

University of São Paulo
"Luiz de Queiroz" College of Agriculture
Center of Nuclear Energy in Agriculture

Aboveground biomass of Atlantic Forest: modeling and strategies for carbon
estimate

Michel Anderson Almeida Colmanetti

Thesis presented to obtain the degree of Doctor in
Science. Area: Applied Ecology

Piracicaba
2018

Michel Anderson Almeida Colmanetti
Bachelor of Science in Biology

Aboveground biomass of Atlantic Forest: modeling and strategies for carbon estimate

Advisor:
Prof. Dr. **HILTON THADEU ZARATE DO COUTO**

Thesis presented to obtain the degree of Doctor in
Science. Area: Applied Ecology

Piracicaba
2018

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Colmanetti, Michel Anderson Almeida

Aboveground biomass of Atlantic Forest: modeling and strategies for carbon estimate / Michel Anderson Almeida Colmanetti. - - Piracicaba, 2018.

127 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz". Centro de Energia Nuclear na Agricultura.

1. Floresta tropical 2. Biomassa 3. Carbono 4. Modelos preditivos I. Título

Dedico ao meu pai e à minha mãe, que com carinho, me apoiaram nessa jornada.

Acknowledgments

I thank God, the Greatest Teacher, and the Our Lady, our lovely Mother.

I thank my father, Altair, and my mother, Sonia, for the advisements, for supporting my decisions and for guidance along the way.

I thank my fiancée, Joyce K. Lucindo, for being patient and bringing happiness to my life.

I thank my brothers, Alex and João Paulo, for talks and advisement along the way.

I thank Professor Ph.D. Hilton Thadeu Zarate do Couto for the opportunity of learning under his advisement and for his friendship.

I thank Professor Ph.D. Aaron for hosting me in Orono, Maine, and for the advisement on chapters 2, 3 and 4.

I thank Robson Machado for hosting me in Orono, Maine, and for the friendship.

I thank Professor Ph.D. Christian Salas Eljatib from Universidad de La Frontera and professors from the University of São Paulo: Ph.D. João L.F. Batista, Ph.D. Ciro A. Righi, Ph.D. Plínio B. de Camargo and Ph.D. Luciano M. Verdade for suggestions along the way.

I thank Anderson R. Santiago for help in the construction of the maps.

I thank Jefferson L. Polizel for technical support and for the friendship.

I thank CAPES (scholarship process number: 88881.133157/2016-01) for the scholarship abroad in Orono, Maine/USA.

I thank the following companies and institutions: DERSA (especially Karina C. Barbosa), Mendes Júnior, and Castro Gardens for logistical support, members from CERAD for technical support (especially Luiz M. Barbosa, Regina T. Shirazuna, Paulo R. Ortiz, Fernando C. de Lima, Carlos Agena and Renata Ruiz), members from the Center for Research on Sustainable Forests - University of Maine.

I thank Professor Ph.D. Tiago C. Barbosa for field working and logistical support.

I thank Forestry Institute for logistical support (especially Marina Sherer).

I thank all guys from the lab CMQ for the friendship.

I also thank the friends and roommates, Renato, Isaias, Elizio, Anderson, Dener, Lucas, José Lucas, Pedro for the friendship and for the beers.

SUMMARY

RESUMO	7
ABSTRACT	8
1. GENERAL INTRODUCTION	9
REFERENCES.....	12
2. ABOVEGROUND BIOMASS AND CARBON OF THE HIGHLY DIVERSE ATLANTIC FOREST IN BRAZIL: COMPARISON OF ALTERNATIVE INDIVIDUAL TREE MODELING AND PREDICTION STRATEGIES	17
ABSTRACT	17
2.1. INTRODUCTION	17
2.2. MATERIALS AND METHODS.....	20
2.2.1. Study site.....	20
2.2.2. Data collection	21
2.2.3. Biomass, woody specific gravity and carbon content determination	23
2.2.4. Data analysis	24
2.2.4.1. Species-specific models.....	24
2.2.4.2. Generalized model.....	25
2.2.4.3. Functional trait models	26
2.2.4.4. Existing models	26
2.2.5. Model evaluation	27
2.2.6. Stand-level Biomass and Carbon Estimates.....	28
2.3. RESULTS	29
2.3.1. Generalized and species-specific models.....	29
2.3.2. Performance Across Species	32
2.3.3. Stand-level Biomass and Carbon	35
2.3.4. Less abundant species biomass prediction	37
2.4. DISCUSSION	39
2.4.1. Generalized and species-specific model	39
2.4.2. Biomass and carbon estimates at stand-level	41
2.4.3. Biomass for less abundant species	42
2.5. CONCLUSIONS	43
REFERENCES.....	45
3. ASSESSING VARIATION IN FOREST-LEVEL ABOVEGROUND BIOMASS AND CARBON ESTIMATES FOR A SPECIES RICH ATLANTIC FOREST IN BRAZIL: A CASE STUDY FOCUSED ON CANTAREIRA STATE PARK.....	53
ABSTRACT	53
3.1. INTRODUCTION	53
3.2. MATERIALS AND METHODS.....	56
3.2.1. Study site.....	56
3.2.2. Data collection	57
3.2.3. Data analysis	60
3.2.3.1. Variance of mean ($Var\mu B$)	61
3.2.3.2. Overall mean (μB)	61
3.2.3.3. Total biomass (τB).....	61
3.2.3.4. Variance of the total ($Var\tau B$).....	62
3.2.3.5. Estimates of sampling error ($E\%$).....	62
3.2.3.6. Stratified sampling	62
3.2.3.7. Individual tree biomass prediction.....	63
3.2.3.7.1. Species-specific approach:	63
3.2.3.7.2. Local generalized models approach:	64
3.2.3.7.3. Existing biomass models approach:	64
3.2.3.8. Forest-level biomass and carbon estimates.....	65
3.3. RESULTS	65
3.3.1. Simple and stratified sampling performances	65

3.3.2. Stand-level estimates of biomass and carbon	68
3.3.3. Forest-level estimates of biomass and carbon	70
3.4. DISCUSSION	71
3.4.1. Sampling efficiency.....	71
3.4.2. Biomass models and the accuracy of predictions	72
3.4.3. Implications for biomass and carbon estimates	74
3.5. CONCLUSIONS	76
REFERENCES	76
4. CALIBRATING INDIVIDUAL TREE BIOMASS MODELS FOR CONTRASTING TROPICAL SPECIES AT A DIVERSE SITE IN THE ATLANTIC FOREST, BRAZIL.....	83
ABSTRACT	83
4.1. INTRODUCTION	83
4.2. MATERIALS AND METHODS	86
4.2.1. Study site.....	86
4.2.2. Data collection	86
4.2.3. Biomass and woody specific gravity determination.....	87
4.2.4. Data analysis	88
4.2.4.1. Species-specific model.....	88
4.2.4.2. Generalized models.....	89
4.2.5. Model calibration procedures	90
4.2.5.1. Linear Mixed-Effect (LME) Approach.....	90
4.2.5.2. Ordinary Least Square (OLS)	91
4.2.6. Species and trees selection for a calibration procedure	91
4.2.7. Validation procedure	92
4.3. RESULTS	93
4.3.1. Species characterization	93
4.3.2. Calibration methods' performance	96
4.3.3. Differences across species and tree selection methods for calibration	101
4.4. DISCUSSION	104
4.4.1. Calibration methods' performance	104
4.4.2. Strategies for tree selection and the interference on LME calibration.....	105
4.5. CONCLUSIONS	107
REFERENCES	107
5. FINAL REMARKS	115
APPENDIX.....	117

RESUMO

Biomassa acima do solo da Mata Atlântica: modelagem e estratégias para a estimativa de carbono

Devido à atual preocupação do potencial efeito do CO₂ nas mudanças climáticas atribuiu-se à biomassa das florestas tropicais uma grande importância como reservatório de carbono. No entanto, a heterogeneidade dos ecossistemas naturais nos trópicos tem significativas implicações para a estimativa de sua biomassa. O presente estudo propõe diferentes modelos de biomassa utilizando amostragem destrutiva para Mata Atlântica, uma floresta altamente diversa. Duas abordagens de modelos: generalizados e espécies-específicos foram ajustados e o desempenho comparado. Em relação aos modelos generalizados, foram testadas diferentes covariáveis, utilizando o diâmetro à altura do peito (*dbh*; em inglês), a altura da base da copa, densidade básica da madeira (*wsg*; em inglês) e os “functional plant traits”. Os modelos espécies-específicos foram ajustados por modelos mistos lineares (LME; em inglês) utilizando as espécies como efeito aleatório e pelos mínimos quadrados (OLS; em inglês). O desempenho dos diferentes modelos e abordagens foi comparado ao desempenho de modelos existentes da literatura. Também foram verificadas diferentes estimativas de biomassa em nível de estande e floresta, assim como as implicações para a quantificação de carbono. Ainda, foram testados dois métodos de calibração para o modelo de biomassa em nível de árvore individual, variando o número de árvores e estratégias para seleção de árvores. Com base nos resultados, o modelo espécies-específicos usando LME apresentou melhor desempenho, podendo ser uma alternativa para as espécies mais abundantes, enquanto o modelo generalizado que inclui *dbh*, *wsg* e “functional plant traits” mostraram-se adequados para espécies menos abundantes. A calibração usando o método LME em alguns casos pode ser usada como uma alternativa para espécies que não possuem equação específica, sendo uma alternativa razoável para florestas tropicais altamente diversas, como a Mata Atlântica.

Palavras-chave: Floresta tropical; Biomassa; Carbono; Modelos preditivos

ABSTRACT

Aboveground biomass of Atlantic Forest: modeling and strategies for carbon estimate

The current concern on potential effect of CO₂ on climate change has assigned to the biomass of the tropical forest the importance as a sink of carbon. However, the heterogeneity of the natural ecosystems in tropics has significant implications for biomass estimation. This study proposed different biomass models using destructive sampling for the highly diverse Atlantic Forest. Models from two different approaches: generalized and species-specific were fitted and had the performance compared. Regarding the generalized models, it was proposed different covariates including diameter at breast height (*dbh*), height to the crown base, woody specific gravity (*wsg*) and functional plant traits. The species-specific models were fitted by linear mixed-models (LME) using species as a random effect and ordinary least square (OLS). The performance of all models and approaches were compared to existing models from the literature. Also, different estimates of biomass in stand- and forest-level, and the implications for carbon quantification were verified. Additionally, two methods for calibration for individual tree-level biomass model were proposed, and different strategies for tree selection were tested. The primary results show that the species-specific model using LME had better performance and can be used for the most abundant species, and models that include *dbh*, *wsg*, and plant traits are suitable for less abundant species. The calibration using the LME method in some cases can be used as an alternative for species that do not have a random effect presented here being a reasonable alternative for diverse tropical forests such as Atlantic Forest.

Keywords: Tropical forest; Biomass; Carbon; Predictive models

1. GENERAL INTRODUCTION

The potential effect of CO₂ on climate regulation has drawn the scientific community's attention to the carbon stocked in forests (IPCC 2006). The aboveground biomass is by far the most studied component once it is the most significant path of organic carbon input into forests.

The aboveground live tree biomass is predicted by models that vary according to levels of specifications in regional biomass conversion factors, stand-level biomass equations, and tree-level biomass equations (Temesgen et al. 2015). The commonly used method is to predict the individual tree biomass using the existing tree-level equations. These equations are often preferred once they require non-destructive sampling leading to less time consuming, less laborious in a field, and consequently reduction of the inventory's cost.

Individual tree biomass models are fitted using destructive sampling, where a relationship between biomass and diameter at breast height (*dbh*) is made (Brown et al. 1989; Chambers et al. 2001; Chave et al. 2001; Scatena et al. 2003; Overman et al. 1994; Burger and Delitti 2008; Nogueira et al. 2008). However, additional covariates as total height (Brown et al. 1989; Scatena et al. 2003; Overman et al. 1994; Burger and Delitti 2008) or woody specific gravity (*wsg*; Brown et al. 1989; Scatena et al. 2003) can also be included.

The generalized models (multi-specific models) are commonly used, having a general relationship between biomass and covariates, where the taxa are not weighted. In this context, the tree selection is a crucial step and widening the *dbh* range, stratifying the data sampling, and sampling a pool of species are recommended. It is expected the generalized model may efficiently predict the tree biomass from other sites within the same forest type since the assumption of interpolation is respected. However, the heterogeneity in the tropical forest leads to implications on biomass estimation, and the limitations associated with this procedure are often ignored.

The natural tropical forests are mosaics based on gap dynamics that have sites with different stages of succession varying species composition and the stand structure (Denslow 1987; Denslow and Guzman 2000; Guariguata and Ostertag 2001; Chazdon et al. 2010). However, selecting abundant species for biomass modeling is a reasonable choice due to the hyper-abundance on a large scale, the changing on most abundant species on smaller scale leads to heterogeneity between sites (Fauster et al. 2015). For example, if the most abundant species in one specific site are taken to fit a generalized model, it is expected accurately predictions of the trees in the same site, but it will unlikely have the same accuracy on predicting trees in sites with different species composition.

Some studies have shown the variation of the average woody specific gravity (Nogueira et al. 2007), and hypsometric relationship (Nogueira et al. 2008; Lima et al. 2012), across different regions in Amazon Forest as consequence of the species variation. Similar patterns seem to drive the biomass aboveground in Africa (Lewis et al. 2013) leading to different aboveground biomass estimates between sites and forests. So, it is expected biased biomass estimates when a generalized model fitted for one specific region is used in other (Nogueira et al. 2008).

The generalized biomass models are more useful and easily applied in tropical forest, particularly for carbon stock quantification at large spatial scales, and we have some efforts to fit predictive models based on massive amount of species that are useful in that scale (Chave et al. 2005, Feldpausch et al. 2011, Chave et al. 2014). However, more accurate approaches are usually based on single species at local or regional scales (Lapi 1991; Meng and Huang 2009; de Miguel et al. 2014; Arias-Rodil et al. 2015; Vismara et al. 2016).

Furthermore, some studies have suggested the improved performance of a species-specific approach for biomass prediction in tropical forests (e.g., Nelson et al. 1999; Sotomayor 2013). Since the most abundant species stock most of the biomass in tropical forests (e.g., van

Breugel et al. 2011; Fauster et al. 2015); the species-specific models may be a suitable approach to predict the biomass of those species. However, the high diversity in these forests leads to a variation in the species composition at different scales (Webb and Peart 2000; Slik et al. 2003; Réjou-Méchain et al. 2008; Eisenlohr and Oliveira-Filho 2014), which could impose a limitation for that approach. Nevertheless, the local calibration of previously developed species-specific biomass model for certain species could be a reasonable alternative for that issue (Vismara 2013), but this approach has received less attention in species rich areas like the Atlantic Forest of Brazil and further studies are required.

Based on the variation on species composition across the sites and the implication on biomass prediction, this study proposes the modeling of the aboveground biomass for Atlantic Forest considering a variation on biomass and diameter at breast height (*dbh*) relationship at species taxon level and suggesting the species-specific models as a feasible strategy. The performances of above and others strategies proposed were evaluated at the tree- and stand-level in the second chapter, and at forest-level in the third chapter. Additionally, a calibration for new species is proposed in the fourth chapter.

The general objectives are:

- To propose and assess the performance of generalized and species-specific models for individual tree biomass predictions, and verify impact on biomass and carbon estimates in stand- and forest level.
- Propose a calibration of species-specific models for aboveground biomass prediction of each new species sampled in an inventory.

The specific objectives are:

- develop generalized biomass equations using different covariates;
- evaluate different species-specific approaches using ordinary least squares (OLS) and linear mixed effects (LME);

- utilize various functional traits and species groups for biomass estimation;
- assess the performance of these various approaches as well as existing equations at the stand- and forest-level carbon estimates;
- to calibrate LME species-specific biomass model using BLUP to predict the random effect for a new species and compare to OLS calibration performance;
- verify the sample sizes, sampling methods of tree selection and *dbh* size on the calibration performances;
- compare all calibration approaches to generalized models performance;

The hypothesized are:

- Species-specific models lead to higher accuracy on biomass and carbon prediction than generalized or existing models.
- Calibrating the species-specific models is the feasible alternative for tropical forests regarding accuracy to predict the biomass for new species.

References

- Arias-Rodil, M.; Castedo-Dorado, F.; Cámara-Obregón, A.; Diéguez-Aranda, U. 2015. Fitting and Calibrating a Multilevel Mixed-Effects Stem Taper Model for Maritime Pine in NW Spain. PLoS ONE 10 (12): 1-20.
- Brown, S., Gillespie, A.J.R., Lugo, A.E., 1989. Biomass estimation methods for tropical forest with applications to forest inventory data. Forest Science 35(4): 881-902.
- Burger, D.M.; Delitti, W.B.C. 2008. Allometric models for estimating the phytomass of a secondary Atlantic Forest area of southeastern Brazil. Biota Neotropica 8(4):131-136.
- Chambers, J.Q.; Santos, J.; Ribeiro, R.J.; Higuchi, N.; 2001. Tree damage, allometric relationships, and above-ground net primary production in central Amazon Forest. Forest Ecology Management 152: 73-84.

- Chave, J.; Riera, B.; Dubois, M., 2001. Estimation of biomass in a Neotropical forest of French Guiana: spatial and temporal variability. *Journal of Tropical Ecology* 17: 79-96.
- Chave, J.; Andalo, C.; Brown, S.; Cairns, M.; Chambers, J.C.; Eamus, D.; Fölster, H.; Fromard, F.; Higuchi, N.; Kira, T.; Lescure, J.; Nelson, B.W.; Ogawa, H.; Puig, H.; Riéra, B.; Yamakura, T. 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* 145: 87-99.
- Chave, J.; Méchain, M.E.; Búrquez, A; et al. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*. DOI: <http://dx.doi.org/10.1111/gcb.12629>
- Chazdon, R.L.; Finegan, B.; Capers, R.S.; Salgado-Negret, B.; Casanoves, F.; Boukili, V.; Norden, N. 2010. Composition and dynamics of functional groups of trees during tropical forest succession in Northeastern Costa Rica. *Biotropica* 42(1):31-40.
- de-Miguel, S.; Mehtätalo, L.; Durkaya, A. 2014. Developing generalized, calibratable, mixed-effects meta-models for large-scale biomass prediction. *Can. J. For. Res.* 44: 648–656.
- Denslow, J.S. 1987. Tropical rainforest gaps and tree species diversity. *Annual Review of Ecology and Systematics* 18:431-51.
- Denslow, J.S.; Guzman, S.G. 2000. Variation in stand structure, light and seedling abundance across a tropical moist forest chronosequence, Panama. *Journal of Vegetation Science* 11: 201-212, 2000.
- Eisenlohr, P.V.; Oliveira-Filho, A.T. 2014. Tree species composition in areas of Atlantic Forest in southeastern Brazil is consistent with a new system for classifying the vegetation of South America. *Acta Botanica Brasilica* 28(2): 227-233.
- Fauset, S., et al., 2015. Hyperdominance in Amazonian forest carbon cycling. *Nature Communications* 6: 7857.

- Feldpausch, T.R. et al. 2011. Height-diameter allometry of tropical forest trees. *Biogeosciences* 8: 1081-1106.
- Guariguata, M.R.; Ostertag, R. 2001. Neotropical secondary forest succession: changes in structural and functional characteristics. *Forest Ecology and Management*, 148, (1):185-206, 2001.
- Intergovernmental Panel on Climate Change [IPCC], 2006. Intergovernmental Panel on Climate Change Guidelines for National Greenhouse Gas Inventories. Chapter 4: Forest Lands. Japan: National Greenhouse Gas Inventories Programme. 83 p.
- Lappi, J. 1991. Calibration of height and volume equations with random parameters. *Forest Science* 37: 781–801.
- Lewis S.L; Sonké, B.; Sunderland, T.; et al. 2013. Aboveground biomass and structure of 260 African tropical forests. *Philosophical Transactions of the Royal Society* 368: 1-14.
- Lima, A.J.N.; Suwa, R.; Ribeiro, G.H.P.M.; Kajimoto, T.; Santos, J.; Silva, R.P.; Souza, C.A.S.; Barros, P.C.B.; Noguchi, H.; Moriyoshi Ishizuka, M.; Higuchi, N. 2012. Allometric models for estimating above- and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil. *Forest Ecology and Management* 277: 163–172.
- Meng, S.X.; Huang, S. 2009. Improved calibration of nonlinear mixed-effects models demonstrated on a height growth function. *Forest Science* 55(3): 238-248.
- Nelson, B.W., Mesquita, R., Pereira, J.L.G., Souza, S.G.A., Batista, G.T. and Couto, L.B., 1999. Allometric regressions for improved estimate of secondary forest biomass in central Amazon. *Forest Ecology and Management* 117:149-167.
- Nogueira, E.M.; Nelson, B.W.; Fearnside, F.M.; França, M.B.; Oliveira, A.C.A. 2008. Tree height in Brazil's 'arc of deforestation': Shorter trees in south and southwest Amazonia imply lower biomass. *Forest Ecology and Management* 255: 2963–2972.

- Nogueira, E.M.N.; Fearnside, P.M.; Nelson, B.W.; França, M.B. 2007. Wood density in forests of Brazil's 'arc of deforestation': Implications for biomass and flux of carbon from land-use change in Amazonia. *Forest Ecology and Management* 248: 119–135.
- Overman, J.P., Witte, H.J., Saldarriaga, J.G., 1994. Evaluation of regression models for aboveground biomass determination in Amazon rainforest. *Journal of Tropical Ecology* 10(2): 207-218.
- Réjou-Méchain, M.; Péllisser, R.; Gourlet-Fleury, S.; Couteron, P.; Nasi, R.; Thompson, J.D. 2008. Regional variation in tropical forest tree species composition in the Central African Republic: an assessment based on inventories by forest companies. *Journal of Tropical Ecology* 24:663–674.
- Scatena, F.N., Silver, W., Siccama, T., Johnson, A., Sánchez, M.J., 1993. Biomass and nutrient content of the Bisley Experimental Watersheds, Luquillo Experimental Forest, Puerto Rico, before and after Hurricane Hugo, 1989. *Biotropica*. 25(1), 15-27.
- Slik, J.W.F.; Poulsen, A.D.; Ashton, P.S.; Cannon, C.H.; Eichhorn, K.A.O.; Kartawinata, K.; Lanniari, I.; Nagamasu, H.; Nakagawa, M.; van Nieuwstadt, M.G.L.; Payne, J.; Saridan, A.; Sidiyasa, K.; Verburg, R.W.; Webb, C.O.; Wilkie, P. 2003. A floristic analysis of the lowland dipterocarp forests of Borneo. *Journal of Biogeography* 30: 1517–1531.
- Sotomayor, J.F.M., 2013. Determinação de biomassa aérea em florestas nativas num ambiente agrícola do Estado de São Paulo. Dissertação (Mestrado em Ciências Florestais) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 130p.
- Temesgen, H.; Affleck, D.; Poudel, K.; Gray, A.; Sessions, J. 2015. A review of the challenges and opportunities in estimating above ground forest biomass using tree-level models. *Scandinavian Journal of Forest Research* 30(4): 326-335.

- van Breugel, M., Ransijn, J., Craven, D., Bongers, F. Hall, J.S., 2011. Estimating carbon stock in secondary forests: Decisions and uncertainties associated with allometric biomass models. *Forest Ecology and Management*. 262, 1648-1657.
- Vismara, E.S. 2013. Avaliação da construção e aplicação de modelos florestais de efeito fixo e efeitos mistos sob o ponto de vista preditivo. Chapter 4: Aplicação de modelos lineares de efeito misto e calibração na predição da biomassa em florestas tropicais multiespecíficas. Thesis, 106.p.
- Vismara, E.S.; Mehtätalo, L.; Batista, J.L.F. 2016. Linear mixed-effects models and calibration applied to volume models in two rotations of *Eucalyptus grandis* plantations. *Can. J. For. Res.* 46: 132–141.
- Webb, C.O.; Peart, D.R. 2000. Habitat associations of trees and seedlings in a Bronean rain forest. *Journal of Ecology* 88: 464-478.

