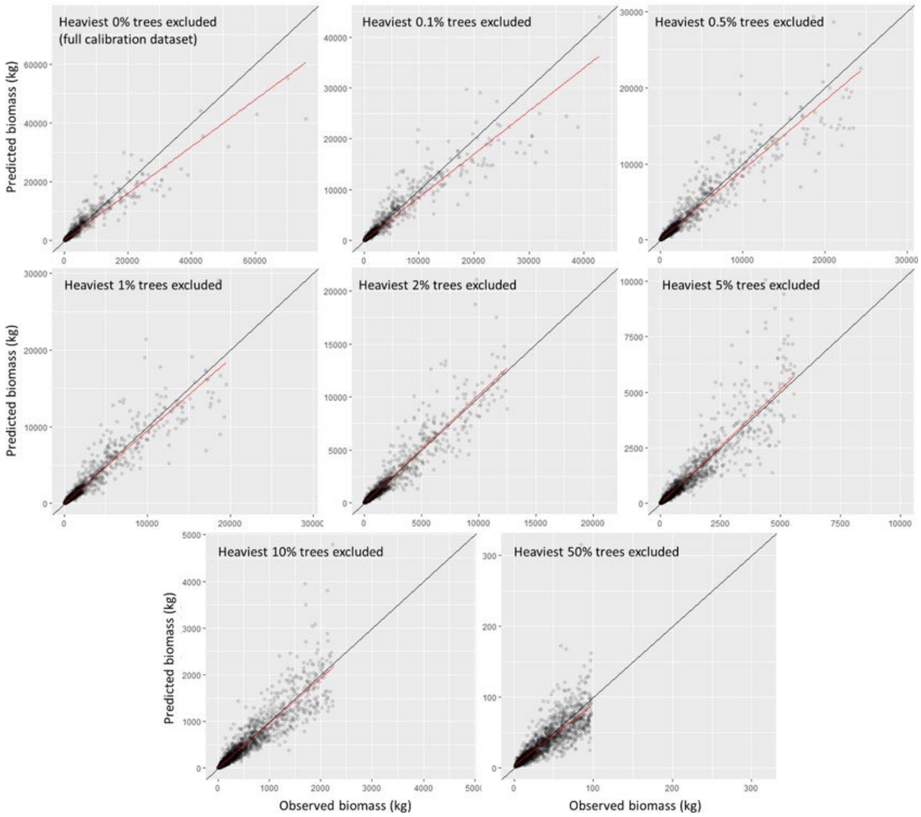

Figure 2S. Observed vs. predicted biomass through models calibrated with datasets reduced by the exclusion of the heaviest trees (see Table 3).

Table 1S. Representativeness in the plot-level biomass of the heaviest tree () in sample units (*su*) with different sizes and sites (The Dbh of the heaviest tree is shown in the top line of each site in Table 2).

Site	AGB^+	$\sum_{i=1}^{n-1}(AGB_i^-)$	su size (m²)				
			10,000	**5,000**	**2,500**	**1,250**	**125**
OF1	24.4	224.7	9.8%	17.8%	30.3%	46.5%	89.7%
OF2	20.8	357.0	5.5%	10.4%	18.9%	31.8%	82.3%
OF3	43.6	253.6	14.7%	25.6%	40.8%	57.9%	93.2%
SF1	9.2	151.1	5.7%	10.8%	19.5%	32.7%	82.9%
SF2	6.1	232.1	2.6%	5.0%	9.5%	17.3%	67.7%

AGB^+ and AGB_i^- are defined in Eq. (8).

Exponents in Table 2*S* should vary within and among the forests. Within, because tree population is dynamic over time and is susceptible to natural and external interventions. The main natural intervention is tree mortality, mainly of big trees. When a big tree dies, a considerable portion of the population biomass ends up migrating to the subsequent lower class of trees. In addition, the fall of a big tree drops many trees that is along the line of the fall, besides those ones pulled by entwined lianas. External interventions, as selective harvest, thinning and forest fires also alter the tree population, consequently affect the exponent.

Table 2S. Taleb's exponent of the fractal phenomena *biomass stored in tropical forests*.

Forest site	% population biomass retained by the:		Range of the exponent (θ)*
	heaviest 1% trees	**heaviest 20% trees**	
OF1 – Ombrophilous Forest 1	29%	83%	1.1–1.4
OF2 – Ombrophilous Forest 2	31%	86%	1.1–1.4
OF3 – Ombrophilous Forest 3	40%	84%	1.1–1.3
SF1 – Seasonal Forest 1	26%	69%	1.3–1.5
SF2 – Seasonal Forest 2	23%	71%	1.3–1.5

* Approximation of exponents tabulated in Taleb (2007).