2. ABOVEGROUND BIOMASS AND CARBON OF THE HIGHLY DIVERSE ATLANTIC FOREST IN BRAZIL: COMPARISON OF ALTERNATIVE INDIVIDUAL TREE MODELING AND PREDICTION STRATEGIES

ABSTRACT

The most common method to estimate forest-level carbon is prediction of individual tree biomass using existing regional and/or national equations with conversion to carbon using a fixed factor (e.g. 0.5). The Brazilian Atlantic Forest is a very important and structurally-diverse biome covering a large area with numerous species present, however, few biomass models exist for this region. This study evaluated alternative individual tree biomass models for the Atlantic Forest using contrasting prediction approaches, which were then applied to estimate stand-level biomass and carbon for a series of plots at a single location. The data comprised 106 destructively sampled trees with a large range in diameter at breast height (*dbh*; 5.4 – 68.5 cm) from the 16 most abundant species in a secondary forest of Atlantic Forest in Serra da Cantareira Mountains of Sao Paulo, Brazil. The approaches examined were species-specific, functional-trait based, generalized, and previously developed equations. Alternative model forms using differing variables including diameter at breast height (*dbh*), woody specific gravity (*wsg*), and height to crown base (*hcb*) were also examined. Species-specific models were developed using both ordinary least squares and linear mixed effects with species as a random effect. The approaches resulted in statistically different estimates of biomass and carbon at both the individual tree- and stand-levels. Compared to a generalized biomass equation for Pan-tropical forests, our results indicate differences of over 46% and 52% for biomass and carbon, respectively, in the secondary Atlantic Forest location examined. Overall, our findings highlight the challenges of accurately estimating biomass and carbon in species rich areas like tropical forests, but offers potential solutions that can be applied to the issue in future analyses.

Keywords: Tropical forest; Functional trait; Generalized models; Destructive sampling

2.1. Introduction

Due to the influence of increasing atmospheric carbon on climate change, significant efforts have been made in order to quantify carbon stock and its dynamics. Some studies indicate that tropical forests is an important sink of carbon (Pan et al., 2011), particularly in the Neotropics where a large forested areas are still preserved and have been broadly studied

(Brown et al., 1984, Brown et al., 1989, Scatena et al., 1993, Overman et al., 1994, Chambers et al., 2001, Nogueira et al., 2008). However, estimates of carbon often rely on empirically-based predictive equations, which have numerous limitations (Weiskittel et al., 2015), especially in the tropics where a greater number of species is present when compared to temperate and boreal regions.

In Brazil, the Atlantic Forest plays an important role in carbon sequestration as it is >16 million ha in area and only approximately 16% of its original area remains (Ribeiro et al., 2009). This forest is a spotlight of Brazilian conservation and restoration efforts due to current high levels of degradation and fragmentation (Myers et al., 2000). Although there is a recent increase in studies of the Atlantic Forest (Lima et al., 2015), biomass models are quite limited and rarely developed due to their time consuming, laborious, and high cost nature as well as restricted destructive sampling related to conservation initiatives (Brazilian Federal Law of Atlantic Forest Protection N. 11,428/2006).

The lack of individual tree biomass models for Atlantic Forest has led some studies (e.g. Rolim et al., 2008, Alves et al., 2010, Shimamoto et al., 2014, Marchiori et al. 2016) to use biomass models developed for other region, as Amazon Forest biomes (Brown et al., 1989, Scatena et al., 1993, Overman et al., 1994, Chambers et al., 2001, Chave et al., 2001), or even more generalized models for tropical forests (Chave et al., 2005, 2014). The generalized models of Chave et al. (2005, 2014) are based primarily on tree diameter, height, and wood specific gravity as well as various bioclimatic variables. Although these generalized equations are likely accurate at broad spatial scales, wood density (Nogueira et al., 2007) and tree diameter-height relations (Nogueira et al., 2008, Lima et al., 2012) often vary significantly at regional spatial scales. However, the implications of using generalized equations to estimate biomass or carbon at specific locations has not been well quantified, particularly in the diverse and relatively productive Atlantic Forest.

In addition to generalized biomass equations, several other types of estimation approaches exist including (1) genus and species-specific (e.g. Nelson et al., 1999, Huy et al., 2016), (2) species groups (e.g. Bastin, et al. 2015, Marra et al., 2016), and (3) based on functional trait (e.g. Marra et al., 2016). Although species-specific approaches are generally preferred as they often have better performance (Nelson et al., 1999), they require more extensive data than species groups and functional trait based methods (Weiskittel et al., 2015). An alternative and potentially more effective approach may be creating a hierarchical modeling approach that would effectively integrate the three approaches identified above as information can be shared across hierarchical levels. This has generally been done by using mixed-effects modeling and treating species group, genus, and species as a random effect. This has been recently used for individual tree height (Lam et al., 2017), height increment (Russell et al., 2014), and stem volume (MacFarlane and Weiskittel, 2016), but fewer applications for tree biomass have been examined (Sotamayor, 2013, Vismara 2013). Mixed-effects is an effective technique for species-specific predictions with a lower bias and root mean square error (RMSE) often outperforming more generalized models (Sharma and Paton, 2007, Crecente-Campo et al., 2013) or species-specific models developed using different methods (e.g. Russell et al. 2014).

However, due to the high diversity of species in tropical forests, species-specific models have generally been developed only for the most abundant species (Nelson et al., 1999). This is logical given that tropical forests are typically composed of a few hyperdominant species (Fauset et al., 2015). For non-dominant species, various strategies are used to provide biomass predictions, which can lead to contrasting estimates depending on the size and abundance of these species. In contrast, the use of species groups, functional traits, or other models that nest species with similar characteristics in common groups may not only be more reliable for the dominant species, but also have utility for the less abundant species. The improved performance

of generalized models has been previously reported (Marra et al., 2016), but further evaluation at multiple scales is needed.

Given the lack of existing individual tree biomass equations (only one existing model Burger and Delitti, 2008) and the application of generalized biomass equations in the Atlantic Forest of Brazil, the goal of this analysis was to develop alternative biomass and carbon estimation approaches specific to this region and evaluate the implications of their use. In particular, we evaluated the performance for the less abundant species given the potentially important implications for stand-level carbon estimates. Specific research objectives were to: (i) develop a generalized biomass equation that functioned across species, (ii) evaluate different species-specific approaches using ordinary least squares (OLS) and linear mixed effects (LME), (iii) utilize various functional traits and/or species groups for biomass estimation, and (iv) assess the performance of these various approaches as well as existing equations from the literature for predicting tree- and stand-level biomass and carbon.

2.2. Materials and Methods

2.2.1. Study site

The study site was a secondary Atlantic Forest located north of the city of São Paulo and southwest of Cantareira State Park (Parque Estadual da Cantareira, in portuguese) at an altitude of approximately 840 m (Figure 1). The vegetation of the region is classified as a moist forest with some species of seasonal semi-deciduous forest with a general predominance of species from the Euphorbiaceae, Lauraceae, Myrtaceae, Meliaceae, Sapotaceae and Rubiaceae families (Baitello et al., 1992). In this study, a general predominance of species was from the families of Myrtaceae, Fabaceae, Euphorbiaceae and Lauraceae. The climate at the region according to Köppen' climate classification adapted to Brazil is Cwa (Alvares et al., 2013), with an average annual temperature of 20.7 °C (minimum average: 15.7°C, maximum average:

25.7°C) and annual rainfall of 1,400 mm (Cepagri, 2017). The topography is characterized by steep slopes, and the soil classification according to soil taxonomy (Soil Survey Staff, 2014) was Ultisol Udic Udults Typic Hapludults.

2.2.2. Data collection

Ten 20 x 40 m rectangular plots (800 m²) were systematically established and installed. All trees ≥ 5 cm diameter at breast height (*dbh*) were measured for *dbh*, height-to-crown base (*hcb*, lowest live branch), and total height (*ht*). All of trees were collected for species identification and herbarium samples were incorporated into the Herbarium of the Botanical Institute of São Paulo, São Paulo City. Additional measurements of location/number of stem forks below *dbh*, potential disease, broken top, and presence of fruit or flowers were also taken. A total of 90 tree species belonging to 60 genera and 33 families were identified and updated according to Brazilian Species Flora List (www.floradobrasil.jbrj.gov.br) and Missouri Botanical Garden (www.tropicos.org) (Appendix A). The families were organized using the system A.P.G. IV (<http://www.mobot.org/MOBOT/research/APweb/>).

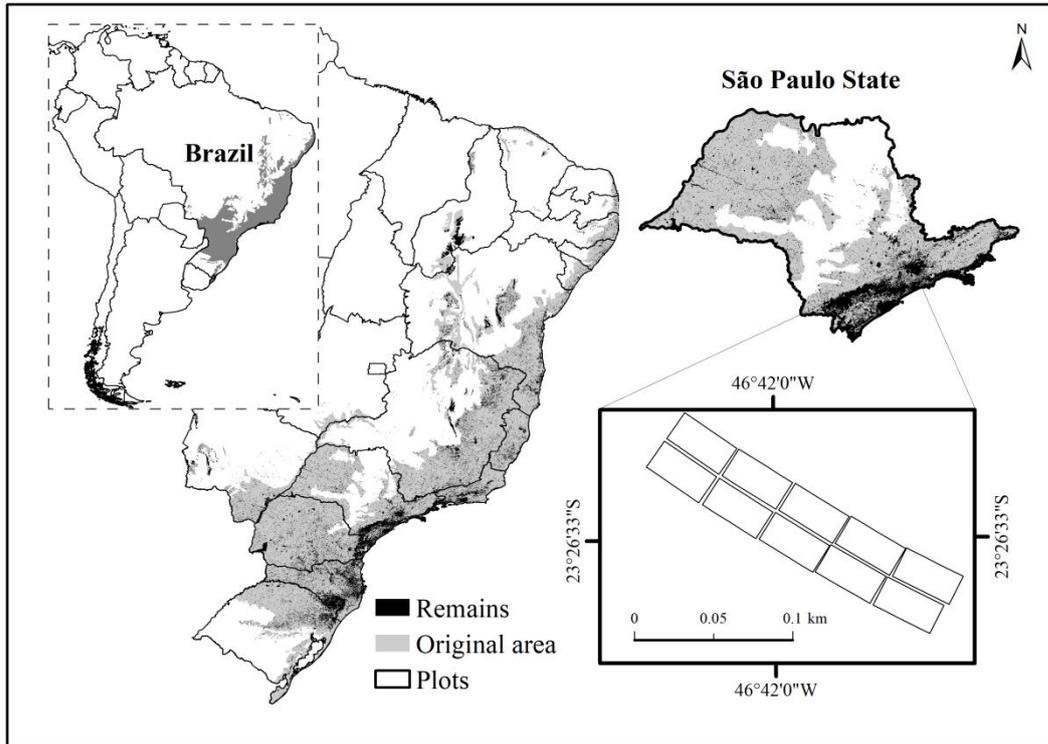


Figure 1. Plots on the study site located in Atlantic Forest near São Paulo, Brazil (sources: SOS Mata Atlântica, 2013, MMA, 2012, IBGE, 2010).

After collecting this plot-level inventory data, 96 trees belonging to the 16 most abundant species (~73% of the trees in each study area) plus 10 of the largest trees were destructively sampled. To ensure a range of *dbh*s, trees of three *dbh* classes (5-10 cm, 10-15 cm, and >15 cm) for each species were systematically selected for sampling. After harvesting, measurements of diameters were taken at various locations along the stem including stem base, breast height, and 50% below and above *hcb*. Samples of wood were collected (four disk of a constant thickness of 5 cm with bark; three for biomass determination and one for carbon determination) along the stem for additional analysis. The crown (foliage and branches) were weighed together, and a sample of both foliage and branch wood was taken. All samples were placed in plastic bags, and taken for biomass and carbon quantification.

2.2.3. Biomass, woody specific gravity and carbon content determination

All harvested trees were weighed completely in the field using a Dynamometer scale. For each tree, three discs were selected from the base, middle, and top of height to crown base. These discs were weighed after harvesting and then again after being dried at 105°C until reaching a constant weight. The same procedure was repeated for the crown samples. An average of the moisture content ($u\%$) for the three discs was calculated for each tree and used to estimate dry stem biomass and then summed with crown biomass to obtain total aboveground dry biomass:

$$(1) \quad db = m \cdot \left(1 - \left(\frac{u\%}{100}\right)\right)$$

where db is dry biomass, $u\%$ is average moisture content of three discs, and m is green biomass.

Wood specific gravity (wsg) was based on the variation of the sample mass using a hydrostatic balance. The method required the determination of stem mass samples at the maximum moisture content when immersed in water and the dry mass at 105°C in a forced-air oven. Calculation of wood specific gravity was as follows:

$$(2) \quad wsg = m_1 \cdot (m_2 - m_3)^{-1}$$

where wsg = woody specific gravity, m_1 = dry biomass at 105°C, m_2 = green biomass, and m_3 = green biomass submerged. An average of woody specific gravity for the three discs was obtained for each tree.

For carbon quantification, the disk at the breast height and the crown sample were dried at 60°C in a forced-air oven and then analyzed in the Leco Corporation Carbon Analyzer (Model C632, St. Joseph, MI) at 1350°C. All these procedures were carried out at the Center of Quantitative Methods at Forest Science Department at the University of São Paulo.

2.2.4. Data analysis

2.2.4.1. Species-specific models

Power equations in a linear form with log-log transformation were used to develop species-specific models of aboveground biomass (eq.(3) and eq.(4)). For this analysis, the fixed effects of models eq.(3) and eq.(4) was used as the generalized model. For this model, dbh and several other variables were all tested used as potential covariates. Species was used as random effect on both the slope and intercept as preliminary analysis indicated that the inclusion of a family and genus level random effects did not improve model performance. The model with only dbh was:

$$agb_{ij} = \phi_{1i} \cdot dbh_{ij}^{\phi_{2i}} \cdot \epsilon_{ij}$$

$$\ln(agb_{ij}) = \phi_{1i} + \phi_{2i} \cdot \ln(dbh_{ij}) + \epsilon_{ij}$$

$$(3) \quad \phi_i = \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} = \beta + b_i$$

$$b_i \sim N(0, \Psi), \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

where agb_{ij} is the aboveground biomass for the tree from the specie i with the dbh j , β are the fixed effects representing the population average of the ϕ_i , b_i is the random effect that represent the deviation from the population average, in this case the species i , with variance-covariance matrix Ψ . The errors were assumed to be independent for different species and within the groups (in this case the species), ϵ_{ij} are assumed to be independent for different i e j and to be independent of the random effects.

The models with two or more covariates was:

$$agb_{ij} = \phi_{1i} \cdot dbh_{ij}^{\phi_{2i}} \dots z_{ij}^{\phi_{ni}} \cdot \epsilon_{ij}$$

$$\ln(agb_{ij}) = \phi_{1i} + \phi_{2i} \cdot \ln(dbh_{ij}) + \dots + \phi_{ni} \cdot \ln(z_{ij}) + \epsilon_{ij}$$

$$(4) \quad \phi_i = \begin{bmatrix} \phi_{1i} \\ \vdots \\ \phi_{ni} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} b_{1i} \\ \vdots \\ b_{ni} \end{bmatrix} = \beta + b_i$$

$$b_i \sim N(0, \Psi), \quad \epsilon_{ij} \sim N(0, \sigma^2)$$

The β and Ψ were estimated by the restricted maximum likelihood estimation using the function *lme* from the *nlme* package (Pinheiro and Bates, 2000) in R (R Core Team 2016). The variance for the ϵ_{ij} was modeled by using the function *weights* from the *nlme* package, where the variance changes according to the covariates (eq.5):

$$(5) \quad \text{Var}(\epsilon_{ij}) = \sigma^2 v_{ij} v^{2\delta},$$

where v_{ij} is the covariate *dbh*, δ is the value for the variance parameter.

The assumptions of homoscedasticity and independence were verified with a visual examination of a scatter plot of standardized residuals versus fitted values, while normality was assessed by normal probability plot of the standardized residuals (Appendix B). Nested models were compared by AIC and likelihood ratio tests. Predicted aboveground biomass was back transformed to original scale to remove the systematical bias using the correction proposed by Sprugel (1983):

$$(6) \quad \widehat{agb} = e^{(\hat{\mu} + SEE^2/2)}$$

$$(7) \quad SEE = \sqrt{\frac{\sum (\ln(y_i) - \widehat{\ln}(y_i))^2}{N - n_p}}$$

where \widehat{agb} is the aboveground biomass, $\hat{\mu}$ is the estimated mean, SEE is the standard error of the estimates, $\ln(y_i)$ is the observed values for biomass, $\widehat{\ln}(y_i)$ is the predicted values for biomass, N is the number of observations, n_p is the number of parameters of the model.

For comparison, species-specific models were derived by using OLS where species was used as an indicator variable. Fitting was completed using the *lm* function in R (R Core Team 2013). The same model form and covariates were used for both OLS and LME. For OLS approach, variance power weighting as a function of *dbh* was tested but no improvement was observed.

2.2.4.2. Generalized model

For the generalized approach, only the fixed effects of models eq.(3) and eq.(4) mixed-effect models were used. Variables such as *dbh*, *wsg* and *hcb* were all tested as potential covariates. An overall average *wsg* was used for each species and was determined from average of all individuals *wsg* for each specie. The variance for the ϵ_{ij} was modeled according to *dbh* (eq.5), and the predicted aboveground biomass was back transformed to original scale using the correction proposed by Sprugel (1983; eq.6).

2.2.4.3. Functional trait models

For the functional trait models, the models eq.(3) and eq.(4) mixed-effect were used with successional stage (pioneer, early successional and late successional) and leaf habit (evergreen and non-evergreen) included as random effects. The variance for the ϵ_{ij} was modeled according to the covariates (eq.5) as described above. Also, the predicted aboveground biomass was back transformed to original scale using the correction proposed by Sprugel (1983) using eq.6.

2.2.4.4. Existing models

For comparison, two additional models were also obtained from the literature. One of them was a model developed for Atlantic Forest (Burger and Delitti 2008) based only on *dbh* (eq.8):

$$(8) \quad \ln(agb) = -3.068 + 2.522 \ln(dbh)$$

The other was a Pantropical model developed from a global database (Chave et al. 2014). The model (eq.9) was based on *dbh* and total height, which was predicted from a Pantropical hypsometric relationship (eq.10) as well as wood specific gravity obtained from the Global Wood Density database (Chave et al., 2009, Zanne et al. 2009), using the values only

for species occurring in South America. For these predictions, a genus-level average was used when species-level information was not available, while an average for the family was used when no information of both taxon levels information were available. Only four species (17 trees) had no *wsg* available for any taxon level, and an average for all species was used for these cases. This strategy was required since Chave et al. (2014) did not provide any model without *wsg*. The Pantropical hypsometric relationship was also defined by bioclimatic variables represented by eq.10.

$$(9) \quad agb = 0.0559(wsg.dbh^2.h)$$

$$(10) \quad \ln(h) = 0.893 - E + 0.760.\ln(dbh) - 0.034.\ln(dbh)^2$$

where E is bioclimatic variable taken from a gridded global layer at a 2.5 arc second resolution available at <http://chave.ups-tlse.fr/pan-tropical-allometry.htm>.

2.2.5. Model evaluation

The performance for the various approaches were compared by computing root mean square error (RMSE) (eq.11) and mean bias (MB, eq.12) by species for all trees used in the fitting procedure.

$$(11) \quad RMSE = \sqrt{\frac{\sum_{i=1}^n (a_{ij} - \hat{a}_{ij})^2}{n}}$$

$$(12) \quad MB = \frac{\sum_{i=1}^n (a_{ij} - \hat{a}_{ij})}{n}$$

where a_{ij} are the observed values of aboveground biomass, \hat{a}_{ij} are the predicted values of aboveground biomass, and n is the number of observations. Ranks were computed based on values of RMSE and MB by species, and an average was obtained for each model. Absolute values were used for MB. In addition, regression based equivalence tests as suggested by Robinson et al. (2005) were used. Equivalence tests reverse the usual null hypothesis, where it is considered the null hypothesis of dissimilarity (H_0 : a difference, i.e., $\mu \neq 0$). It means that if

two one-sided confidence intervals around the mean difference are entirely contained within the interval of equivalence, the H_0 is rejected. According to Robinson et al. (2005), the test is particularly suitable for validation procedure. For this analysis, region of equivalence was established as 25% of the mean (mean $\pm 25\%$) and slope ($1 \pm 25\%$). The analysis was done using the function *equiv.boot* from the package *equivalence* in R (R Development Core Team 2013).

2.2.6. Stand-level Biomass and Carbon Estimates

Biomass predictions for all trees in ten fixed-area plots were obtained by using all of the above described approaches. When a species did not have specific parameters (less abundant species), the generalized models were used to predict them. In this case, the best functional trait model (*dbh* and *wsg* model) was used to estimate stand-level biomass, and the wood specific gravity (*wsg*) for the species was obtained from the global database (Chave et al. 2009, Zanne et al. 2009). For global database, only the *wsg* for species occurring in South America were used, but a genus-level average was used when a specie-level information was not available, and when no information of both taxon levels information were available 42 trees from 12 species, the generalized model based only on *dbh* was used in these cases.

The functional trait model predictions were done using only leaf habit when a successional group was not available (e.g. non-evergreen pioneer). When neither information about the groups was available nor genus level identification available, the generalized model based on *dbh* only was used.

Carbon was calculated by multiplying the predicted biomass by the carbon content. Three alternatives were used: (1) a single value of 0.45 obtained from an average of all sampled species, (2) an average of carbon content for each sampled species and alternative 1 for non-sampled species, and (3) a carbon content fraction of 0.47 as suggested for tropical and subtropical forests (IPCC 2006). For biomass, the various estimates were compared using

analysis of variance (ANOVA), while a two-way ANOVA was used with the different carbon conversion factors as a factor for carbon estimates. The differences among the methods was evaluated using Tukey's HSD comparison test at a 95% level by using the *HSD.test* function from the package *agricolae* in R (R Development Core Team 2013).

2.3. Results

2.3.1. Generalized and species-specific models

The *dbh* in the dataset ranged from 5.4 to 68.5 cm. The largest species (*dbh* >40 cm) were *P. glabrata*, *C. speciosa*, *T. rubrivenium* and *P. gonoacantha* (Table 1). The species with lowest average wood specific gravity was *C. speciosa* (0.31 g cm^{-3}) and the highest was *C. oblongifolia* (0.68 g cm^{-3}). The species with the highest average crown/stem biomass relationship were *A. sidifolia*, *C. oblongifolia*, *C. sylvestris*, *P. glabrata*, and *P. gonoacantha*. A range of 42.9 to 47.5 % for average carbon content across species was observed, and a total average of ~45 % was obtained.

Table 1. Summary of the destructively sampled individual tree data from Atlantic forest at Serra da Cantareira-SP/Brazil. Heights were collected after the trees were harvested. *dbh* – diameter at breast height; *ht* – total height; *hcb* – high to crown base; *agb* – aboveground biomass.

Species	<i>dbh</i> (cm)		<i>wsg</i> (g cm ⁻³)		<i>ht</i> (m)		<i>hcb</i> (m)		<i>agb</i> (kg)		Carbon (%)	
	Range	Mean (sd)	Range	Mean (sd)	Range	Mean (sd)	Range	Mean (sd)	Range	Mean (sd)	Range	Mean (sd)
<i>Alchornea sidifolia</i> Müll. Arg. ¹	8.2 - 34.3	18.8 (12.0)	0.44 - 0.47	0.45 (0.01)	6.9 - 13.6	9.9 (2.6)	2.3 - 6.3	4.2 (1.6)	15.0 - 568.6	182.4 (233.4)	43.5 - 47.2	45.6 (1.3)
<i>Allophylus petiolulatus</i> Radlk. ³	7.6 - 20.8	13.1 (5.7)	0.50 - 0.55	0.53 (0.02)	9.4 - 12.8	10.8 (1.4)	3.8 - 7.9	6.2 (1.4)	12.7 - 199.5	67.0 (71.2)	43.9 - 45.9	44.8 (0.7)
<i>Cabralea canjerana</i> (Vell.) Mart. ³	7.0 - 11.8	9.3 (1.9)	0.35 - 0.45	0.42 (0.04)	8.6 - 11.2	9.3 (1.0)	3.2 - 8.4	5.2 (1.8)	7.6 - 25.8	17.5 (7.74)	44.1 - 47.2	45.4 (1.3)
<i>Casearia sylvestris</i> Sw. ²	8.2 - 20.4	13.2 (4.7)	0.54 - 0.59	0.55 (0.02)	7.7 - 10.6	9.4 (0.9)	2.5 - 6.0	4.0 (1.2)	18.1 - 137.5	64.1 (48.5)	43.7 - 46.1	45.0 (0.7)
<i>Ceiba speciosa</i> (A. St.-Hil.) Ravenna ^{2*}	6.0 - 67.8	32.1 (20.8)	0.19 - 0.37	0.31 (0.06)	5.9 - 17.3	12.3 (3.9)	2.7 - 11.4	8.4 (2.9)	2.3 - 1,114.6	294.8 (396.0)	41.2 - 44.7	42.9 (1.2)
<i>Croton floribundus</i> Spreng. ¹	7.6 - 30.0	16.8 (9.4)	0.45 - 0.58	0.53 (0.05)	9.5 - 15.8	12.4 (2.3)	6.4 - 9.1	7.1 (1.0)	21.9 - 415.9	153.6 (161.3)	43.7 - 47.0	45.3 (1.2)
<i>Cupania oblongifolia</i> Mart. ²	6.0 - 30.6	12.9 (9.1)	0.60 - 0.76	0.68 (0.05)	6.1 - 10.8	8.0 (1.7)	1.9 - 4.8	3.1 (1.2)	9.7 - 445.9	100.4 (170.5)	42.4 - 45.4	44.1 (1.0)
<i>Jacaranda puberula</i> Cham. ^{2*}	6.3 - 20.6	12.9 (5.8)	0.30 - 0.39	0.34 (0.03)	7.4 - 12.6	9.7 (2.1)	2.8 - 7.7	5.8 (1.9)	4.0 - 148.9	44.2(54.9)	44.6 - 48.5	47.5 (1.5)
<i>Machaerium villosum</i> Vogel ^{3*}	7.7 - 27.2	15.3 (7.6)	0.56 - 0.61	0.58 (0.02)	8.6 - 14.6	11.3 (2.7)	5.0 - 8.9	6.8 (1.6)	12.9 - 300.7	104.6 (120.6)	44.4 - 46.5	45.5 (0.8)
<i>Myrcia splendens</i> (Sw.) DC. ²	6.2 - 14.9	11.8 (3.4)	0.48 - 0.56	0.51 (0.03)	4.3 - 10.4	8.6 (2.4)	3.5 - 6.2	4.8 (1.1)	9.4 - 80.5	51.7 (27.6)	44.4 - 46.2	45.1 (0.8)
<i>Nectandra oppositifolia</i> Ness ²	7.3 - 26.5	15.4 (7.5)	0.39 - 0.51	0.45 (0.05)	8.8 - 14.0	11.4 (2.2)	5.7 - 9.5	7.3 (1.5)	11.4 - 251.8	101.7 (96.9)	46.1 - 48.9	47.5 (1.0)
<i>Pera glabrata</i> (Schott) Poep. ex Baill.	5.4 - 68.4	25.0 (23.6)	0.47 - 0.65	0.57 (0.06)	5.7 - 16.5	10.2 (3.9)	2.9 - 8.0	4.6 (1.8)	5.6 - 1,917.1	466.0 (701.8)	42.9 - 45.2	44.4 (0.8)
<i>Piptadenia gonoacantha</i> (Mart.) J. F. Macbr. ^{2*}	7.1 - 42.7	20.3 (13.2)	0.54 - 0.64	0.58 (0.03)	8.6 - 18.3	13.9 (3.7)	4.0 - 11.8	8.0 (2.9)	10.6 - 711.6	250.1 (293.4)	45.0 - 46.4	45.5 (0.6)
<i>Sessea brasiliensis</i> Toledo ³	8.2 - 22.8	15.3 (4.6)	0.38 - 0.59	0.50 (0.07)	8.0 - 14.2	11.2 (2.4)	3.0 - 9.9	6.8 (2.7)	10.1 - 122.6	70.0 (42.4)	44.6 - 49.8	46.6 (1.7)
<i>Tetrorchidium rubrivenium</i> Poep. ²	8.6 - 58.3	27.1 (20.1)	0.36 - 0.45	0.40 (0.03)	9.6 - 17.3	14.1 (3.0)	5.1 - 9.6	7.7 (1.8)	16.0 - 1,493.8	407.4 (536.1)	45.7 - 48.1	47.2 (0.8)
<i>Vochysia tucanorum</i> Mart. ²	8.4 - 18.8	12.9 (4.7)	0.40 - 0.47	0.43 (0.03)	8.5 - 16.0	11.6 (2.7)	5.0 - 11.0	7.4 (2.1)	18.8 - 91.7	46.1 (34.6)	44.0 - 46.1	44.7 (0.9)
Overall	5.4 - 68.5	17.3 (13.0)	0.19 - 0.76	0.49 (0.10)	4.3 - 18.3	10.9 (3.0)	1.9 - 11.8	6.2 (2.4)	2.3 - 1,917.1	158.5 (297.7)	41.2 - 49.8	45.4 (1.6)

¹Pioneer, ²Early Secondary, ³Late Secondary, *Non-evergreen species

The generalized models were obtained from the fixed part of the linear mixed-effects models using only dbh (eq.3), and dbh and wsg (eq.4), where both equations are already corrected for log-transformation:

$$(13) \quad agb = \exp^{-2.245+2.388.\ln(dbh)}, \sigma = 0.1524$$

$$(14) \quad agb = \exp^{-1.293+2.389.\ln(dbh)+1.373.\ln(wsg)}, \sigma = 0.1633$$

where agb is the aboveground biomass; dbh is the diameter at breast height; wsg is the woody specific gravity, and σ is the standard deviation of the random effect.

No additional model improvements were observed when including other potential covariates like hcb for the LME approach. The species-specific parameters for both OLS and LME are provided in Table 2.

Table 2. Species-specific parameters estimates using linear mixed effects (LME) and ordinary least squares (OLS). The intercept for all models was corrected for log-transformation. Standard error of the intercept and slope in LME was common to all species and equal to 0.338 and 8.864 e⁻⁰⁷, and for OLS equal to 0.157 and 0.046 for β_0 , β_1 , respectively. β_0 , β_1 and β_2 are parameters related to intercept, dbh and $species$, respectively.

Specie	LME		OLS		
	β_0	β_1	β_0	β_1	β_2
<i>A. petiolulatus</i>	-2.192	2.389	-2.185	2.389	
<i>A. sidifolia</i>	-2.246	2.389	-2.185	2.389	-0.055 (0.144)
<i>C. canjerana</i>	-2.487	2.389	-2.185	2.389	-0.316 (0.150)
<i>C. floribundus</i>	-1.896	2.389	-2.185	2.389	0.298 (0.150)
<i>C. oblongifolia</i>	-2.07	2.389	-2.185	2.389	0.138 (0.144)
<i>C. speciosa</i>	-3.293	2.389	-2.185	2.389	-1.137 (0.144)
<i>C. sylvestris</i>	-2.093	2.389	-2.185	2.389	0.090 (0.144)
<i>J. puberula</i>	-2.693	2.389	-2.185	2.389	-0.521 (0.149)
<i>M. splendens</i>	-2.052	2.389	-2.185	2.389	0.153 (0.149)
<i>M. villosum</i>	-2.29	2.389	-2.185	2.389	-0.082 (0.150)
<i>N. oppositifolia</i>	-2.174	2.389	-2.185	2.389	0.028 (0.150)
<i>P. glabrata</i>	-2.104	2.389	-2.185	2.389	0.052 (0.144)
<i>P. gonoacantha</i>	-2.229	2.389	-2.185	2.389	-0.016 (0.145)
<i>S. brasiliensis</i>	-2.38	2.389	-2.185	2.389	-0.193 (0.144)
<i>T. rubrivenium</i>	-2.327	2.389	-2.185	2.389	-0.161 (0.142)
<i>V. tucanorum</i>	-2.338	2.389	-2.185	2.389	-0.172 (0.150)

The plant functional trait models were based on two classifications, namely successional stage and leaf habit. However, species belonging to the non-evergreen pioneer group were not present in the harvested trees sampled. The parameters were corrected for the log-transformation (Table 3). For the evergreen group, the early and late-successional species have a smaller intercept in comparison to pioneers. For the non-evergreen species, the late successional species intercept was smaller than the early successional. When *wsg*, was included in the model, an improvement in both RMSE and MB was observed.

Table 3. Parameter estimates using linear mixed-effect for functional trait models. The intercept for all models was corrected for log-transformation. G.D – linear mixed-models based on functional trait (only *dbh* as covariate), G.DW – linear mixed-models based on functional trait (*dbh* and *wsg* as covariate). Standard error of the intercept and slope in LME (σ) was common to all species and equal to 0.209 and $2.128e^{-08}$ for β_0 and β_1 , respectively, for G.D, and 0.158 and $1.08e^{-08}$ for β_0 and β_1 , respectively, for G.DW. β_0 , β_1 and β_2 are parameters related to Intercept, *dbh* and *wsg*.

Random Effect	Random Effect	G.D		G.DW		
		β_0	β_1	β_0	β_1	β_2
Leaf Habit	Non-evergreen	-2.349	2.339	-1.691	2.403	1.151
	Evergreen	-2.114	2.339	-1.384	2.404	1.152
Successional Stage inside Leaf Habit groups	Non- Evergreen/E.Successional	-2.486	2.339	-1.573	2.405	1.153
	Non- evergreen/L.Successional	-2.098	2.339	-1.572	2.406	1.154
	Evergreen/Pioneer	-2.085	2.339	-1.431	2.407	1.155
	Evergreen/E.Successional	2.147	2.339	-1.521	2.408	1.156
	Evergreen/L.Successional	-2.341	2.339	-1.631	2.409	1.157

2.3.2. Performance Across Species

In terms of RMSE and MB (Table 4), the best model was the plant trait model with *dbh* and *wsg* and the species-specific LME model, respectively. Based on the overall mean rank for both RMSE and MB, the best model was the species-specific LME model followed closely by the plant trait model with *dbh* and *wsg*. The functional traits model with just *dbh* underperformed the generalized model with only *dbh*, while both generalized equations tended to outperform the OLS

species-specific approach. The equations from Burger and Delitti (2008) and Chave et al. (2014) had the lowest performance as they tended to underestimate and overestimate, respectively.

At the species level, overestimation was most prevalent for *C. speciosa*, *T. rubrivenium*, *P. gonoacantha* and *P. glabrata* (Table 4). Chave et al. (2014), the generalized model, and Burger and Delitti (2008) all overestimated *C. speciosa*, which was the species with lowest *wsg*, more than other models

In terms of the regression-based equivalence tests using a region of equivalence of 25%, all examined models did not differ for the intercept or slope, except the equation from Chave et al. (2014). This result indicates similarity between predicted and observed values for an area of equivalence of 25%.

Table 4. Relative root mean square (RMSE) and mean bias (MB, observed - predicted) for the generalized, species-specific, and existing models from the literature for predicting individual tree total aboveground biomass in the Atlantic Forest at Serra da Cantareira of Sao Paulo, Brazil. D – Generalized model (only *dbh*) based on fixed-effects from linear mixed effects (LME), DW– Generalized model (*dbh* and *wsg*) based on fixed-effects from LME, LME - Linear Mixed-Effects Species-Specific, OLS - Ordinary Least Squares Species-Specific, G.D – LME for functional trait model with DBH as only covariate, G.DW – LME for functional trait model with *dbh* and *wsg* as covariates, B&D – Burger and Delitti (2008) model for Atlantic Rain Forest, Chave – Chave et al. (2014) Pantropical model. The values in bold were the best method for each species.

Species	RMSE								MB							
	Generalized		Species Specific		Traits		Existing		Generalized		Species Specific		Traits		Existing	
	D	D.W	LME	OLS	G.D	G.DW	B&D	Chave	D	D.W	LME	OLS	G.D	G.DW	B&D	Chave
<i>A. petiolulatus</i>	24.28	23.41	23.26	23.18	27.93	23.81	43.06	22.51	4.09	-4.9	0.72	0.24	9.68	2.26	26.75	-0.99
<i>A. sidifolia</i>	47.28	51.69	47.25	47.41	61.99	47.39	87.02	62.9	-2.78	13.18	-2.57	-3.53	-26.04	-3.35	47.03	-21.87
<i>C. canjerana</i>	6.63	2.89	2.2	2.1	5.03	2.32	4.27	8.92	-5.71	-1.8	-0.72	-0.46	-4.21	-0.99	3.68	-7.54
<i>C. floribundus</i>	29.87	19.5	58.78	60.82	18.75	28.65	85.04	21.56	23.7	4.24	-30.58	-32.06	3.52	-8.03	66.2	12.29
<i>C. oblongifolia</i>	27.52	55.15	2.27	4.55	16.6	39.89	72.23	74.85	14.57	-27.01	0.73	-1.24	8.06	-20.16	38.67	-31.66
<i>C. speciosa</i>	577.76	128.21	125.09	133.82	158.15	75.09	329.01	668.64	-386.02	-81.12	56.31	63.12	-106.77	-26.49	-209.85	-432.12
<i>C. sylvestris</i>	17.57	22.4	19.93	19.85	17.96	21.87	33.73	54.55	4.6	-8.69	-5.15	-5.01	-1.95	-8.34	26.35	-36.14
<i>J. puberula</i>	27.91	25.04	24.36	24.71	25.29	30.81	23.59	26.37	-16.59	6.38	5.35	5.89	6.29	13.42	5.32	-13.93
<i>M. splendens</i>	15.57	13.97	13.52	13.76	13.42	13.66	29.4	39.52	8.76	5.28	-0.35	-1.4	3.59	4.34	25.14	-30.55
<i>M. villosum</i>	18.73	29.17	23.48	21.07	30.12	20.99	59.98	124.32	7.38	-20.54	11.67	9.56	16.31	9.78	40.49	-83.81
<i>N. oppositifolia</i>	12.85	16.1	15.94	17.44	16.4	13.83	46.72	59.99	4.38	11.18	-2.76	-4.48	-4.89	8.4	37.66	-34.22
<i>P. glabrata</i>	247.7	499.73	394.83	361.56	277.65	365.97	115.74	1266.62	-80.36	-220.58	-163.13	-144.72	-104.24	-160.65	57.02	-611.92
<i>P. gonoacantha</i>	92.86	144.35	93.19	94.64	169.01	95.05	140	354.77	21.33	-47.33	17.72	11.23	112.2	31.69	90.08	-181.97
<i>S. brasiliensis</i>	29.88	34.42	24.44	24.46	24.89	27.97	28.71	50.86	-10.51	-16.21	-0.35	-0.43	-3.23	-8.42	18.27	-29.99
<i>T. rubrivinium</i>	301.93	211.52	258.15	249.46	327.45	205.88	212.81	774.73	-107.5	1.64	-66.93	-57.62	-133.44	1.78	28.55	-418.49
<i>V. tucanorum</i>	15.11	5.51	9.13	8.01	22.3	8.2	11.21	26.01	-10.63	-2.22	-5.59	-4.6	-16.94	-4.91	10.22	-17.87
Rank	4.25	4.44	3.00	3.38	4.62	3.62	5.62	7.06	4.44	4.69	2.75	2.75	4.56	3.75	5.94	7.12

2.3.3. Stand-level Biomass and Carbon

Stand-level biomass and carbon varied widely using the alternative estimation approaches and showed statistically significant differences (Figure 2). The equation of Bugar and Delitti (2008) had the lowest estimates ($72.2 \pm 21.5 \text{ Mg ha}^{-1}$, mean \pm SD), followed by trait-based model with *dbh* only ($93.3 \pm 21.9 \text{ Mg ha}^{-1}$), OLS ($103.7 \pm 29.1 \text{ Mg ha}^{-1}$), and the generalized model without *wsg* ($107.6 \pm 28.8 \text{ Mg ha}^{-1}$), LME ($108.7 \pm 31 \text{ Mg ha}^{-1}$), generalized with *wsg* ($112.1 \pm 34 \text{ Mg ha}^{-1}$) and trait-based with *wsg* ($116.3 \pm 22.4 \text{ Mg ha}^{-1}$).

Comparing the species-specific models for the most abundant species, OLS ($104.3 \pm 29.4 \text{ Mg ha}^{-1}$) produced lower estimates than LME ($108.7 \pm 31 \text{ Mg ha}^{-1}$) and the functional trait model with *dbh* and *wsg* ($116.3 \pm 22.4 \text{ Mg ha}^{-1}$). For the existing models from the literature, the Chave et al. (2014) equation ($157.9 \pm 48.8 \text{ Mg ha}^{-1}$) was more than 100% higher those provided by the Bugar and Delitti (2008) equation, and 35.8 % greater than the functional trait model with *dbh* and *wsg*, the second largest estimate.

The largest difference between the carbon estimates (Figure 3) were observed for the previously existing models, which were $32.5 \pm 9.7 \text{ (Mg ha}^{-1}\text{)}$ for Bugar and Delitti (2008) and 74.2 ± 22.9 for Chave et al. (2014), using 0.45 (average factor based on destructively sampled trees) and 0.47 (IPCC factor), respectively. The models had estimates that were intermediate of those estimates such as $42 \pm 9.9 \text{ (Mg ha}^{-1}\text{)}$ for functional trait-based with *dbh*, $46.9 \pm 13.2 \text{ (Mg ha}^{-1}\text{)}$ for OLS species-specific, $48.4 \pm 13 \text{ (Mg ha}^{-1}\text{)}$ for generalized model with *dbh*, $48.6 \pm 13.8 \text{ (Mg ha}^{-1}\text{)}$ for LME species-specific, $50.4 \pm 15.3 \text{ (Mg ha}^{-1}\text{)}$ for generalized model with *dbh* and *wsg* and $52.4 \pm 10.1 \text{ (Mg ha}^{-1}\text{)}$ for functional trait-based with *dbh* and *wsg*, and using the 0.45 carbon conversion. Using 0.47 carbon conversion factor resulted in an average increase of 2.2 Mg ha^{-1} (4.4%) in comparison to 0.45 carbon convert factor, and 2 Mg ha^{-1} (4%) in comparison to carbon

conversion factor by species when compared to other stand-level carbon estimates. The species-specific carbon content factor had a slightly higher average value of 0.2 Mg ha⁻¹ (~0.5%) in comparison to the constant 0.45 carbon factor.

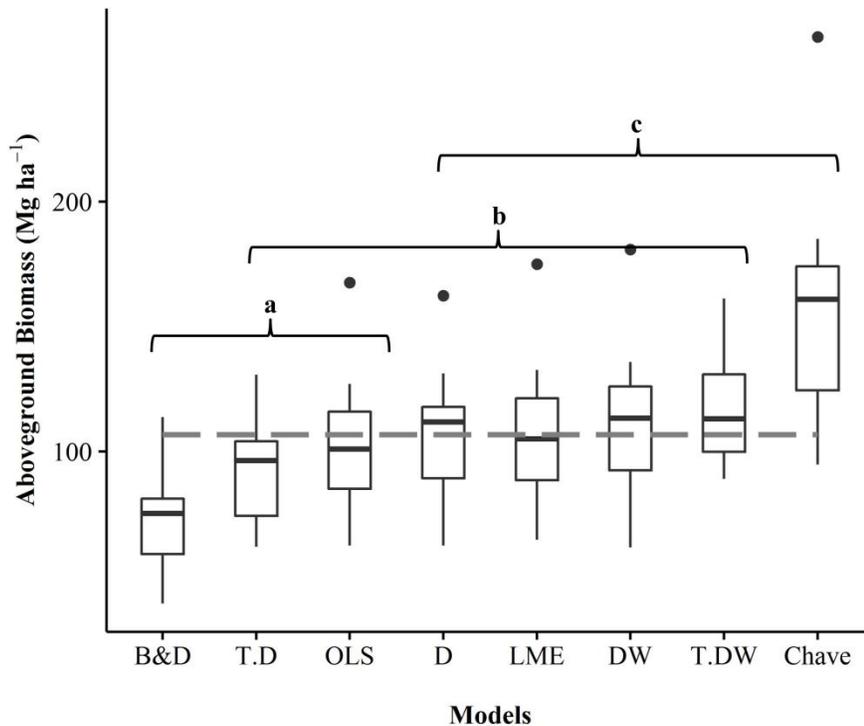


Figure 2. Stand-level estimates of biomass (Mg ha⁻¹) using different approaches including Burger and Delitti (2008) – Atlantic Forest model, T.D – LME model based on functional trait with *dbh* as covariate, OLS - ordinary least squares species-specific model, D – generalized model (only *dbh*) based on fixed-effects from linear mixed effects (LME), LME – LME species-specific model, DW – generalized model (*dbh* and *wsg*) based on fixed-effects from LME, T.DW – LME model based on functional trait with *dbh* and *wsg* as covariates, Chave et al. (2014) – Pantropical model. Broken gray line represents a median of all biomass estimates (106.7 Mg ha⁻¹). The letters are the statistical groups according to Tukey’s HSD comparison test at 95% level of confidence.

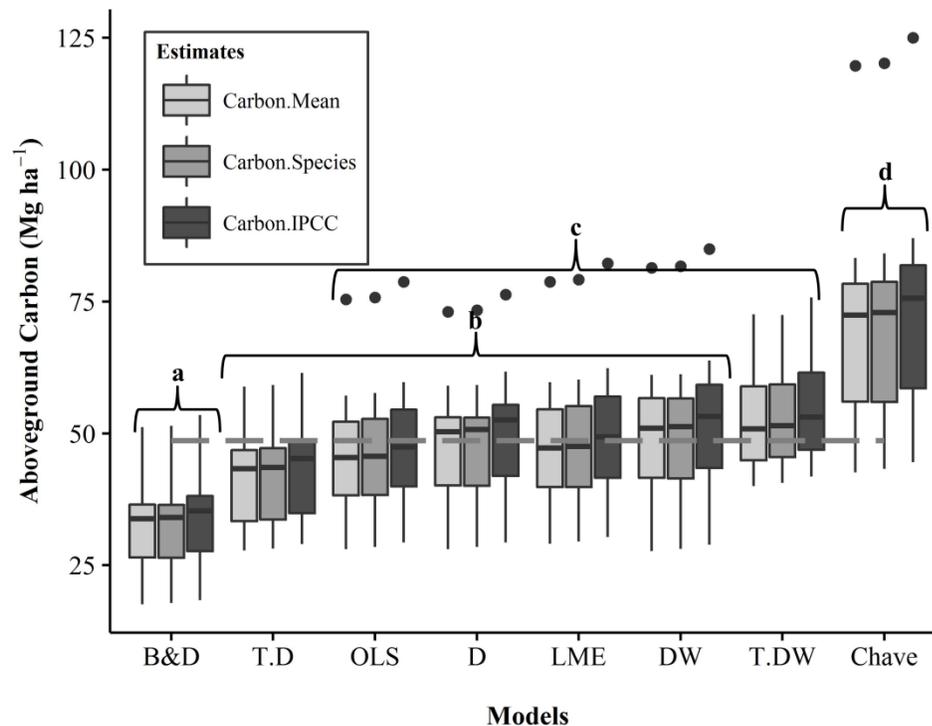


Figure 3. Stand-level estimates of carbon (Mg ha^{-1}) using different approaches including Burger and Delitti (2008) – Atlantic Forest model, T.D – LME model based on functional trait with *dbh* as covariate, OLS - ordinary least squares species-specific model for abundant species and functional trait based for less abundant (*dbh* and *wsg*), D – generalized model (only *dbh*) based on fixed-effects from linear mixed effects (LME), LME – LME species-specific model for abundant species and functional trait based for less abundant (*dbh* and *wsg*), DW – generalized model (*dbh* and *wsg*) based on fixed-effects from LME, T.DW – LME model based on functional trait with *dbh* and *wsg* as covariates, Chave et al. (2014) – Pantropical model, Carbon.Mean – observed biomass to carbon transformation factor based on destructively sampled trees (0.45), Carbon.Species – observed biomass to carbon transformation factor based by species, Carbon.IPCC – IPCC biomass to carbon transformation factor (0.47). Broken gray line represents a median of all carbon estimates (48.1 Mg ha^{-1}). The letters are the statistical groups according to Tukey’s HSD comparison test at 95% level of confidence.

2.3.4. Less abundant species biomass prediction

Different estimates of stand-level biomass (Figure 4) were obtained using the various estimation strategies, particularly for the less abundant trees, 74 species and approximately 27% of total trees. The less abundant species biomass corresponded to 18% of the total biomass (median of less abundant species 19 Mg ha^{-1} / median of the total biomass 106.7 Mg ha^{-1}). The smallest estimate was from functional trait-based model with *dbh* only ($19.8 \pm 11.9 \text{ Mg ha}^{-1}$), and the

greatest was using generalized model with *dbh* and *wsg*, using a family-level average of *wsg* ($29.7 \pm 19.1 \text{ Mg ha}^{-1}$). The functional trait model using *dbh* and *wsg* with species and genus average for *wsg* had similar estimates of $23.2 \pm 14.2 \text{ Mg ha}^{-1}$ and $22.9 \pm 15.6 \text{ Mg ha}^{-1}$, respectively. Higher estimates were obtained using average of *wsg* at family-level, than using specie- or genus-level.

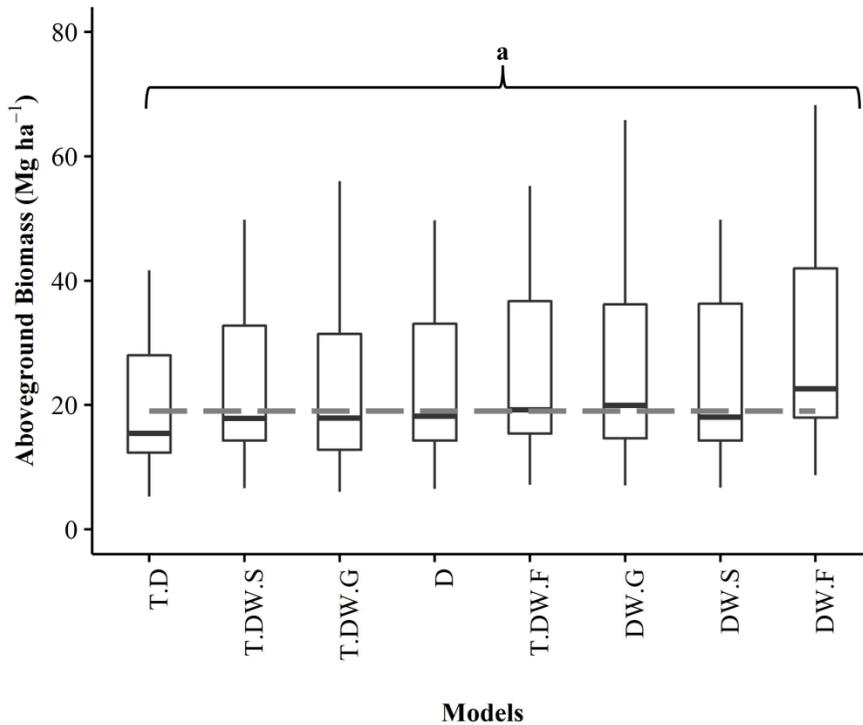


Figure 4. Stand-level estimates of biomass (Mg ha^{-1}) for all of the less abundant species in the studied plots. D – generalized model using only *dbh*, DW.S– generalized model using *dbh* and *wsg* (specie-level average of *wsg*), DW.G– generalized model using *dbh* and *wsg* (genus-level average of *wsg*), DW.F– generalized model using *dbh* and *wsg* (family-level average of *wsg*), T.D – functional trait-based using only *dbh*, T.WD.S – functional trait-based using in *dbh* and *wsg* (specie-level average of *wsg*), T.WD.G – functional trait-based using in *dbh* and *wsg* (genus-level average of *wsg*), T.WD.F – functional trait-based using in *dbh* and *wsg* (family-level average of *wsg*). Broken gray line represents a median of all biomass estimates (19 Mg ha^{-1}). No statistical differences across groups according to Tukey’s HSD comparison test at 95% level of confidence was observed.

2.4. Discussion

2.4.1. Generalized and species-specific model

The overall best model in this analysis was the LME species-specific model, which reaffirms the prior finding that species-specific approaches generally outperform more generalized models (Nelson et al., 1999). In addition, species-specific developed using LME was superior to the species-specific models from OLS, which indicates that additional information is gained by using species as a random effect. Across methods, models generally performed best with the inclusion of both *dbh* and *wsg* as covariates. Although the total height (*ht*) is often used in biomass models, we did not evaluate this variable in our analysis since tree crowns have significant overlap in tropical forests so measurements often have low efficiency and accuracy. In contrast, *hcb* is much easier to measure and likely a better predictor of total biomass. However, *hcb* or other similar metrics did not generally improve model performance for the species examined.

For species-specific models, *wsg* was not included since this trait varies more between than within species (Baker et al. 2004; Chave et al. 2006). In addition, no improvement had been reported by adding *wsg* in species-specific (Nelson et al., 1999) or genus-level models (Huy et al., 2016), which is similar to the findings in this analysis. Although the incorporation of additional hierarchical groups could potentially improve the LME species-specific models in this analysis such as family or genus (Huy et al., 2016), no additional improvements were found likely due to the relatively limited number of families and genus destructively sampled. Future analyses may consider the incorporation of these hierarchical levels as Lam et al. (2017) also found significant improvement in prediction accuracy of total tree height of various tropical forest species with their inclusion.

The functional trait approach using *dbh* only did not perform well. However, when *wsg* was added, much better performance was observed and in fact, it was the best performing approach overall following the species-specific models. Similar improvements in model performance with the inclusion of *wsg* have been found too (e.g. Marra et al., 2016). Besides the importance of variables like *dbh* and *wsg* on the improvement of the performance, the random effects were also important (leaf habit and successional groups). Although Powers and Tiffin (2010) suggested the leaf habit had little importance in forming functional groups and other groupings such as classifying the species as Fabaceae and non Fabaceae in their analysis were more important, our results suggested high usefulness of the leaf habit trait. Using leaf habit and successional groups as a random effects in a hierarchical structure performed much better than models without them. In addition, they also performed better than a legume/non-legume as well as just a successional group models (not presented). Although leaf habit alone was not sufficient for predictions in this analysis, other characteristics when added to leaf habit may provide reasonable ecological groups (Powers and Tiffin 2010) as found in this analysis. In the future, likely other functional traits could be combined with leaf habit for further evaluation. For instance, these characteristics could include crown architecture type or shade-tolerance since some Atlantic Forest species have similar characteristics (Larcher et al. 2012). Some traits such as *wsg* do have a rather narrow distribution of values within certain shade-tolerance groups (Bastin et al. 2015), which suggest the need for this type of information. Like Marra et al. (2016), the addition of functional trait groups improved the robustness of the model and warrants further exploration.

Overall, the existing models generally did not perform well at this site. Although the Burguer and Delitti (2008) model performed better than the generalized Pan-tropics model of Chave et al. (2014). Burger and Delitti (2008) model underestimated biomass for all species, except

C. speciosa. Likely, this underestimation was due to the small range of *dbh* (1.6-47.8 cm) used in their sample dataset and the fact that they were all obtained from a young secondary forest (30 years old). In contrast, the Chave et al. (2014) model overestimated biomass for all species, except *C. floribundus*. The greater number of larger trees harvested from across the tropics including both primary and secondary forests may have caused the Chave et al. (2014) model to overestimate for the smaller diameter classes. The same trend was observed by van Breugel et al. (2011) for a young secondary tropical forest in central Panama. Potentially, the fundamental idea of a generalized biomass equation for all tropical species biomass is unlikely and the application scale of the approach should probably be limited. Regardless, additional examination of the approach is warranted, but a general caution of broad scale application is suggested.

2.4.2. Biomass and carbon estimates at stand-level

All of our estimates of stand-level biomass and carbon were lower than previous estimates at other similar sites (Table 5), except for those provided by Chave et al. (2014) model ($157.9 \pm 48.8 \text{ Mg ha}^{-1}$). As previously mentioned, the lack of existing biomass models for Brazilian Atlantic Forest has led some studies to use models from different forest types, which we believe can lead to biased estimates. These estimates can also differ due to varying levels of conservation (Rolim et al. 2005), variation in stem density (Marchiori et al., 2016), and the presence of a larger *dbh* range (Marchiori et al., 2016). Regardless, we believe the Chave et al. (2014) approach overestimated biomass and carbon medians by over 49% and 68%, respectively, for the Atlantic Forest examined in our analysis, highlighting the significant potential for error when equations are applied to a different area than what they were developed for and/or based on generalized relations.

Table 5. Biomass estimates using two models from this study (LME and T.DW), and from literature for different sites of Atlantic Forest. All estimates were obtained only for trees. NA – Not available.

Study	Biomass (Mg ha ⁻¹)	dbh (cm)	Stems (ha ⁻¹)	Forest Type	Status of Conservation
LME	108.7	5.4-68.5	1,490	Moist	Secondary
T.DW	116.3	5.4-68.5	1,490	Moist	Secondary
Marchiori et al. (2016)	155.6	4.8-108.2	1,704	Montane Ombrophilous Dense	Secondary
Alves et al. (2010)	271.3	4.8-NA	1,230	Montane	NA
Rolim et al. (2005)	344.5	10-NA	NA	NA	Primary

When compared to the biomass estimation approaches, relatively small differences were observed for the contrasting approaches used to estimate carbon. The IPCC carbon factor for tropical and subtropical forest is 0.47, which lead to an overestimation of 4.4%. Since the IPCC carbon conversion factor is more general factor and likely built on fewer species, we recommended our carbon conversion factor of 0.45 (based on the mean of the 16 species examined) for the Atlantic Forest.

2.4.3. Biomass for less abundant species

Less abundant species can be a challenge for biomass estimation in the tropics and can comprise a significant portion of the total biomass. In our analysis, relatively few abundant species comprised the majority of the total biomass similar to other studies (e.g. van Breugel et al. 2011; Fauset et al. 2015). However, species-specific models may not always be the case and an effective approach for these less abundant species is needed. We believe the use of species-specific models for the most abundant species and a functional trait model used for the less abundant species is a rather feasible strategy for stand-level biomass and carbon estimation. Although a significant number of species can be rather challenging for identification in tropical forests, the use of genera

level information and the extensive database for *wsg* (Chave et al., 2009; Zanne et al. 2009) may provide important information for biomass and ultimately improve the estimation at the stand-level.

Using genus-level average *wsg* is a rather feasible strategy, especially for tropical forests where species-level identification is difficult or due to the lack of prior *wsg* information (Slik, 2006). Our results suggested similar estimates for the less abundant species using average *wsg* at either the species- or genus levels, reaffirming the appropriateness of this strategy. When a species- or genus-level information is not available, likely a family-level approach could be a viable alternative, but the usage of broad taxonomic groups such as families may cause an increased bias in determined biomass estimates. Actually, family-level information explained a limited amount (34 %) of *wsg* variance in a prior analysis (Chave et al. 2006), and is potentially not a reasonable strategy for the less abundant species.

Some studies (e.g. Bastin et al. 2015) had highlighted significant differences in *wsg* from various existing databases so we suggest the actual determination of *wsg* for the most abundant species, while using the prior information only for the less abundant species. In fact, observed specie-level *wsg* was the most important trait across species in this analysis and was quite important as predictor of biomass. For instance, the general overestimation of *C. speciosa* biomass was only observed when specie-level *wsg* was not included as a predictor. Thus, we believe predicting biomass for secondary sites with a high abundance of species with low *wsg* in the higher *dbh* classes can lead to an overall overestimation of stand-level biomass.

2.5. Conclusions

The LME species-specific model performed the best of all the approaches examined in this analysis and should be considered as a viable approach in future analyses. Generalized and

functional trait-based approaches also performed well, but only after the inclusion of *wsg* as a predictor, which highlights its importance in the highly diverse tropical forests similar to the one examined in this analysis. Neither of the two previously developed models examined performed well, indicating the challenge of applying biomass equations to different locations in the tropics. When compared to biomass estimation, much smaller differences (1-4%) were observed for the examined alternative approaches for estimating carbon from biomass. Despite the differences in carbon not being statistically different, we believe the use of a more conservative value of 0.45 rather than 0.47 carbon conversion factor from the IPCC is warranted, particularly when being applied at larger scales. Overall, the analysis highlights the challenges of estimating both biomass and carbon in the highly species-rich tropical forest, but provides a framework that may help to improve estimation in future studies. In particular, the continual development and evaluation of functional trait based approaches across a broader range of traits and species are suggested.

We recommend the following steps to predict the biomass at stand-level:

- Use the species-specific models fit by LME for most the abundant species.
- Use the functional plant trait model based on *dbh* and *wsg* for the least abundant species. A genus-level of information can be used when a species-level is not available.
- When no information of species or genus is available, we suggest using a local generalized model based on *dbh*.

References

- Alvares, C.A., J.L. Stape, P.C. Sentelhas, J.L. de Moraes Gonçalves, and G. Sparovek. 2013. "Köppen's climate classification map for Brazil." *Meteorologische Zeitschrift* 22(6): 711-728. doi: 10.1127/0941-2948/2013/0507
- Alves, L.F., S.A. Vieira, M.A. Scaranello, P.B. Camargo, F.A.M. Santos, C.A. Joly, and L.A. Martinelli. 2010. "Forest structure and live aboveground biomass variation along an elevational gradient of tropical Atlantic moist forest (Brazil)." *Forest Ecology Management* 260: 679-691. doi: 10.1016/j.foreco.2010.05.023
- Baitello, J.B., O.T. Aguiar, F.T. Rocha, J.A. Pastore, and R. Esteves. 1992. "Florística e fitossociologia do estrato arbóreo de um trecho da Serra da Cantareira (Núcleo Pinheirinho) - SP." *Revista do Instituto Florestal* 4, 291-297.
- Baker, T.R., et al. 2004. "Variation in wood density determines spatial patterns in Amazonian forest biomass." *Global Change Biology* 10: 545-562. doi: 10.1111/j.1365-2486.2004.00751.x .
- Bastin, J.F., A. Fayolle, Y. Tarelkin, J.V. den Bulcke, T. de Haulleville, F. Mortier, H. Beeckman, et al., 2015. "Wood specific gravity variations and biomass of central African tree species: the simple choice of the outer wood." *PLoS ONE* 10(11): 1-16. doi: 10.1371/journal.pone.0142146.
- Brown, S., and A.E. Lugo. 1984. "Biomass of tropical forests: A new estimate based on forest volumes." *Science* 223(4642): 1290-1293. doi: 10.1126/science.223.4642.1290.
- Brown, S., A.J.R. Gillespie, and A.E. Lugo. 1989. "Biomass estimation methods for tropical forest with applications to forest inventory data." *Forest Science* 35(4), 881-902.

- Burger, D.M., and W.B.C Delitti. 2008. "Allometric models for estimating the phytomass of a secondary Atlantic Forest area of southeastern Brazil." *Biota Neotropica* 8(4), 131-136. doi: 10.1590/S1676-06032008000400012.
- Cepagri. 2017. Centro De Pesquisas Meteorológicas E Climáticas Aplicadas À Agricultura. Available at: <<http://www.cepagri.unicamp.br/>>. Assessed at August 8 2017.
- Chambers, J.Q., J. Santos, R.J. Ribeiro, and N. Higuchi. 2001. "Tree damage, allometric relationships, and above-ground net primary production in central Amazon Forest." *Forest Ecology and Management* 152: 73-84.
- Chave, J., B. Riera, and M. Dubois. 2001. "Estimation of biomass in a Neotropical forest of French Guiana: spatial and temporal variability". *Journal of Tropical Ecology* 17: 79-96. doi: 10.1017/S0266467401001055.
- Chave, J., et al. 2005. "Tree allometry and improved estimation of carbon stocks and balance in tropical forests." *Oecologia* 145: 87-99. Doi: 10.1007/s00442-005-. 0100-x
- Chave, J., H.C. Muller -Landau, T.R. Baker, T.A. Easdale, H.T. Steege, and C.O. Webb. 2006. "Regional and phylzogenetic variation of wood density across 2456 neotropical tree species." *Ecological Applications* 16(6): 2356–2367. Doi; 10.1890/1051-0761(2006)016[2356:RAPVOW]2.0.CO;2
- Chave J., D.A. Coomes, S. Jansen, S.L. Lewis, N.G. Swenson, and A.E. Zanne. 2009. "Towards a worldwide wood economics spectrum." *Ecology Letters* 12(4): 351-366. <http://dx.doi.org/10.1111/j.1461-0248.2009.01285.x>
- Chave, J., et al., 2014. "Improved allometric models to estimate the aboveground biomass of tropical trees." *Global change biology* 20(10): 3177-3190. doi: 10.1111/gcb.12629.

- Crescente-Campo, F., J.J. Corral-Rivas, B. Vargas-Larreta, and C. Wehenkel. 2013. "Can random components explain differences in the height-diameter relationship in mixed uneven-aged stands?" *Annals of Forest Science* 71: 51-70. doi: 10.1007/s13595-013-542 0332-6.
- Fauset, S., M.O. Johnson, M. Gloor, T.R. Baker, A. Monteagudo, R.j.W. Brienen, T.R. Feldpausch, et al. 2015. "Hyperdominance in Amazonian forest carbon cycling." *Nature Communications* 6: 7857. doi: 10.5521/FORESTPLOTS.NET/2015_1 10.5521/FORESTPLOTS.NET/2015_1
- Huy, B., K.P. Poudel, K. Kralicek, N.D. Hung, P.V. Khoa, V.T. Phuong, V.T., and H. Temesgen. 2016. "Allometric equations for estimating the tree aboveground biomass in tropical Dipterocarp Forest of Vietnam." *Forests* 7(180): 1-19. doi: 10.3390/f7080180.
- IBGE. 2010. Instituto Brasileiro de Geografia - Diretoria de Geociências. Base vetorial contínua (BCIM). Rio de Janeiro. Escala: 1:1.000.000. Available at: <http://geoftp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bcim/>. Accessed at 19 October 2017.
- IPCC. 2006. Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change Guidelines for National Greenhouse Gas Inventories. Chapter 4: Forest Lands. Japan: National Greenhouse Gas Inventories Programme.
- Lam, T.Y., J.A. Kershaw, Z.S.N. Hajar, K.A. Rahman, A.R. Weiskittel, and M.D. Potts. 2016. "Evaluating and modelling genus and species variation in height-to-diameter relationships for Tropical Hill Forests in Peninsular Malaysia." *Forestry* 90(2): 268-278. doi: 10.1093/forestry/cpw051
- Larcher, L., M.R.T. Boeger, and M.C.M. Marques. 2012. "Biomass allocation and shade tolerance in tree species of the Atlantic Forest." *Botany* 90: 830-838. doi: 10.1139/b2012-053.
- Lima, A.J.N., R. Suwa, G.H.P. de Melo, T. Kajimoto, J dos Santos, R.P. da Silva, C.A.S. de Souza, et al. 2012. "Allometric models for estimating above- and below-ground biomass in Amazonian

forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil.” *Forest Ecology and Management* 277: 163–172. doi: 10.1016/j.foreco.2012.04.028.

Lima, R.A.F., G. Pitta, M.O. Melito, C. Bello, L.F. Magnago, V.P. Zwiener, D.D. Saraiva, et al., 2015. “How much do we know about the endangered Atlantic Forest? Reviewing nearly 70 years of information on tree community surveys” *Biodiversity and Conservation* 24(9): 2135–2148. doi : 10.1007/s10531-015-0953-1.

MacFarlane, W.D., and A.R. Weiskittel. 2016. “A new method for capturing stem taper variation for trees of diverse morphological types.” *Canadian Journal of Forest Research* 46: 804–815. doi: 10.1139/cjfr-2016-0018

Marra, D.M., N. Higuchi, S. E. Trumbore, G.H.P.M. Ribeiro, J. dos Santos, V.M.C. Carneiro, A.J.N. Lima et al. 2016. “Predicting biomass of hyperdiverse and structurally complex central Amazonian forests – a virtual approach using extensive field data.” *Biogeoscience* 13: 1553–1570. doi: 10.5194/bg-13-1553-2016.

Marchiori, N.M., H.R. Rocha, J.Y. Tamashiro, and P.M.A. Aidar. 2016. “Tree community composition and aboveground biomass in a secondary Atlantic Forest, Serra do Mar State Park, São Paulo, Brazil.” *Cerne* 22(4): 501-514. Doi: 10.1590/01047760201622042242.

MMA. 2012. Ministério do Meio Ambiente - Mapa da área de aplicação da Lei 11.428 de 2006. Escala 1:5.000.000. Available at: <http://www.mma.gov.br/estruturas/202/_arquivos/shape_mata_atlantica_ibge_5milhoes_policonica_sirgas2000shp_202.zip>. Accessed at: 19 October 2017.

Myers, N, R.A. Mittermeier, C.G. Mittermeier, G.A.B. Fonseca, and J. Kent, J. 2000. “Biodiversity hotspots for conservation priorities.” *Nature* 403: 853-858. doi: 10.1038/35002501.

- Nelson, B.W., R. Mesquita, J.L.G. Pereira, S.G.A. Souza, G.T. Batista, and L.B. Couto. 1999. "Allometric regressions for improved estimate of secondary forest biomass in central Amazon." *Forest Ecology and Management* 117:149-167. doi: 10.1016/S0378-1127(98)00475-7.
- Nogueira, E.M.N., P.M. Fearnside, B.W. Nelson, and M.B. França. 2007. "Wood density in forests of Brazil's 'arc of deforestation': Implications for biomass and flux of carbon from land-use change in Amazonia." *Forest Ecology and Management* 248: 119–135. doi: 10.1016/j.foreco.2007.04.047.
- Nogueira, E.M., B.W. Nelson, F.M. Fearnside, M.B. França, and A.C.A. Oliveira. 2008. "Tree height in Brazil's 'arc of deforestation': Shorter trees in south and southwest Amazonia imply lower biomass." *Forest Ecology and Management* 255: 2963–2972. doi: 10.1016/j.foreco.2008.02.002.
- Overman, J.P., H.J. Witte, and J.G. Saldarriaga. 1994. "Evaluation of regression models for aboveground biomass determination in Amazon rainforest." *Journal of Tropical Ecology* 10(2): 207-218. doi: 10.1017/S0266467400007859.
- Pan, Y., R.A. Birdsey, J. Fang, R. Houghton, P.E. Kauppi, W.A. Kurz, O.L. Phillips, et al. 2011. "A large and persistent carbon sink in the world's forests." *Science* 333: 988-993. doi: 10.1126/science.1201609.
- Pinheiro, J.C., and D.M. Bates. 2000. *Mixed-effects models in S and S-plus*. Springer.
- Powers, J.S., and P. Tiffin. 2010. "Plant functional type classifications in tropical dry forests in Costa Rica: leaf habit versus taxonomic approaches." *Functional Ecology* 24: 927-936. doi: 10.1111/j.1365-2435.2010.01701.x.
- Ribeiro, M.C., J.P. Metzger, A.C. Martensen, F.J. Ponzoni, and M.M. Hirota. 2009. "The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation." *Biological Conservation* 142: 1141–1153. doi: 10.1016/j.biocon.2009.02.021.

- Robinson, A.P., R.A. Duursma, and J.D. Marshall. 2005. "A regression-based equivalence test for model validation, shifting the burden of proof." *Tree Physiology* 25: 903–913. doi: 10.1093/treephys/25.7.903.
- Rolim, S.G., R.M. Jesus, H.E.M. Nascimento, H.T.Z. Couto, and J.C. Chambers. 2005. "Biomass change in an Atlantic tropical moist forest: the ENSO effect in permanent sample plots over a 22-year period." *Oecologia* 142: 238–246. doi: 10.1007/s00442-004-1717-x.
- Russell, M.B, A.R. Weiskittel, and J.A. Kershaw. 2014. "Comparing strategies for modeling individual-tree height and height-to-crown base increment in mixed-species Acadian forests of northeastern North America." *European Journal of Forest Research* 133: 1121- 1135. doi: 10.1007/s10342-014-0827-1.
- Scatena, F.N., W. Silver, T. Siccama, A. Johnson, and M.J. Sánchez. 1993. "Biomass and nutrient content of the Bisley Experimental Watersheds, Luquillo Experimental Forest, Puerto Rico, before and after Hurricane Hugo, 1989." *Biotropica* 25(1): 15-27. doi: 10.2307/2388975.
- Sharma, M., and J. Paton. 2007. "Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach." *Forest Ecology and Management* 249: 187- 198. doi: 10.1016/j.foreco.2007.05.006.
- Shimamoto, C.Y., P. Botosso, and M.C.M. Marques. 2014. "How much carbon is sequestered during the restoration of tropical forests? Estimates from tree species in the Brazilian Atlantic Forest." *Forest Ecology and Management* 329: 1-9. doi: 10.1016/j.foreco.2014.06.002.
- Slik, J.W.F. 2006. "Estimating species-specific wood density from the genus average in Indonesian trees." *Journal of Tropical Ecology* 22: 481-482. doi: 10.1017/S0266467406003324.
- Sotomayor, Jaime .F.M. 2013. "Determinação de biomassa aérea em florestas nativas num ambiente agrícola do Estado de São Paulo." Ms. diss., University of São Paulo.

- Soil Survey Staff. 2014. Keys to Soil Taxonomy. 12th Edition. Washington, DC: US Department of Agriculture–Natural Resources Conservation Service.
- SOS Mata Atlântica, Instituto Nacional de Pesquisas Espaciais. 2013. Atlas dos remanescentes florestais da Mata Atlântica, período de 2013 a 2014. Available at: <<http://www.sosmatatlantica.org.br>>. Accessed at: 19 October 2017.
- Sprugel, D.G. 1983. “Correcting for bias in log-transformed allometric equations.” *Ecology* 64(1): 209-210. doi: 10.2307/1937343.
- van Breugel, M., J. Ransijn, D. Craven, F. Bongers, and J.S. Hall. 2011. “Estimating carbon stock in secondary forests: Decisions and uncertainties associated with allometric biomass models.” *Forest Ecology and Management* 262: 1648-1657. doi: 10.1016/j.foreco.2011.07.018.
- Vismara, Edgar.S. 2013. “Avaliação da construção e aplicação de modelos florestais de efeito fixo e efeitos mistos sob o ponto de vista preditivo”. Ms.diss., University of São Paulo.
- Zanne A.E., G. Lopez-Gonzalez, D.A. Coomes, J. Ilic, S. Jansen, S.L. Lewis, R.B. Miller, N.G. Swenson, M.C. Wiemann, and J. Chave. 2009. Towards a worldwide wood economics spectrum. Dryad Digital Repository. <http://dx.doi.org/10.5061/dryad.234>
- Weiskittel, A.R., D.W. MacFarlane, P.J. Radtke, D.L. Affleck, H. Temesgen, C.W. Woodall, J.A. Westfall, and J.W. Coulston. 2015. “A call to improve methods for estimating tree biomass for regional and national assessments.” *Journal of Forestry* 113(4): 414-424. doi: 10.5849/jof.14-091.

3. ASSESSING VARIATION IN FOREST-LEVEL ABOVEGROUND BIOMASS AND CARBON ESTIMATES FOR A SPECIES RICH ATLANTIC FOREST IN BRAZIL: A CASE STUDY FOCUSED ON CANTAREIRA STATE PARK

ABSTRACT

High accuracy and precision of biomass and carbon estimates is important, but uncertainty of these values is often not reported and can lead to bias. The objective of this study was to compare different biomass and carbon estimates for the species-rich Cantareira State Park of the Atlantic Forest in Brazil. Several different models that varied in tree-level accuracy were tested using a network of plots, which were sampled using variable-radius methods. For comparison, both species-specific and generalized tree-level models were used, while the latter method included two local and two previously existing generalized models. The models were based on diameter at breast height (*dbh*) and wood specific gravity (*wsg*), which were found to have the highest accuracy when compared to alternative models. The estimates differed significantly, particularly for the two previously existing models, and differences were up to over 44% in some cases. Overall, our results highlight the sensitivity of tropical forest biomass/carbon estimates to the approach used and suggest the need to both validate and potentially calibrate regional models when no local models are available.

Keywords: Tropical forest; Species-specific models; Generalized models

3.1. Introduction

Significant efforts to quantify the tropical forest biomass and carbon have been conducted (Pan et al. 2011, Saatchi et al. 2011, Baccini et al. 2012). The estimation of carbon at large-scales often requires multiple steps ranging from field inventory sampling to remote sensing analysis, but this process has several sources of error, which may lead to different estimates. For example, the field inventory can include several uncertainties associated with tree-level measurements as well as the sampling intensity and design used, which can lead to biased forest-level carbon stock estimates (Clark & Kellner 2012). However, more limited assessments of forest-level biomass and

carbon uncertainties have been conducted in species-rich tropical forests where estimation is much more difficult.

In particular, individual tree biomass prediction is a significant challenge for tropical forests due to the complexities of model development and number of species present. For temperate forests, family- or genus-specific models often exist (Chojnacky et al. 2014) and generally have much better performance than generalized or species-group equations (Jenkins et al. 2003). A similar approach focused at the species-level has emerged as a reasonable alternative for tropical forests, but some studies suggest the species-specific approach only for the most abundant species due to the high diversity of species present (Nelson et al. 1999, Marra et al. 2016). The most abundant species are often responsible for majority of the biomass at large scales (Fauset et al. 2015). For the less abundant species, a generalized model based on diameter and woody specific gravity seems to be a satisfactory alternative (Colmanetti et al. 2018a, *In review*).

However, some tropical forest biomass and carbon quantification efforts are somewhat limited by biomass model availability. In this case, generalized models developed from an extensive database with a large number of species such as pan-tropical models (Chave et al. 2005, 2014) are a reasonable option. However, a central issue of using models outside their area of development is the accuracy of their predictions. If no validation data is available, the assessment of accuracy is largely speculative. The Atlantic Forest is highly fragmented (Ribeiro et al. 2009) but is also a high protected forest (Brazilian Federal Law of Atlantic Forest Protection N. 11,428/2006) where destructive sampling to fit a new model or calibrate a previous one is highly restricted.

Many studies used models from other regions (e.g., Rolim et al. 2005, Alves et al. 2010, Shimamoto et al. 2014, Marchiori et al. 2016), since just a few models have been proposed for this

biome. Only one study in the recent past (Burger & Delitti 2008), and a more recent one (Colmanetti et al. 2018a, *In review*) have developed biomass equations for this region. Both highlight significant positive and negative bias for the Burger & Delitti (2008) and Chave et al. (2014) models, respectively, was found by Colmanetti et al. (2018a, *In review*) at the stand-level. To evaluate the accuracy of previously developed models at broad spatial scale can aid the verification of biomass and carbon stocks, which is particularly important for understanding the environmental services for a biome not commonly studied for biomass, such as the Atlantic Forest in Brazil.

To investigate the relevance of different models for individual tree biomass prediction and their influence of total biomass/carbon forest-level estimates, we proposed examination of Cantareira State Park. The park size is approximately 8,000 ha having a significant green area, located in the vital watershed and making part of the Cantareira System, a critical water provider for the city of São Paulo, the most prominent metropolis in South America, with over 20 million of citizens in the metropolitan area (IBGE, 2017). The park also makes part of the Biosphere Reserve of the Green Belt ("Cinturão Verde," in Portuguese) of the city of São Paulo elected by UNESCO and it is an essential green area in the São Paulo metropolitan area, but is under the intense urbanization pressure.

Given the importance of the individual tree biomass prediction on Atlantic Forest biomass estimates, the goal of this study was to assess the influence of different approaches (generalized and species-specific) and previously developed models on forest-level biomass estimation and the implications for carbon quantification for Atlantic Moist Dense Forests. The specific research objectives were to: (i) to estimate the biomass of the park using species-specific, local generalized models, and existing models from literature to predict individual tree biomass; (ii) evaluate the

precision of those estimates as well as the sampling efficiency; and (iii) compare the estimates and examine the implications of carbon estimates at a forest-level.

3.2. Materials and methods

3.2.1. Study site

The studied site was the Cantareira State Park (Parque Estadual da Cantareira, in Portuguese), that was a mosaic of secondary and primary Atlantic Forest located north of São Paulo City - Brazil. The vegetation of the region is classified as a moist forest with some species of the seasonal semi-deciduous forest with a general predominance of species from the Euphorbiaceae, Lauraceae, Myrtaceae, Meliaceae, Sapotaceae and Rubiaceae families (Baitello et al. 1992). In this study, a general predominance of species was from the families of Euphorbiaceae, Fabaceae, Sapindaceae, and Lauraceae (Appendix C). The climate in the region according to Köppen' climate classification adapted to Brazil is Cwa (Alvares et al. 2013), an average annual temperature of 20.7 °C (minimum average: 15.7°C, maximum average: 25.7°C) and an annual rainfall of 1,400 mm (Cepagri 2017). The topography is characterized by steep slopes, with altitude varying from 775 to 1,215 m, and the soil classification according to soil taxonomy (Soil Survey Staff, 2014) was Ultisol Udic Udults Typic Hapludults.

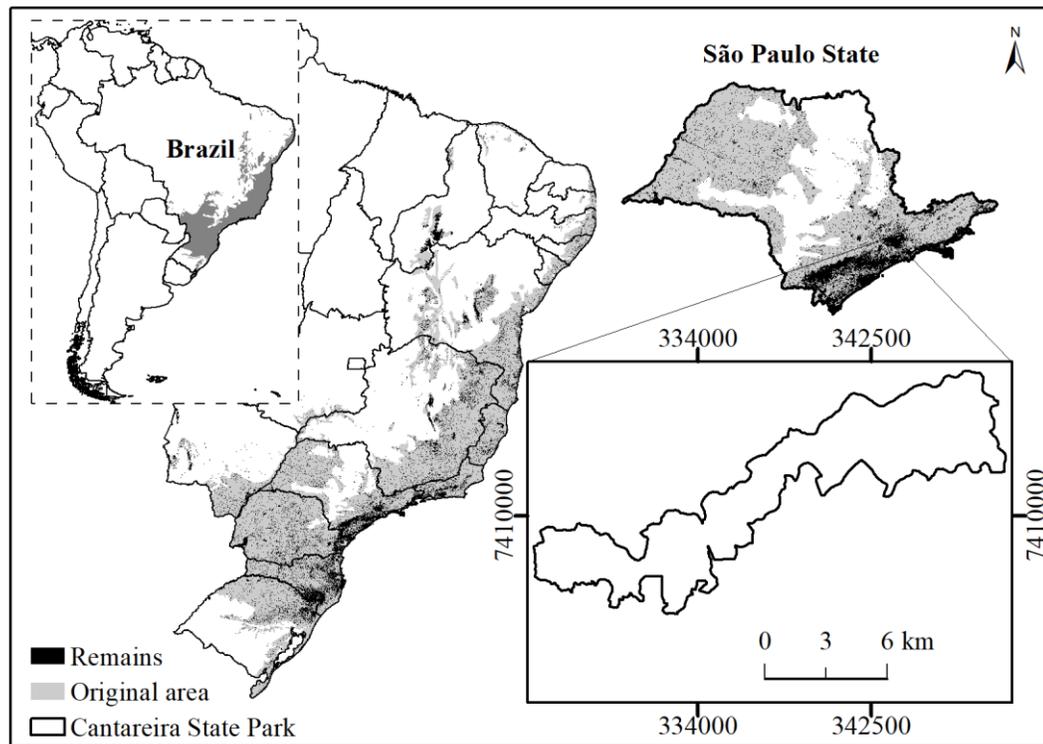


Figure 1. Study site in Cantareira State Park, Brazil (sources: SOS Mata Atlântica 2013, MMA 2012 IBGE, 2010). Coordinates in UTM – WGS84.

3.2.2. Data collection

Double sampling using the Bitterlich method, with a basal area factor equal to three ($F_G = 3$) for the Relascope was used. A systematic location of points using east and west orientation with points spread along the major trail in the park was used. In the first-phase of the double sampling, the auxiliary variable, in this case, the basal area was obtained by using the Relascope to count the number of trees. For the second-phase, the *dbh* of all trees in the stand were measured, and the individual tree aboveground biomass was predicted using different models. The stand aboveground biomass (variable of interest) was determined by summing the individual tree biomass. This phase was the subset of the first-phase, and the trees were identified *in situ* at the species-level. When the identification at that level was difficult, the trees were identified at genus- or family-level. The

sample size was 48 points in the first-phase and 18 in the second-phase (27 % of the total of points). All the sampling techniques used were according to Shiver & Borders (1996) and Batista et al. (2016).

The points were all spread along several transects with four to five points each, except by two transect with only two points. The localization points in the middle of the main trail were taken as reference for transect in strata S2 and S3. For stratum S1, one localization point on the beginning of the strata was used as the reference. A distance (d) of 50 m was defined for the Bitterlich points and was based on Eq.1, where the information of the expected maximum tree diameter was from Baitello et al. (1992).

$$(1) \quad d = \frac{\max(dbh)}{\sqrt{F_G}}$$

where $\max(dbh)$ is the expected maximum tree diameter to be sampled and F_G is the basal area factor ($F_G = 3$).

The stratified sampling was used, and the information for that was obtained from management plan of the park (Management Plan 2009), where the vegetation was characterized based on satellite images. The management plan proposed for the whole park area of 7,916.52 ha was across four administration centers (Águas Claras, Cabuçu, Engordador and Pedra Grande). The dominant forest type of the park is Moist Dense Forest, having the following subdivision, according to IBGE (1992): Moist Dense Montane Forest, Alluvial Moist Dense Montane Forest, Small Size Moist Dense Montane Forest, Shrubland and Secondary Areas (or Anthropic Areas). The present study focused on Moist Dense Montane Forest since it covers the most of the total area (92.5 % of the total area). The Moist Dense Stands Montane Forest subdivision was stratified according to its status of conservation as (i) vegetation with high tree size, uniform canopy and low or without alteration (4.5 % of the total area); (ii) vegetation with high tree size, non-uniform

canopy and low alteration (23 % of the total area), (iii) vegetation with medium tree size, non-uniform canopy and strong alteration (10 % of the total area); (iv) secondary vegetation with medium tree size, varying the canopy and strong alteration (22.5 % of the total area); (v) secondary vegetation with a high density of trees and low canopy (32.5 % of the total area). In this study, only the strata ii, iii and iv (named as S1, S2, and S3, respectively) were sampled, accounting for 55.5 % of the total park area (Fig. 2). We highlight the steep slope associated with remote sites imposed a strong limitation to locomotion inside the forest avoiding to take samples in all strata.

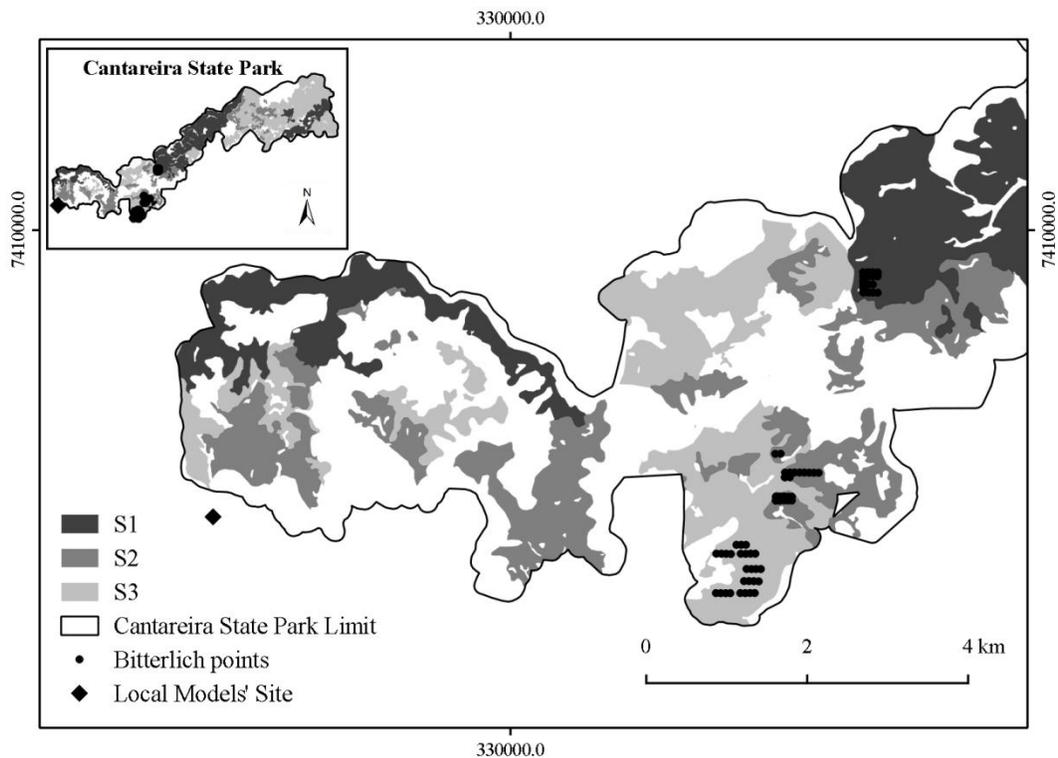


Figure 2. Cantareira State Park, Brazil. S1, S2, and S3 are the strata sampled; Local Models' Site is the localization of the destructive sampling used to fit the local generalized and species-specific models in Colmanetti et al. (2018a, *In review*). Coordinates in UTM – WGS84.

3.2.3. Data analysis

The following equations were all obtained from (Batista et al. 2016). The auxiliary variable (G'_j ; stand basal area) in the first-phase was calculated for each stand according to Bitterlich (1984), and obtained:

$$(2) \quad G'_j = N_j \cdot F_G$$

where N_j is the number of trees of the stand j ; and F_G is expansion factor ($F=3$).

The second-phase was dependent on the first-phase and considered as a subset. For this phase, besides the auxiliary variable, the variable of interest, which was the stand-level estimate of biomass, were taken. The stand-level biomass (B_j):

$$(3) \quad B_j = \sum_{i=1}^{N_j} b_{ij} \cdot f_{Dij}$$

Where b_{ij} is the individual tree biomass for the tree i in stand j , and f_{Dij} is the individual tree expansion factor for the tree i in stand j . The biomass was predicted by using several different models and approaches, as described below.

The individual tree expansion factor (f_{Dij}) was obtained as:

$$(4) \quad f_{Dij} = \frac{F_G}{g_{ij}}$$

where: g_{ij} is the transversal area for each tree i in the stand j , and the F_G is the basal area factor equal three.

In the second-phase, a ratio estimator (\hat{R} eq.5) was obtained by the relationship between biomass (B_j) and basal area (G_j), and used to obtain the biomass in the first-phase.

$$(5) \quad \hat{R} = \frac{B_j}{G_j}$$

where the variables were defined previously.

The biomass from the first-phase was obtained by the relationship between the stand basal area and the ratio estimator:

$$(6) \quad B_j = G'_j \cdot \hat{R}$$

3.2.3.1. Variance of mean ($\widehat{Var}(\hat{\mu}_B)$)

The variance of the mean was obtained using eq.7:

$$(7) \quad \widehat{Var}(\hat{\mu}_B) = \frac{\hat{\sigma}_Y^2 + \hat{R}^2 \cdot \hat{\sigma}_{X_1}^2 - 2 \cdot \hat{R} \cdot \hat{\sigma}_{XY}}{n_2} + \frac{2 \cdot \hat{R} \cdot \hat{\sigma}_{XY} - \hat{R}^2 \cdot \hat{\sigma}_{X_1}^2}{n_1} + \frac{\hat{\sigma}_Y^2}{N}$$

where: $\hat{\sigma}_Y^2$ is the variance of biomass in the stands in the second-phase; \hat{R} is the ratio estimator; $\hat{\sigma}_{X_1}^2$ is the variance of basal area in first-phase; $\hat{\sigma}_{XY}$ is the covariance of basal area and biomass in the second-phase; n_1 is the number of points in the first-phase; n_2 is the number of points in the second-phase; N for the Bitterlich method is the number of points in the forest, in this case, is infinity.

3.2.3.2. Overall mean ($\hat{\mu}_B$)

The total biomass was obtained by summing the biomass by each stratum:

$$(8) \quad \hat{\mu}_B = \hat{R} \cdot \hat{\mu}_{X_1}$$

where $\hat{\mu}_{X_1}$ is the mean of the basal area in the first-phase.

3.2.3.3. Total biomass (\hat{t}_B)

The total biomass was obtained by summing the biomass by each stratum:

$$(9) \quad \hat{t}_B = N \cdot \hat{\mu}_B$$

where $\hat{\mu}_B$ is the mean biomass; N is the total area in the hectare (ha).

3.2.3.4. Variance of the total ($\widehat{Var}(\hat{\tau}_B)$)

The variance of the total was obtained by:

$$(10) \quad \widehat{Var}(\hat{\tau}_B) = N^2 \cdot \widehat{Var}(\hat{\mu}_B)$$

where the variables were defined previously.

3.2.3.5. Estimates of sampling error ($E\%$)

The relative sampling error was:

$$(11) \quad E\% = \frac{t_{0.05} \cdot \sqrt{\widehat{Var}(\hat{\mu}_B)}}{\hat{\mu}_B} \cdot 100$$

where the $t_{0.05}$ value was approximately two and the other variables were defined previously.

3.2.3.6. Stratified sampling

For the stratified sampling, all the procedures above described were calculated for each stratum, separately. The biomass of the total for stratified sampling ($\hat{\tau}_{BSS}$) was obtained by summing the total of each stratum (S1, S2, and S3):

$$(12) \quad \hat{\tau}_{BSS} = \sum(\hat{\tau}_{S1} + \hat{\tau}_{S2} + \hat{\tau}_{S3})$$

The variance of the total for the stratified sampling ($\widehat{Var}(\hat{\tau}_{BSS})$) was obtained by summing the variance for the different strata:

$$(13) \quad \widehat{\mathbf{Var}}(\hat{t}_{BSS}) = \sum \left(\widehat{\mathbf{Var}}(\hat{t}_{S1}) + \widehat{\mathbf{Var}}(\hat{t}_{S2}) + \widehat{\mathbf{Var}}(\hat{t}_{S3}) \right)$$

The variance of the mean for the stratified sampling ($\widehat{\mathbf{Var}}(\hat{\mu}_{BSS})$) was obtained by:

$$(14) \quad \widehat{\mathbf{Var}}(\hat{\mu}_{BSS}) = \frac{\widehat{\mathbf{Var}}(\hat{t}_B)}{N^2}$$

3.2.3.7. Individual tree biomass prediction

The individual tree biomass (b_{ij} ; eq.3) was predicted using different models. Three approaches and five models for each tree were used to predict the biomass, which provided several alternative estimates of the total biomass for the park. They are:

3.2.3.7.1. Species-specific approach:

The species-specific biomass models fitted by Colmanetti et al. (2018a, *In review*) were used. For this approach, species as a random effect for the 16 abundant species was used. In our study, we used the species-specific models proposed by Colmanetti et al. (2018a, *In review*), and for species with no specific models, we used the generalized model with *dbh* and *wsg* also proposed by the same authors (eqs. 15 and 16). We used the average of the *wsg* for genus for those species with no species-level information, and the generalized model using *dbh* only for those species with no species- or genus-level of information.

3.2.3.7.2. Local generalized models approach:

The two local generalized biomass models fitted by Colmanetti et al. (2018a, *In review*), which occurred in a study surrounding the park for the same vegetation type, were used. The two models including *dbh* only and *dbh* and *wsg*, were used:

$$(15) \quad agb = \exp(-2.245 + 2.388 \cdot \ln(dbh))$$

$$(16) \quad agb = \exp(-1.293 + 2.389 \cdot \ln(dbh) + 1.373 \cdot \ln(wsg))$$

3.2.3.7.3. Existing biomass models approach:

Two existing models from the literature were used to predict the biomass. One of them was a model developed for the Atlantic Forest (Burger & Delitti 2008) based only on *dbh*:

$$(17) \quad abg = \exp(-3.068 + 2.522 \cdot \ln(dbh))$$

The other was a Pantropical model developed from a global database (Chave et al. 2014). The model (eq.18) was based on *dbh* and height and *wsg*. The height was predicted from a Pantropical hypsometric relationship (eq.19), and the wood specific gravity information was obtained from the Global Wood Density database (Chave et al. 2009, Zanne et al. 2009), using the values only for species occurring in South America. For these predictions, a genus-level average was used when species-level information was not available, while an average for the family was used when no information of both taxon levels information was available. For species with no *wsg* available for any taxon level, an average for all species was used for these cases. This strategy was required since Chave et al. (2014) did not provide any model without *wsg*.

$$(18) \quad agb = 0.0559(wsg \cdot dbh^2 \cdot h)$$

$$(19) \quad \ln(h) = 0.893 - E + 0.760 \cdot \ln(dbh) - 0.034 \cdot \ln(dbh)^2$$

where E is bioclimatic variable taken from a gridded global layer at a 2.5 arc second resolution (Tropical Forest Biodiversity Group).

3.2.3.8. Forest-level biomass and carbon estimates

We calculated the individual tree carbon by multiplying the predicted biomass by the carbon content. Two alternatives were used a single value of 0.45 obtained from an average of all sampled species according to Colmanetti et al. (2018a, *In review*), and a carbon content fraction of 0.47 as suggested for tropical and subtropical forests (IPCC 2006). For biomass, the various estimates were compared using analysis of variance (ANOVA), while a two-way ANOVA was used with the different carbon conversion factors with a factor for carbon estimates. The differences among the methods were evaluated using Tukey's HSD comparison test at a 95% level by using the *HSD.test* function from the package *agricolae* in R (R Development Core Team 2016).

3.3. Results

3.3.1. Simple and stratified sampling performances

A higher basal area was observed in second-phase with a greater standard deviation in strata S1 and S2 as well as overall.

Table 1. Summary of sampling of Atlantic Forest at Cantareira State Park.

Stratum	# Species	# Genera	# Families	First-phase			Second-phase			
				Basal area (m ² ha ⁻¹)	# Points	# Trees	Basal Area (m ² ha ⁻¹)	Average wsg	# Points	# Trees
S1	29	28	22	36.72±9.25	12	12±3	41.00±16.84	0.48±0.03	5	13.4±5.5
S2	26	24	21	37.48±10.61	16	12.2±3.5	41.82±12.34	0.50±0.04	6	13.7±4
S3	33	27	20	37.79±11.75	20	12.4±3.8	42.84±10.3	0.54±0.05	7	14±3.4
Overall	59	53	34	37.42±10.58	48	12.2±3.5	47.99±12.23	0.51±0.05	18	13.7±4

The *dbh* range varied according to each stratum: 5.0 – 102.5 cm for S1, 5.5 – 91.0 cm for S2, and 5.6 – 66.2 for S3 (Fig. 3). The two less altered strata, S1 and S2, had a predominance of large trees (*dbh* > 60 cm).

A general lower sampling error was observed for non-stratified sampling (Table 2). High correlation values were also observed for all models for the basal area (m² ha⁻¹) and biomass (Mg ha⁻¹) for the second-phase Bitterlich points. The models from Burger & Delitti (2008) and Chave et al. (2014) had lower and intermediate sampling error, respectively.

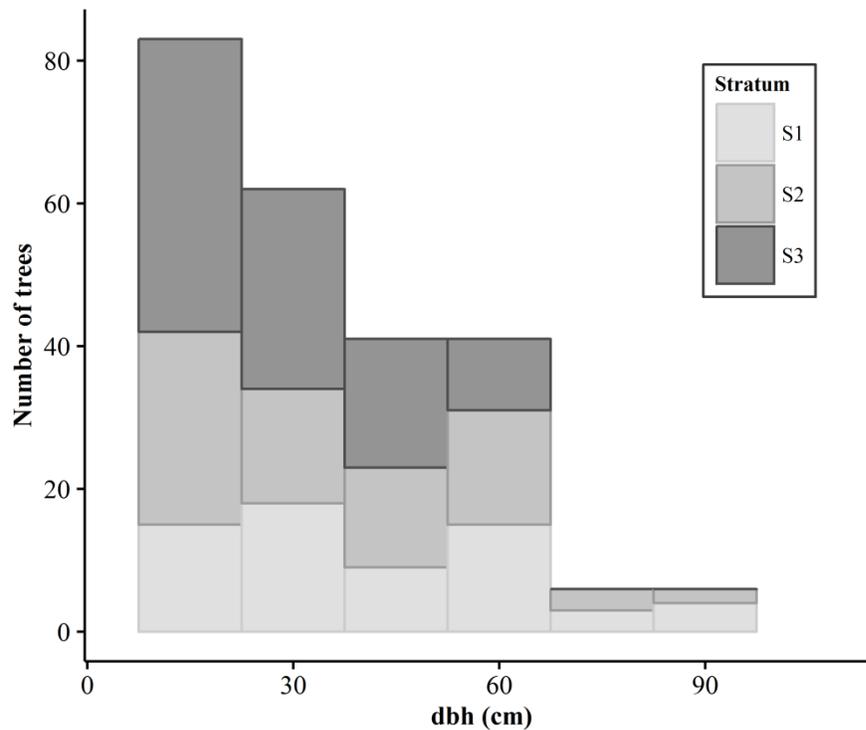


Figure 3. Diameter distribution in the second-phase sampling of Atlantic Forest at Cantareira State Park. S1, S2, and S3 are the studied strata.

Table 2. Observed percent sampling error for different sampling methods on the estimation of the biomass of Cantareira State Park using contrasting approaches to predict individual tree biomass. Pearson correlation for the basal area (m^2) and biomass ($kg\ ha^{-1}$) in the second-phase sampling. Equations included Burger & Delitti (B&D; 2008), Chave et al. (2014), generalized based on *dbh*, generalized with *dbh* and wood specific gravity, and a species-specific as a function of *dbh*.

Model	Sampling Error (%) for Systematic Sampling	Sampling Error (%) for Stratified Systematic Sampling	Pearson Correlation (Second-Phase)
B&D	8.9	8.5	0.95**
Chave	10.7	11.3	0.94**
<i>dbh</i>	8.3	8.2	0.97**
<i>dbh</i> and <i>wsg</i>	11.1	12.3	0.96**
Species-specific	11.0	12.2	0.96**
Overall Mean	10.0	10.5	

** Statistically significance of Pearson correlation at 95% level of confidence.

3.3.2. Stand-level estimates of biomass and carbon

Similar stand-level estimates of biomass were observed for local models (Fig. 4), $190.6 \pm 53.9 \text{ Mg ha}^{-1}$ (mean \pm SD), $202.1 \pm 57.1 \text{ Mg ha}^{-1}$, and $203.8 \pm 57.6 \text{ Mg ha}^{-1}$, respectively for the generalized models using *dbh* only, species-specific, and generalized model using *dbh* and *wsg*, respectively, based on systematic sampling with no stratification. Burger & Delitti (2008) and Chave et al. (2014), had the smallest and highest estimates, $134.4 \pm 38.0 \text{ Mg ha}^{-1}$ and $279 \pm 78.9 \text{ Mg ha}^{-1}$, respectively.

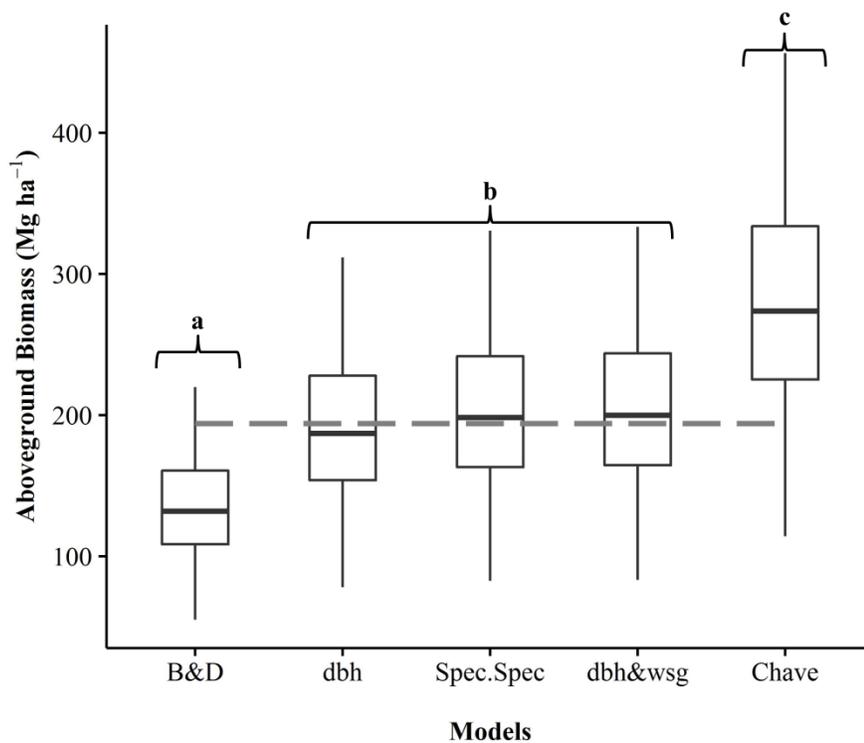


Figure 4. Stand-level estimates of biomass (Mg ha^{-1}) using different models including B&D – existing Atlantic Forest biomass model from Burger & Delitti (2008); *dbh* – local generalized model (only *dbh*) from Colmanetti et al. (2018a, *In review*); *Spec.Spec* – Species-Specific approach from Colmanetti et al. (2018a, *In review*); *dbh&wsg* – local generalized model (*dbh* and *wsg*) from based Colmanetti et al. (2018a, *In review*); Chave – existing Pantropical biomass model from Chave et al. (2014). The broken gray line represents a median of all biomass estimates ($193.93 \text{ Mg ha}^{-1}$). The letters are the statistical groups according to Tukey’s HSD comparison test at 95% level of confidence.

The estimates of carbon also were similar for local models (Fig. 5), $85.8 \pm 24.2 \text{ Mg ha}^{-1}$ (Mean \pm SD), $90.9 \pm 25.7 \text{ Mg ha}^{-1}$ and $91.7 \pm 25.9 \text{ Mg ha}^{-1}$ using 0.45 carbon conversion factor, respectively for generalized model using *dbh* only, species-specific, and generalized model using *dbh* and *wsg*, respectively, based on non-stratified sampling. Burger & Delitti (2008) and Chave et al. (2014), had the smallest and highest estimates, $60.5 \pm 17.1 \text{ Mg ha}^{-1}$ and $125 \pm 35.5 \text{ Mg ha}^{-1}$, respectively. No significant difference was observed for two carbon conversion factors.

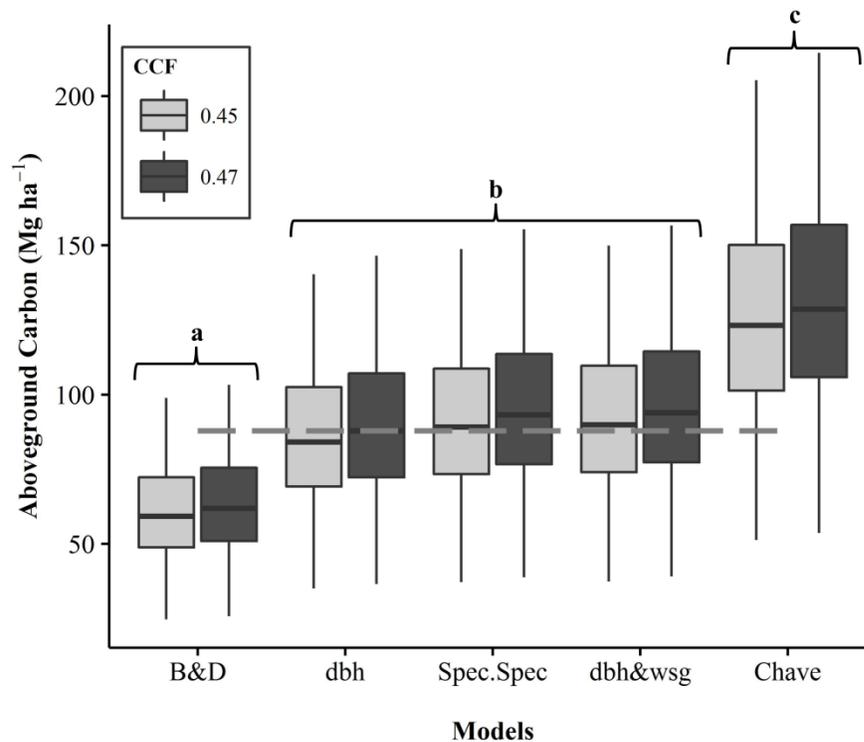


Figure 5. Stand-level estimates of carbon (Mg ha^{-1}) using different models including B&D – existing Atlantic Forest biomass model from Burger & Delitti (2008); dbh – local generalized model (only *dbh*) from Colmanetti et al. (2018a, *In review*); Spec.Spec – Species-Specific approach from Colmanetti et al. (2018a, *In review*); dbh&wsg – local generalized model (*dbh* and *wsg*) from based Colmanetti et al. (2018a, *In review*); Chave – existing Pantropical biomass model from Chave et al. (2014). CCF – Carbon Content Fraction of 0.45 based on Colmanetti et al. (2018a, *In review*); and 0.47 based on IPCC (2006). The broken gray line represents a median of all biomass estimates (87.9 Mg ha^{-1}). The letters are the statistical groups according to Tukey's HSD comparison test at 95% level of confidence.

3.3.3. Forest-level estimates of biomass and carbon

The total biomass varied widely across the models used to predict the individual tree biomass in the second-phase using non-stratified sampling (Table 3). As observed, the existing models had the extremes estimates, where Burger & Delitti (2008) and Chave et al. (2014) had the lower and higher estimates, respectively. The local generalized and species-specific models had more similar estimates.

The Burger and Delitti (2008) was 34% lower and Chave et al. (2014) was 38% higher than the biomass estimates from the species-specific approach. Using the existing models and the IPCC carbon conversion factor (CCF = 0.47), Burger and Delitti (2008) was likewise 31% (or $122.1 \cdot 10^3$ Mg less) lower and Chave et al. (2014) was 44% (or $176.57 \cdot 10^3$ Mg more) higher than the carbon estimates of the Species-specific approach, which used a CCF of 0.45.

Table 3. Estimates of total aboveground biomass and carbon (Mg \pm SD) of the Cantareira State Park by varying the models used to predict the individual tree biomass, and using non-stratified sampling. CCF is the carbon conversion factor, which was 0.45 according to Colmanetti et al. (2018a, *In review*) and 0.47 according to IPCC (2006). Equations included Burger & Delitti (B&D; 2008), Chave et al. (2014), generalized based on *dbh*, generalized with *dbh* and wood specific gravity, and a species-specific as a function of *dbh*.

Model	Total Biomass (10^6 Mg)	Total Carbon (10^3 Mg) C.C.F.= 0.45	Total Carbon (10^3 Mg) C.C.F. = 0.47	Sampling Error (%)
B&D	0.59 ± 0.03	265.65 ± 33.56	277.46 ± 33.10	8.9
Chave	1.23 ± 0.07	551.61 ± 87.96	576.13 ± 86.99	10.7
<i>dbh</i> only	0.84 ± 0.03	376.81 ± 50.10	396.56 ± 49.41	8.3
<i>dbh</i> and <i>wsg</i>	0.9 ± 0.05	402.82 ± 67.51	420.72 ± 66.77	11.1
Species-specific	0.89 ± 0.05	399.56 ± 66.25	417.32 ± 65.52	11.0

3.4. Discussion

3.4.1. Sampling efficiency

Bitterlich method (Bitterlich 1984) is a practical approach to basal area estimation and provides estimates with low error sampling. The method has been broadly used along the last century for temperate forests and recently for tropical forests (e.g., Nascimento et al. 2004, Bryan et al. 2010, Yang et al. 2017). The method requires adequate visualization throughout the understory that can be a challenge in a tropical forest where the understory is more abundant than in a temperate forest. However, some primary and advanced secondary forest can have a lower abundance of small trees in the understory (Guariguata & Ostertag 2001) and allow the method to be used. In this study, the less degraded secondary sites allowed sampling using the method and was very efficient (S1 and S2), while it was not particularly challenging for more degraded sites such as S3.

In general, the Bitterlich method is less time consuming and less laborious. Also, the relascope is very useful for accounting for slope since it does not require plot installation reducing the costs. Additionally, the double sampling method as used in this study can enhance sampling precision. According to Shiver & Borders (1996), double sampling only 25 – 35% of the total of points is adequate for precise estimates of both the volume and basal area, which we confirmed was also the case for biomass in our analysis. In our study, the correlation between basal area and biomass in the second-phase was quite high, supporting the idea that reducing the number of double-sampling to 27 % of the total of points within the range originally suggested by Shiver and Borders (1996).

Our results also suggested the estimates using the non-stratified sampling were more effective than stratified, especially for the species-specific biomass estimation approach. Although

the stratification is essential in large scales (Guitet et al. 2015), where different biomes (Scolforo et al. 2016), forest types (Alves et al. 2010, Djuikouo et al. 2010), or varying the successional stages (Nyirambangutse et al. 2017) play an important role on biomass and carbon quantification, we had no better performance for smaller scale like the park area studied. We speculate that using only satellite images with no field data collection may is not an effective method for stratification, and consequently no differences for aboveground biomass could be observed across the strata previously proposed.

Even though satisfactory estimates of both forest-level biomass and carbon were shown in this study, we had important limitations that should be recognized. The *dbh* range from datasets of the destructively sampled trees for both studies (5.4 – 68.5 cm for Colmanetti et al. 2018a; 1.6 – 47.8 cm for Burger & Delitti 2008) had much narrower *dbh* range in comparison to the *dbh* range in this study (Figure 2). For example, models from Colmanetti et al. (2018a, *In review*) predicted 23% of the total basal area (5% of the total of trees) by extrapolation, while model from Burger & Delitti (2008) predicted over 66% (27% of the total of trees). Another issue was the real location of points inside the forest. The strong steep slopes limited GPS signals and led us to use a more robust method of localization based on compass and measuring tape. So the distance of 50 m from each Bitterlich point was based on the real distance on the ground surface and not based on the maps.

3.4.2. Biomass models and the accuracy of predictions

Pantropical models have been extensively used for tropical forests (e.g., Alves et al. 2010; Shimamoto et al. 2014; Marchiori et al. 2016). The good fit of these models (Chave et al. 2005, 2014) can be mathematically supported by the huge number of trees and the large *dbh* range, and

consequently a reduction of the variance of the coefficients. We believe this reasonable fit associated with the general lack of biomass models has contributed their wide use, especially for Atlantic Forest. However, the unrestricted use of these models for Atlantic Forest is an important limitation. In general, the use of models fitted with a given dataset and then used elsewhere goes beyond the acceptable limits of the prediction and is likely extrapolation. In this context, any previous model requires local calibration. Otherwise, the accuracy of the estimates may be dubious. Even though no significant bias was observed for Brazilian Amazon Forest (Nogueira et al. 2008), we have clear evidence of positive bias by the generalized Pantropical models for Atlantic Forest at both the tree- and stand-levels (Colmanetti et al. 2018a, *In review*), and now in forest-level as suggested in this study.

The low accuracy is often associated with errors in the initial stages of inventory prediction that are simplified in the following steps in this analysis: (i) prediction of the individual tree biomass (kg tree^{-1}); (ii) summing the individual biomass values per plot, followed by a standardizing for the land area covered by plot leading to stand-level estimates ($\pm \text{Mg ha}^{-1}$); and (iii) forest-level estimate (or total estimate) based stand-level estimates ($\pm \text{Mg ha}^{-1}$). The precision of the forest-level is essentially the performance of the initial biomass estimate based on the variance (σ^2). Although it is possible to obtain high precision (low variance and low sampling error) as obtained by the model of Burger & Delitti (2008) and intermediate values for Chave et al. (2014), the overall accuracy is unknown if further verification is not performed. In the verification process, the accuracy was obtained by comparing the predicted and the observed biomass of each tree using destructive sampling.

Based on Colmanetti et al. (2018a, *In review*), these existing models from the literature have low accuracy for individual tree-level prediction. In our study, we used the same models proposed by Colmanetti et al. (2018a, *In review*), although no validation destructive sample was

taken, due to the legal protection of the park, we are assuming the best performance for those models on individual predictions in tree-level for the trees in the park once both study sites have the same forest type, and similar species composition.

The two models that used only *dbh* as a predictor had the smallest sampling error. However, they did not have the same high accuracy for individual tree biomass prediction for species in Atlantic Forest according to Colmanetti et al. (2018a, *In review*). In contrast, the species-specific and generalized models based on *dbh* and *wsg* had very similar estimates, but much higher accuracy. We believe the higher *wsg* for some species resulted in the higher estimate. For instance, *Cupania oblongifolia* have higher *wsg* than *Cecropia hololeuca*, and the predictions of their biomass are much different in the model using the *dbh* and *wsg* when compared to the *dbh* only model. Using *wsg* is an important predictor for generalized models (Bastin et al. 2015), and can lead to different estimates in a biomass estimate at a forest-level. This can often be a reasonable choice when a species-specific model is not available.

3.4.3. Implications for biomass and carbon estimates

The Atlantic Forest is highly degraded and fragmented forest (Ribeiro et al. 2009), so any information about the carbon stock in the original forest or the CO₂ release by deforestation in the past (Baccini et al. 2012) is merely speculative. In contrast, quantifying the current carbon remains is an important subject, but it is still quite challenging. For example, evaluation of the potential impacts of tropical forest carbon on global climate requires accurate carbon quantification (Clark & Kellner 2012). The lack of accuracy on the lower levels of the inventory (I.e. tree- and plot-level estimates) can lead to highly contrasting estimates at forest-level. Consequently, selection more accurate models such species-specific or local generalized models is crucial for the overall

estimation of carbon. Uncertainties of accuracy lead us to evaluate quantitatively the importance of alternative approaches for Atlantic Forest biomass, which has important implications for various ecosystem services such carbon stock (e.g., Vieira et al. 2011), carbon sequestering by reforestation (e.g., Shimamoto et al. 2014), or the release of carbon by degradation/disturbance (e.g. Lindner and Sattler 2012), and consequently relevance for global carbon cycle.

Aiming to quantify the biomass and carbon at large scales as the national-level, Gibbs et al. (2007) assumed the Pan-tropical models such as Chave et al. (2005) were ‘the best available’ method to estimate forest biomass and can be used accurately to estimate forest carbon stocks across a wide range of forest types, but occasionally not for all of them. Chave et al. (2014) enhanced the model proposed by Chave et al. (2005) and also included the destructive sampling dataset from Atlantic Forest. However, our results do not corroborate to Gibbs et al. (2007) assumption on Pantropical model performance. The positive bias in tree-level biomass observed by Colmanetti et al. (2018a, *In review*) had rather severe consequences for the forest-level estimate.

Biomass and carbon stocks still remain a significant source of uncertainty. Improvement on biomass estimates can support the understanding about the importance of the forest on the global carbon cycle. The current limitation of assessment on biomass and carbon quantification can be overcome, but some steps on the inventories must be improved, including the validation and consequently calibration of the models. A previously study highlighted the potential of calibrating individual tree biomass for specie-specific models (Colmanetti et al. 2018b, *In review*), and it should be considered for regional models in tropical forests. We strongly believe some efforts in this way can help to obtain more accurate estimates and from there start to make mitigation incentives such REDD+ a reality.

3.5. Conclusions

A previous study carried by Colmanetti et al. (2018a, *In review*) highlighted the higher accuracy for the species-specific and local generalized model based on *dbh* and *wsg*, but our results show a lower precision for these models when applied at larger spatial scales. The total biomass of the studied strata in the park was 0.89 Mg 10⁶ and carbon was 399.56 Mg³ for species-specific approach varying widely from estimates based on existing models. We determined and emphasized the importance of a validation procedure for applications using an existing model when no species-specific or local generalized models are available. If low accuracy is verified, a calibration of a previous model can be a reasonable choice. Future work is needed to better determine and assess these findings across a broader range of forest types.

References

- ALVARES CA, STAPE JL, SENTELHAS PC, DE MORAES GONÇALVES JL & SPAROVEK G. 2013. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift* 22(6): 711-728. Doi: 10.1127/0941-2948/2013/0507.
- ALVES LF, VIEIRA AS, SCARANELLO MA, CAMARGO PB, SANTOS FAM, JOLY CA & MARTINELLI LA. 2010. Forest structure and live aboveground biomass variation along an elevational gradient of tropical Atlantic moist forest (Brazil). *Forest Ecology Management* 260: 679-691. Doi: 10.1016/j.foreco.2010.05.023.
- BACCINI A, GOETZ SJ, WALKER WS ET AL. 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature*. Doi: 10.1038/NCLIMATE1354.

- BAITELLO JB, AGUIAR OT, ROCHA FT, PASTORE JA & ESTEVES R. 1992. Florística e fitossociologia do estrato arbóreo de um trecho da Serra da Cantareira (Núcleo Pinheirinho) - SP. *Revista do Instituto Florestal* 4: 291-297.
- BASTIN JF, FAYOLLE A, TARELKIN Y ET AL. 2015. Wood specific gravity variations and biomass of central African tree species: the simple choice of the outer wood. *PLoS ONE* 10(11): 1-16.
- BATISTA JLF, DO COUTO HTZ, SILVA FILHO DF. 2016. *Quantificação de Recursos Florestais. Árvores, Arvoredos e Florestas*. Oficina de Textos, São Paulo.
- BITTERLICH W. 1984. *The Relascope Idea: Relative Measurement in Forestry*. Page Bros, Norwich.
- BRYAN J, SHEARMAN P, ASH J & KIRKPATRICK J.B. 2010. Impact of logging on aboveground biomass stocks in lowland rain forest, Papua New Guinea. *Ecological Applications* 20(8): 2096-2103. Doi: 10.1890/09-1818.1.
- BURGER DM & DELITTI WBC. 2008. Allometric models for estimating the phytomass of a secondary Atlantic Forest area of southeastern Brazil. *Biota Neotropica* 8(4): 131-136. doi: 10.1590/S1676-06032008000400012.
- CEPAGRI. 2017. Centro De Pesquisas Meteorológicas E Climáticas Aplicadas À Agricultura. Available at: <<http://www.cepagri.unicamp.br/>>. Assessed at August 8 2017.
- CHAVE J, ANDALO C, BROWN S ET AL. 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* 145: 87-99. Doi: 10.1007/s00442-005-0100-x.
- CHAVE J, COOMES DA, JANSEN S, LEWIS SL, SWENSON NG, ZANNE AE. 2009. Towards a worldwide wood economics spectrum. *Ecology Letters* 12(4): 351-366. <http://dx.doi.org/10.1111/j.1461-0248.2009.01285.x>.

- CHAVE J, RÉJOU-MÉCHAIN M, BÚRQUEZ A ET AL. 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology* 20(10): 3177-3190. Doi: 10.1111/gcb.12629.
- CHOJNACKY DC, HEATH LS & JENKINS JC. 2014. Updated generalized biomass equations for North American tree species. *Forestry* 87: 129–151. Doi:10.1093/forestry/cpt053.
- CLARK DB & KELLNER JR. 2012. Tropical forest biomass estimation and the fallacy of misplaced concreteness. *Journal of Vegetation Science* 23: 1191–1196. Doi: 10.1111/j.1654-1103.2012.01471.x.
- COLMANETTI MAA, WEISKITTEL A, BARBOSA LM ET AL. 2018a. Aboveground biomass and carbon of the highly diverse Atlantic Forest in Brazil: Comparison of alternative individual tree modeling and prediction strategies. *Carbon Management* In review.
- COLMANETTI MAA, FERRAÇO SCOLFORO H, WEISKITTEL A & do COUTO HTZ. 2018b. Calibrating individual tree biomass models for contrasting tropical species at a diverse site in the Atlantic Forest, Brazil. *Canadian Journal of Forest Research*. In review.
- DJUIKOUO MNK, DOUCET JL, NGUEMBOU CK, LEWIS SL & SONKÉ B. 2010. Diversity and aboveground biomass in three tropical forest types in the Dja Biosphere Reserve, Cameroon. *African Journal of Ecology* 48(4): 1053–1063. Doi: 10.1111/j.1365-2028.2010.01212.x.
- FAUSET S, JOHNSON MO, GLOOR M ET AL. 2015. Hyperdominance in Amazonian forest carbon cycling. *Nature Communications* 6: 7857. Doi: 10.5521/FORESTPLOTS.NET/2015_1 10.5521/FORESTPLOTS.NET/2015_1.

- GIBBS HK, BROWN S, NILES JO, FOLEY JA. 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters* 2: 1-13 Doi: 10.1088/1748-9326/2/4/045023.
- GUARIGUATA MR & OSTERTAG R. 2001. Neotropical secondary forest succession: changes in structural and functional characteristics. *Forest Ecology and Management* 148: 185-206.
- GUITET S, HÉRAULT B, MOLTO Q, BRUNAUX O & COUTERON P. 2015. Spatial Structure of Above-Ground Biomass Limits Accuracy of Carbon Mapping in Rainforest but Large Scale Forest Inventories Can Help to Overcome. *PLoS One* 10(9): 1-22. Doi: 10.1371/journal.pone.0138456.
- IBGE. 1992. *Manual Técnica da Vegetação Brasileira*. Fundação Instituto Brasileiro de Recursos Naturais e Estudos Ambientais, Rio de Janeiro.
- IBGE. 2010. Instituto Brasileiro de Geografia - Diretoria de Geociências. (Base vetorial contínua (BCIM). Rio de Janeiro. Escala: 1:1.000.000; Accessed at 19 October 2017). http://geofpt.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bcim/
- IBGE. 2017. Instituto Brasileiro de Geografia. Accessed at Jan. 08/2018. Available at: <https://cidades.ibge.gov.br/>
- IPCC (Intergovernmental Panel on Climate Change). 2006. Intergovernmental Panel on Climate Change Guidelines for National Greenhouse Gas Inventories. Chapter 4: Forest Lands. Japan: National Greenhouse Gas Inventories Programme.
- JENKINS JC, CHOJNACKY DC, HEATH LS & BIRDSEY RA. 2003. National-Scale Biomass Estimators for United States Tree Species. *Forest Science* 49(1): 12-35.
- LINDNER A & SATTLER D. 2012. Biomass estimations in forests of different disturbance history in the Atlantic Forest of Rio de Janeiro, Brazil. *New Forests* 43:287–301. Doi: 10.1007/s11056-011-9281-9.

- MANAGEMENT PLAN. 2009. *Parque Estadual da Cantareira: Plano de Manejo*.
- MARRA DM, HIGUCHI N, TRUMBORE SE ET AL. 2016. Predicting biomass of hyperdiverse and structurally complex central Amazonian forests – a virtual approach using extensive field data. *Biogeoscience* 13: 1553-1570. Doi: 10.5194/bg-13-1553-2016.
- MARCHIORI NM, ROCHA HR, TAMASHIRO JY & AIDAR PMA. 2016. Tree community composition and aboveground biomass in a secondary Atlantic Forest, Serra do Mar State Park, São Paulo, Brazil. *Cerne* 22(4): 501-514. Doi: 10.1590/01047760201622042242.
- MMA. 2012. Ministério do Meio Ambiente. (Mapa da área de aplicação da Lei 11.428 de 2006. Escala 1:5.000.000; Accessed at: 19 October 2017). http://www.mma.gov.br/estruturas/202/_arquivos/shape_mata_atlantica_ibge_5milhoes_policonica_sirgas2000shp_202.zip>.
- NASCIMENTO ART, SCARIOT A, SILVA JA & SERVILHA AC. 2004. Estimativas de área basal e uso do relascópio de Bitterlich em amostragem de Floresta Estacional Decidual. *Ciência Florestal* 14(2): 169-176. Doi: 10.5902/198050981816.
- NELSON BW, MESQUITA R, PEREIRA JLG, SOUZA SGA, BATISTA GT & COUTO LB. 1999. Allometric regressions for improved estimate of secondary forest biomass in central Amazon. *Forest Ecology and Management* 117:149-167. Doi: 10.1016/S0378-1127(98)00475-7.
- NOGUEIRA EM, FEARNside PM, NELSON BW, BARBOSA IRI & KEIZER EWH. 2008. Estimates of forest biomass in the Brazilian Amazon: New allometric equations and adjustments to biomass from wood-volume inventories. *Forest Ecology and Management* 256: 1853–1867. Doi:10.1016/j.foreco.2008.07.022.

- NYIRAMBANGUTSE B, ZIBERA E, UWIZEYE FK ET AL. 2017. Carbon stocks and dynamics at different successional stages in an Afromontane tropical forest. *Biogeosciences* 14: 1285–1303. Doi: 10.5194/bg-14-1285-2017.
- PAN Y, BIRDSEY RA, FANG J ET AL. 2011. A large and persistent carbon sink in the world's forests. *Science* 333: 988-993. Doi: 10.1126/science.1201609.
- R CORE TEAM. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org/>.
- RIBEIRO MC, METZGER JP, MARTENSEN AC, PONZONI FJ & HIROTA MM. 2009. The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation* 142: 1141–1153. Doi: 10.1016/j.biocon.2009.02.021.
- ROLIM SG, JESUS RM, NASCIMENTO HEM, COUTO HTZ & CHAMBERS J.C. 2005. Biomass change in an Atlantic tropical moist forest: the ENSO effect in permanent sample plots over a 22-year period. *Oecologia* 142: 238–246. Doi: 10.1007/s00442-004-1717-x.
- SAATCHI SS, HARRIS NL, BROWN S ET AL. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *PNAS*. 108 (24): 9899–9904. Doi: 10.1073/pnas.1019576108.
- SCOLFORO HF, SCOLFORO JRS, de MELLO JM, de MELLO CR, MORAIS VA 2016. Spatial interpolators for improving the mapping of carbon stock of the arboreal vegetation in Brazilian biomes of Atlantic forest and Savanna. *Forest Ecology and Management* 376: 24–35. Doi: 10.1016/j.foreco.2016.05.047.
- SHIMAMOTO CY, BOTOSSO P & MARQUES, MCM. 2014. How much carbon is sequestered during the restoration of tropical forests? Estimates from tree species in the Brazilian Atlantic Forest. *Forest Ecology and Management* 329: 1-9. Doi: 10.1016/j.foreco.2014.06.002.

- SHIVER BD & BORDERS BE. 1996. Chapter 7: Double sampling or two-phase sampling. Pp 190-229 in Shiver BD & Borders BE. (eds) *Sampling techniques for forest resource inventory*. John Wiley & Sons, New York.
- SOIL SURVEY STAFF. 2014. Keys to Soil Taxonomy. 12th Edition. Washington, DC: US Department of Agriculture–Natural Resources Conservation Service.
- SOS Mata Atlântica, Instituto Nacional de Pesquisas Espaciais. 2013. (Atlas dos remanescentes florestais da Mata Atlântica, período de 2013 a 2014; Accessed at: 19 October 2017). <http://www.sosmatatlantica.org.br>
- VIEIRA SA, ALVES LF, DUARTE-NETO PJ ET AL. 2011. Stocks of carbon and nitrogen and partitioning between above- and belowground pools in the Brazilian coastal Atlantic Forest elevation range. *Ecology and Evolution* Doi: 10.1002/ece3.41.
- YANG W, KOBAYASHI H, NASAHARA KN, SUZUKI R & KONDOH A. 2017. Quantitative Evaluation of Bitterlich Sampling for Estimating Basal Area in Sparse Boreal Forests and Dense Tropical Forests. *Open Journal of Forestry* 7: 143-156. Doi: 10.4236/ojf.2017.72009.
- ZANNE AE, LOPEZ-GONZALEZ G, COOMES DA ET AL. 2009. Towards a worldwide wood economics spectrum. *Dryad Digital Repository* <http://dx.doi.org/10.5061/dryad.234>.

4. CALIBRATING INDIVIDUAL TREE BIOMASS MODELS FOR CONTRASTING TROPICAL SPECIES AT A DIVERSE SITE IN THE ATLANTIC FOREST, BRAZIL

ABSTRACT

Some studies have suggested improved performance of a species-specific approach for biomass prediction. The calibration of previously developed species-specific model for certain species could be a reasonable alternative for highly diversity forest like the Atlantic Forest of Brazil. The primary research goal of this study was evaluation of the potential to calibrate an individual tree aboveground biomass model for a new species by using linear mixed-effect (LME). The performance of LME calibration approach was compared to ordinary least square (OLS) calibration. Comparisons were made across varying sample sizes, tree size classes, and sample selection methods. Our results suggested that calibration was not efficient for all cases, but was required when the new species had a wood specific gravity outside of the range from the data used to fit the original model. LME calibration method had significantly better performance than the OLS across all sample sizes, tree size classes, and sample selection methods examined. One to three randomly selected trees was sufficient to calibrate a biomass model for new species using LME, while stratification by tree size did not provide any improvement. Overall, the results highlight the potential of model calibration to improve both biomass and carbon estimates in tropical forests.

Keywords: Predictive models; Tropical forest; Atlantic Forest; Destructive sampling; BLUP

4.1. Introduction

Regarding the importance of tropical forests in the global carbon cycle, many studies quantified the carbon stocked in their aboveground biomass at local (e.g., Vieira et al. 2011; Vicent et al. 2015), regional (e.g., Nogueira et al. 2008a; Nogueira et al. 2008b; Scolforo et al. 2015; Scolforo et al. 2016), continental (e.g., Lewis et al. 2009), or global scales (e.g., Pan et al. 2011, Saatchi et al. 2011). Methods used for biomass estimation have varied according to levels of specificity as regional biomass conversion factors as well as stand- and tree-level biomass equations have been used, but only the latter method focuses on the basic unit in traditional

inventory and consequently, it demands more detailed input data (Temesgen et al. 2015). Therefore, the tree-level approach is expected to provide improvements to other methods and may provide increased accuracy of biomass estimates.

Some studies have suggested a better performance of species-specific models for individual tree biomass prediction (Nelson et al 1999; van Breugel et al. 2011), while mixed-effect models with species as a random effect can improve estimates in certain situations (Sotomayor 2013; Vismara 2013). Besides biomass, height increment (eg. Russell et al. 2014), height-diameter (eg. Crecente-Campo et al. 2014, Lam et al. 2017) and stem taper have been improved by using species as random effect (eg. MacFarlane and Weiskittel 2016; Scolforo et al. 2018), highlighting the potential importance of this approach.

Mixed-effect models often lead to higher accuracy and precision, but are generally restricted to the same species and/or individuals used in the model fitting. This issue can be rather challenging for areas with high species diversity like tropical forests. Generally, the large abundance of a few hyper dominant species in these forests concentrates the majority of the biomass and carbon stocks (Fauset et al. 2015) so sampling often focuses solely on those species. This strategy is a feasible approach but requires the need for species-specific models (Nelson et al 1999; van Breugel et al. 2011; Scolforo et al. 2017) and there is less clarity on how to address biomass and carbon quantification for the least abundant species.

Although abundant species play an important role on biomass quantification at large spatial scales, significant variation in the most abundant species are often observed in tropical forests (Fauset et al. 2015). This variation can be seen across contrasting forest types (Eisenlohr and Oliveira-Filho 2014), regions (Slik et al. 2003; Réjou-Méchain et al. 2008), or even locally (Webb and Peart 2000). Given this variation, species-specific models may not be an effective

approach, particularly if being applied across broad spatial scales that include various forest types. Vismara (2013) highlighted this potential issue and proposed a calibration of the mixed-effects model for each new species sampled in an inventory.

Calibrating single or multi-level mixed-effects models have been rather commonly used for single species stands (Lappi 1991; de-Miguel et al. 2014; Arias-Rodil et al. 2015; Vismara et al. 2016). This approach is based on prediction of random effects using the best linear unbiased predictor (BLUP), where the variance-covariance matrix is altered by new observations using a single or (e.g., Temesgen et al. 2008, de-Miguel et al. 2014, Vismara et al. 2016) multiple levels (e.g., Crecente-Campo et al. 2014, Arias-Rodil et al. 2015) with one (Temesgen et al. 2008, Vismara et al. 2016) or more random effects on parameters (Temesgen et al. 2008; Crecente-Campo et al. 2014, Arias-Rodil et al. 2015). Calibration using crossed random effect has also been used (Vismara 2013). When compared to more traditional calibration methods, the use of mixed-effect has been commonly shown to be superior (Temesgen et al. 2008; Crecente-Campo et al. 2014, Arias-Rodil et al. 2015).

The calibration of mixed-effect models can be additionally improved by increasing the number of trees (Temesgen et al. 2008; de-Miguel et al. 2014; Vismara, 2016), widening the diameter at breast height (*dbh*) range (Temesgen et al. 2008), increasing *dbh* size (Crecente-Campo et al. 2014) or using stratified sampling (Temesgen et al. 2008; de-Miguel et al. 2014). Furthermore, mixed-effects model calibration has been found to perform better than new species-specific models fitted by ordinary least squares (OLS; de-Miguel et al. 2014). Although the calibrations methods have been used for single species stands, these procedures have rarely been used for natural tropical forest where predictive gains could be rather high given the number of species present.

To evaluate the use of mixed-effect model calibration in tropical forests, we assessed the local calibration of a biomass model for new species using a destructively obtained tree-level

measurements in the highly diverse Atlantic Forest of Brazil where the abundant species often varies across sites. Specific research objectives were to: (i) calibrate LME biomass model using BLUP to predict the random effect for a new species; (ii) compare this approach to OLS calibration performance; (iii) assess the influence of various sample sizes, selection of sample tree methods, and *dbh* size on calibration performances; and (iv) compare all of these calibration approaches to more generalized models.

4.2. Materials and Methods

4.2.1. Study site

The study site was a secondary Atlantic Forest located north of city of São Paulo and southwest of Cantareira Mountains State Park (Parque Estadual da Cantareira, in Portuguese) at an altitude of approximately 840 m. The area was covered by vegetation classified as a moist forest with some species of seasonal semi-deciduous forest. A general predominance of species was from the families of Myrtaceae, Fabaceae, Euphorbiaceae and Lauraceae.

The climate at the region according to Köppen's climate classification adapted to Brazil is Cwa (Alvares et al. 2013), with an average annual temperature of 20.7 °C (minimum average: 15.7°C, maximum average: 25.7°C) and annual rainfall of 1,400 mm (Cepagri 2017). The topography is characterized by steep slopes, and the soil classification according to soil taxonomy (Soil Survey Staff, 2014) was Ultisol Udic Udults Typic Hapludults.

4.2.2. Data collection

Trees ≥ 5 cm *dbh* were measured in ten 20 x 40 m rectangular plots (800 m²). The species were identified and updated according to Brazilian Species Flora List (Flora do Brasil 2016) and Missouri Botanical Garden (Mobot 2016), and families were organized according to A.P.G. IV system. The collected species were incorporated into the Herbarium of the Botanical Institute of São Paulo, São Paulo City.

In total 106 trees, 96 trees belonging to the 16 most abundant species plus 10 of the largest trees were destructively sampled. To ensure a range of *dbh*s, trees of three *dbh* classes (5-10 cm, 10-15 cm, and >15 cm) were selected for each species. After harvesting, samples of wood were collected (three disk of a constant thickness of 5 cm with bark) along the stem (base, and 50% below and above height to crown base) for woody specific gravity (*wsg*) analysis. The crown (foliage and branches) was weighed together, and a sample of both foliage and branch wood was taken. All samples were placed in plastic bags, and taken for biomass quantification.

4.2.3. Biomass and woody specific gravity determination

All harvested trees were weighed completely in the field using a Dynamometer scale. For each tree, three discs were selected from the base, middle, and top of height to crown base. These discs were weighed after harvesting and then again after being dried at 105°C until reaching a constant weight. The same procedure was repeated for the crown samples. An average of the moisture content (*u*%) for the three discs was calculated for each tree and used to estimate dry stem biomass and then summed with crown biomass to obtain total aboveground dry biomass:

$$(1) \quad db = m. \left(1 - \left(\frac{u\%}{100} \right) \right)$$

where *db* is dry biomass, *u*% is average moisture content of three discs, and *m* is green biomass.

Wood specific gravity (wsg) was based on the variation of the sample mass using a hydrostatic balance. The method required the determination of stem mass samples at the maximum moisture content when immersed in water and the dry mass at 105°C in a forced-air oven. Calculation of wood specific gravity was as follows:

$$(2) \quad wsg = m_1 \cdot (m_2 - m_3)^{-1}$$

where wsg = woody specific gravity, m_1 = dry biomass at 105°C, m_2 = green biomass, and m_3 = green biomass submerged. An average of woody specific gravity for the three discs were obtained for each tree.

4.2.4. Data analysis

4.2.4.1. Species-specific model

Power equation in linear form with log-log transformation was used to develop species-specific models of aboveground biomass (eq. 3). Species was used as random effect on both the intercept and slope.

$$(3) \quad \begin{aligned} agb_{ij} &= \phi_{1i} \cdot dbh_{ij}^{\phi_{2i}} \cdot \epsilon_{ij} \\ \ln(a_{ij}) &= \phi_{1i} + \phi_{2i} \cdot \ln(dbh_{ij}) + \epsilon_{ij} \\ \phi_i &= \begin{bmatrix} \phi_{1i} \\ \phi_{2i} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} = \beta + b_i \\ b_i &\sim N(0, \Psi), \quad \epsilon_{ij} \sim N(0, \sigma^2) \end{aligned}$$

where agb_{ij} is the aboveground biomass for tree from the specie i with the j th dbh , β are the fixed effects representing the population average of the ϕ_i , b_i is the random effect that represent the deviation from the population average, in this case the species i , with variance-covariance

matrix Ψ . The errors were assumed to be independent for different species and within the groups, ϵ_{ij} are assumed to be independent for different observations and the random effects. The β and Ψ were estimated by the restricted maximum likelihood estimation using the function `lme` from the `nlme` package (Pinheiro et al. 2016) in R (R Core Team 2016). The variance for the ϵ_{ij} was modeled by using the function `weights` from the `nlme` package, where the variance changes according to a specific covariate (eq.4):

$$(4) \quad \text{Var}(\epsilon_{ij}) = \sigma^2 v_{ij} v^{2\delta},$$

where: v_{ij} is the covariate *dbh*, δ is the value for the variance parameter.

Predicted aboveground biomass was back transformed to original scale to remove the systematic bias using the correction proposed by Sprugel (1983):

$$(5) \quad \widehat{agb} = e^{(\hat{\mu} + SEE^2/2)}$$

$$(6) \quad SEE = \sqrt{\frac{\sum(\log(y_i) - \widehat{\log}(y_i))^2}{N - n_p}}$$

where \widehat{agb} is the aboveground biomass, $\hat{\mu}$ is the estimated mean, and the $\hat{\sigma}^2$ is the variance estimated from the random effect, *SEE* is the standard error of the estimates, $\log(y_i)$ is the observed values for biomass, $\widehat{\log}(y_i)$ is the predicted values for biomass, *N* is the number of observations, n_p is the number of parameters of the model.

4.2.4.2. Generalized models

The fixed-effects portion from the species-specific models were used as a generalized model (i.e. multi-specific model). The models included *dbh* only and the enhanced model with *dbh* and *wsg* were using the species that were not used in the calibration procedures.

4.2.5. Model calibration procedures

4.2.5.1. Linear Mixed-Effect (LME) Approach

Eq. (3) can be linearized and re-written as:

$$(7) \quad \ln(a_{ij}) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) \cdot \ln(dbh_{ij}) + \epsilon_{ij}$$

The calibration procedure was based on the prediction of the random effects using the best linear unbiased predictor (BLUP; Lappi 1991). Similarly, Eq. (7) can be simplified as:

$$(8) \quad y_i = \mu + Zb_i + \epsilon_i$$

where μ is the fixed part of Eq. (7), b_i is a vector of random effects for each species i , Z is the design matrix, and ϵ_i is the vector of random residuals.

The calibration procedure used the destructive sample of at least one tree from a species not used in the fitting of eq.(3). In eq.(9), we have the variance-covariance matrix of the random effects Ψ ($\text{var}(b_i)$) in eq.(3) is represented by D , and the variance of the random residuals multiplying I matrix by $R = \sigma^2 I$. D and Z are the square $p \times p$ matrix and the design matrix $n \times p$, respectively, where n is the number of trees used for calibration and p the number of the random parameters. The first column of Z is filled by values of 1, and others contain the covariates measurements that are assumed to have the random effects (dbh in this case) for every tree from the new species i .

$$(9) \quad \begin{bmatrix} b_i \\ y_i \end{bmatrix} \left(\begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} D & DZ' \\ ZD & ZDZ' + R \end{bmatrix} \right)$$

The BLUP of the random effects for the new species b_i can be obtained as:

$$(10) \quad BLUP(b_i) = \hat{b}_i = DZ'(ZDZ' + R)^{-1} (y_i - \mu)$$

where the y_i is the average aboveground biomass of the new observations in logarithmic scale and μ is the average predicted values for the same observations using the fixed part from eq.3.

The variance can be predicted by:

$$(11) \quad \text{var}(\hat{b}_i - b_i) = D - DZ'(ZDZ' + R)^{-1}ZD$$

According to Robinson (1991), the BLUP is biased towards to the prediction of the new observations, in this case, the new species i . Based on de-Miguel et al. (2014), the bias decreased according to the increase of observations so in this case when we increased the number of trees from the new species.

4.2.5.2. Ordinary Least Square (OLS)

The LME calibration was compared to an alternative calibration method, using the OLS correction factor (eq.12), according to Temesgen et al. (2008), adapted for biomass for the new species i .

$$(12) \quad k_i^* = \frac{\sum_{j=1}^{n_{ij}}(\hat{a}_{ij} \cdot a_{ij})}{\sum_{j=1}^{n_{ij}} \hat{a}_{ij}^2}$$

where: k_i^* is the correction factor for the slope \hat{a}_{ij} is the predicted biomass for each biomass from the fixed part of eq.3, in linear form, for the new biomass observation, a_{ij} , for each tree from species i .

The generalized model (dbh only) calibrated by OLS can be described as:

$$(13) \quad \ln(a_{ij}) = k_i^* \cdot (\phi_{1i} + \phi_{2i} \cdot \ln(dbh_{ij}))$$

4.2.6. Species and trees selection for a calibration procedure

Eq.(3) was estimated 16 times using contrasting approaches. For each fitting, one species was kept out from the data used in the species-specific biomass model, and then used in the calibration procedure, while the fixed part of this model was used as a generalized model. For example, in the calibration of *A. petiolulatus*, Eq.(3) was fitted using the 15 other remaining species.

For each species, different strategies of tree selection for calibration were used with varying sample sizes using: (i) random selection starting with one tree to all trees sampled according to species (6-8 trees); (ii) stratification according to *dbh* ($5 \leq dbh \leq 10\text{cm}$; $10 < dbh \leq 15\text{cm}$; $15\text{cm} < dbh$) with at least one tree from each class; (iii) sequential sampling starting with smallest *dbh* for sample size with one tree, following the second smallest *dbh* for sample size at two trees, until all of trees been used in the calibration and (iv) sequential sampling with the largest *dbh* for sample size of one tree, after following the second largest *dbh* for sample size at two trees until all of the trees been used in the calibration.

We used bootstrapping resampling with 1000 samplings per different sample sizes and different sampling types: (i) completely random selection starting with 1, 2, 3, ..., 8 trees; (ii) stratified selection according to the classes used in the data collection ($5 \leq dbh < 10\text{cm}$; $10 \leq dbh < 15\text{cm}$; $15\text{cm} \leq dbh$), and a random sampling inside each strata beginning with 3 and ending with 6 trees; (iii) systematic sampling from 1 to 8 trees. The mean and standard deviation were obtained after the sampling selection.

4.2.7. Validation procedure

The calibrated models using both OLS and LME were compared with their generalized models using root mean square error (RMSE; eq.14) and mean bias (MB; eq.15) both on the relative

scale, which was calculated by dividing the values by the mean observed biomass and multiplying by 100. The mean and standard deviation were calculated for RMSE based on bootstrap sampling.

$$(14) \quad RMSE(\%) = \sqrt{\frac{\sum_{i=1}^n (a_{ij} - \hat{a}_{ij})^2}{n}} \cdot \frac{100}{\frac{\sum_{i=1}^n a_{ij}}{n}}$$

$$(15) \quad MB(\%) = \frac{\sum_{i=1}^n (a_{ij} - \hat{a}_{ij})}{\sum_{i=1}^n a_{ij}} \cdot 100$$

where: a_{ij} are the observed values of aboveground biomass; \hat{a}_{ij} are the predicted values of aboveground biomass; n is the number of observations.

4.3. Results

4.3.1. Species characterization

Observations from 16 contrasting species were available for this analysis and each species had six to eight trees harvested with a range of tree sizes sampled (Table 1). *C. speciosa* and *C. oblongifolia*, had the smallest and highest *wsg* across the species, respectively, while the *C. canjerana* and *C. speciosa* had the smallest and highest *dbh*, respectively (Fig. 1).

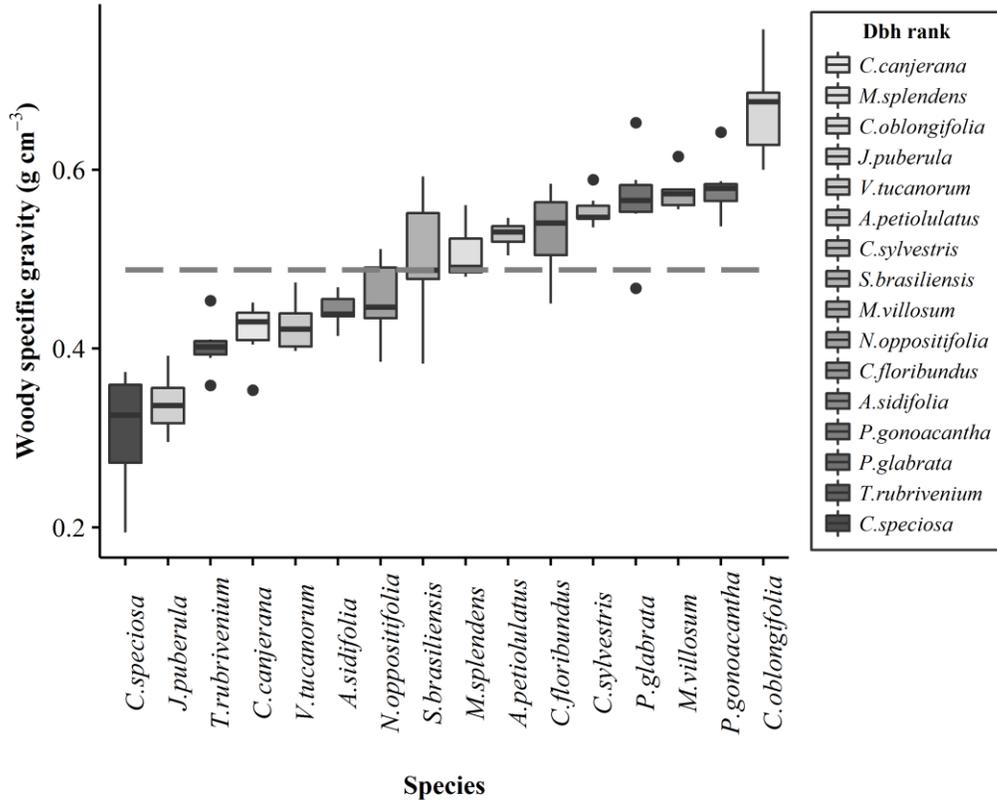


Figure 1. Wood specific gravity (*wsg*) across 16 species from the Atlantic Forest in Brazil. Broken gray line represents a median of all *wsg* (0.49 g cm^{-3}). All 16 species are ranked in gray scale according to average *dbh*.

Table 1. Summary of the destructively sampled individual tree data from Atlantic forest at Serra da Cantareira-SP/Brazil.

Species	Number of trees	dbh (cm)		wsg (g cm ⁻³)		agb (kg)	
		Range	Mean (sd)	Range	Mean (sd)	Range	Mean (sd)
<i>Alchornea sidifolia</i> Müll. Arg.	7	8.2 - 34.3	18.8 (12.0)	0.44 - 0.47	0.45 (0.01)	15.0 - 568.6	182.4 (233.4)
<i>Allophylus petiolulatus</i> Radlk.	6	7.6 - 20.8	13.1 (5.7)	0.50 - 0.55	0.53 (0.02)	12.7 - 199.5	67.0 (71.2)
<i>Cabralea canjerana</i> (Vell.) Mart.	6	7.0 - 11.8	9.3 (1.9)	0.35 - 0.45	0.42 (0.04)	7.6 - 25.8	17.5 (7.74)
<i>Casearia sylvestris</i> Sw.	7	8.2 - 20.4	13.2 (4.7)	0.54 - 0.59	0.55 (0.02)	18.1 - 137.5	64.1 (48.5)
<i>Ceiba speciosa</i> (A. St.-Hil.) Ravenna	8	6.0 - 67.8	32.1 (20.8)	0.19 - 0.37	0.31 (0.06)	2.3 - 1,114.6	294.8 (396.0)
<i>Croton floribundus</i> Spreng.	6	7.6 - 30.0	16.8 (9.4)	0.45 - 0.58	0.53 (0.05)	21.9 - 415.9	153.6 (161.3)
<i>Cupania oblongifolia</i> Mart.	7	6.0 - 30.6	12.9 (9.1)	0.60 - 0.76	0.68 (0.05)	9.7 - 445.9	100.4 (170.5)
<i>Jacaranda puberula</i> Cham.	6	6.3 - 20.6	12.9 (5.8)	0.30 - 0.39	0.34 (0.03)	4.0 - 148.9	44.2(54.9)
<i>Machaerium villosum</i> Vogel	6	7.7 - 27.2	15.3 (7.6)	0.56 - 0.61	0.58 (0.02)	12.9 - 300.7	104.6 (120.6)
<i>Myrcia splendens</i> (Sw.) DC.	6	6.2 - 14.9	11.8 (3.4)	0.48 - 0.56	0.51 (0.03)	9.4 - 80.5	51.7 (27.6)
<i>Nectandra oppositifolia</i> Ness	6	7.3 - 26.5	15.4 (7.5)	0.39 - 0.51	0.45 (0.05)	11.4 - 251.8	101.7 (96.9)
<i>Pera glabrata</i> (Schott) Poepp. ex Baill.	7	5.4 - 68.4	25.0 (23.6)	0.47 - 0.65	0.57 (0.06)	5.6 - 1,917.1	466.0 (701.8)
<i>Piptadenia gonoacantha</i> (Mart.) J. F. Macbr.	7	7.1 - 42.7	20.3 (13.2)	0.54 - 0.64	0.58 (0.03)	10.6 - 711.6	250.1 (293.4)
<i>Sessea brasiliensis</i> Toledo	7	8.2 - 22.8	15.3 (4.6)	0.38 - 0.59	0.50 (0.07)	10.1 - 122.6	70.0 (42.4)
<i>Tetrorchidium rubrivenium</i> Poepp.	8	8.6 - 58.3	27.1 (20.1)	0.36 - 0.45	0.40 (0.03)	16.0 - 1,493.8	407.4 (536.1)
<i>Vochysia tucanorum</i> Mart.	6	8.4 - 18.8	12.9 (4.7)	0.40 - 0.47	0.43 (0.03)	18.8 - 91.7	46.1 (34.6)
Overall	106	5.4 - 68.5	17.3 (13.0)	0.19 - 0.76	0.49 (0.10)	2.3 - 1,917.1	158.5 (297.7)

4.3.2. Calibration methods' performance

Improved performance for the calibrated models (LME or OLS) was not observed for all species (Fig. 2 and 3). The generalized models (*dbh* only) had a better performance than the calibration methods for *A. petiolulatus*, *A. sidifolia*, *C. sylvestris*, *N. oppositifolia* and *A. petiolulatus*, while *A. sidifolia*, *J. puberula*, and *T. rubrivenium* saw better performance with the generalized model that included *dbh* with *wsg* (Fig 1). In general, the LME calibration method had better performance with a smaller standard deviation than OLS. Except for *C. speciosa*, LME calibration had a similar performance across the various sample sizes examined (Fig 1 and 2).

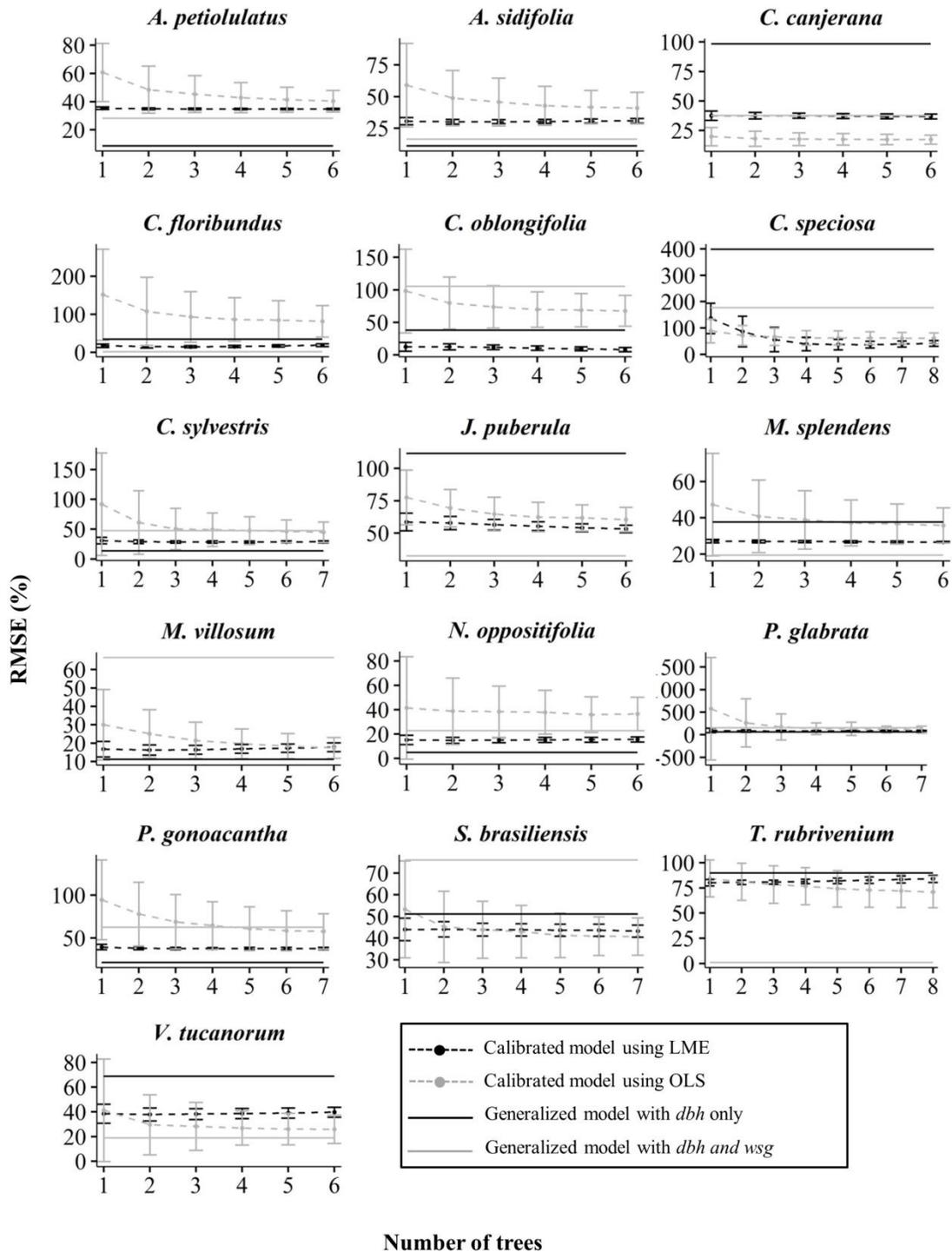


Figure 2. Percent root mean square error (RMSE) for calibration of aboveground biomass prediction across 16 species with varying sample sizes from the Atlantic Forest in Brazil using linear mixed-effects (LME) and ordinary least square (OLS) as well as generalized models with diameter at breast height (*dbh*) and *dbh* with wood specific gravity (*wsg*). The error bars are \pm two standard deviations.

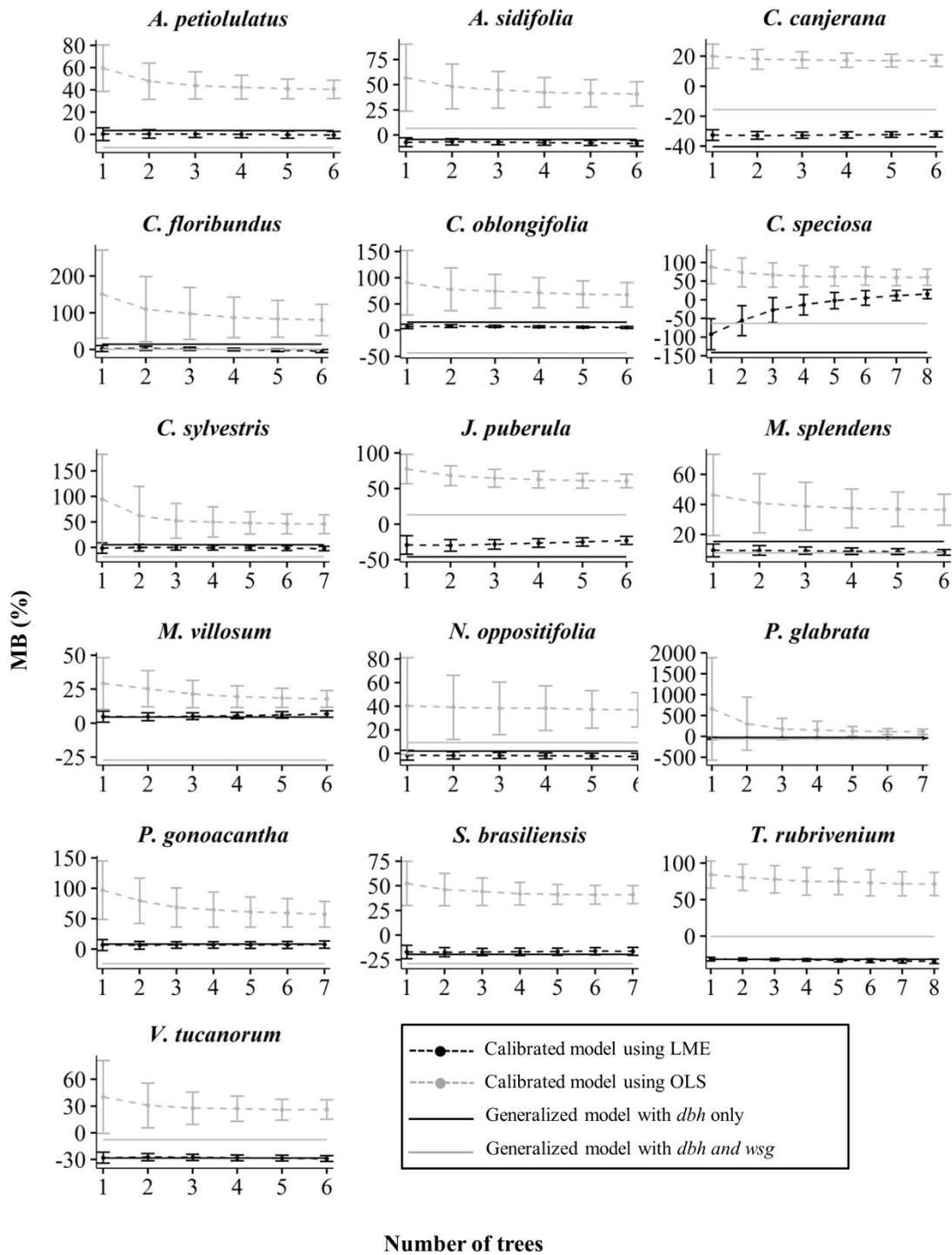


Figure 3. Percent mean bias (MB) for calibration of aboveground biomass prediction across 16 species with varying sample sizes from the Atlantic Forest in Brazil using linear mixed-effects (LME) and ordinary least square (OLS) as well as generalized models with diameter at breast height (*dbh*) and *dbh* with wood specific gravity (*wsg*). The error bars are \pm two standard deviations.

The calibrated random effect varied according to the increasing of the number of trees used for calibration (Table 2). The general pattern of low values of random effect for intercept (β_0) and high values for slope (β_1) was observed, using one tree for calibration. The increasing of the number of trees in the calibration induced the increasing of values for intercept and the reductions in slope toward to the empirical values. The empirical values were obtained in the species-specific models based on LME and using the completed dataset, including all trees of the sixteen species.

Table 2. Empirical and calibrated random effects of 16 species in Atlantic Forest, Brazil.

Species	Calibrated random effect						Empirical random effect	
	One tree in calibration		Three trees in calibration		Six trees in calibration		β_0	β_1
	β_0	β_1	β_0	β_1	β_0	β_1		
<i>A. petiolulatus</i>	9.4e-15	0.01	0.011	0.009	0.019	0.007	0.111	4.4e-12
<i>A. sidifolia</i>	4.3e-15	0.007	0.012	0.004	0.027	0.002	0.058	6.8e-12
<i>C. canjerana</i>	-1.2e-14	-0.024	-0.005	-0.022	-0.011	-0.021	-0.183	-1.1e-11
<i>C. floribundus</i>	2.6e-14	0.038	0.066	0.022	0.158	0.013	0.408	-4.9e-11
<i>C. oblongifolia</i>	2.1e-14	0.028	0.023	0.022	0.062	0.017	0.234	3.3e-11
<i>C. speciosa</i>	-4.5e-14	-0.062	-0.594	-0.012	-0.954	0.0001	-0.990	1.2e-10
<i>C. sylvestris</i>	1.7e-14	0.022	0.016	0.014	0.036	0.011	0.211	-3.2e-11
<i>J. puberula</i>	-2.6e-14	-0.046	-0.049	-0.029	-0.117	-0.021	-0.389	1.7e-11
<i>M. splendens</i>	1.3e-14	0.026	0.012	0.022	0.028	0.02	0.252	1.4e-11
<i>M. villosum</i>	-1.7e-15	-0.001	-0.018	0.004	-0.043	0.007	0.014	6.2e-11
<i>N. oppositifolia</i>	4.2e-15	0.01	0.006	0.01	0.019	0.009	0.130	1.9e-11
<i>P. glabrata</i>	1.2e-14	0.023	0.033	0.007	0.081	0.001	0.200	-1.8e-10
<i>P. gonoacantha</i>	6.4e-15	0.005	-0.006	0.006	-0.012	0.007	0.075	8.4e-11
<i>S. brasiliensis</i>	-7.2e-15	-0.007	-0.009	-0.004	-0.015	-0.003	-0.076	8.3e-12
<i>T. rubrivenium</i>	-7.1e-16	-0.001	0.008	-0.002	0.021	-0.003	-0.023	1.1e-10
<i>V. tucanorum</i>	1.1e-15	-0.001	0.012	-0.006	0.028	-0.008	-0.034	-4.1e-11

4.3.3. Differences across species and tree selection methods for calibration

Across the different strategies for tree selection examined, stratified sampling according to three *dbh* classes and random selection had similar performance when using three to six trees for LME calibration (Fig. 4 and 5). Some species had better performance using sequential sampling with the smallest *dbh* for the calibration (*C. canejerana*, *C. oblongifolia*, *C. speciosa*, *J. Puberula*, *M. splendens* and *S. brasiliensis*), but others were improved with sequential sampling with the largest tree (*C. sylvestris*, *M. villosum*, *T. rubrivenium* and *V. tucanorum*) for RMSE. A general improved performance of using calibration with varying the *dbh* sizes was observed.

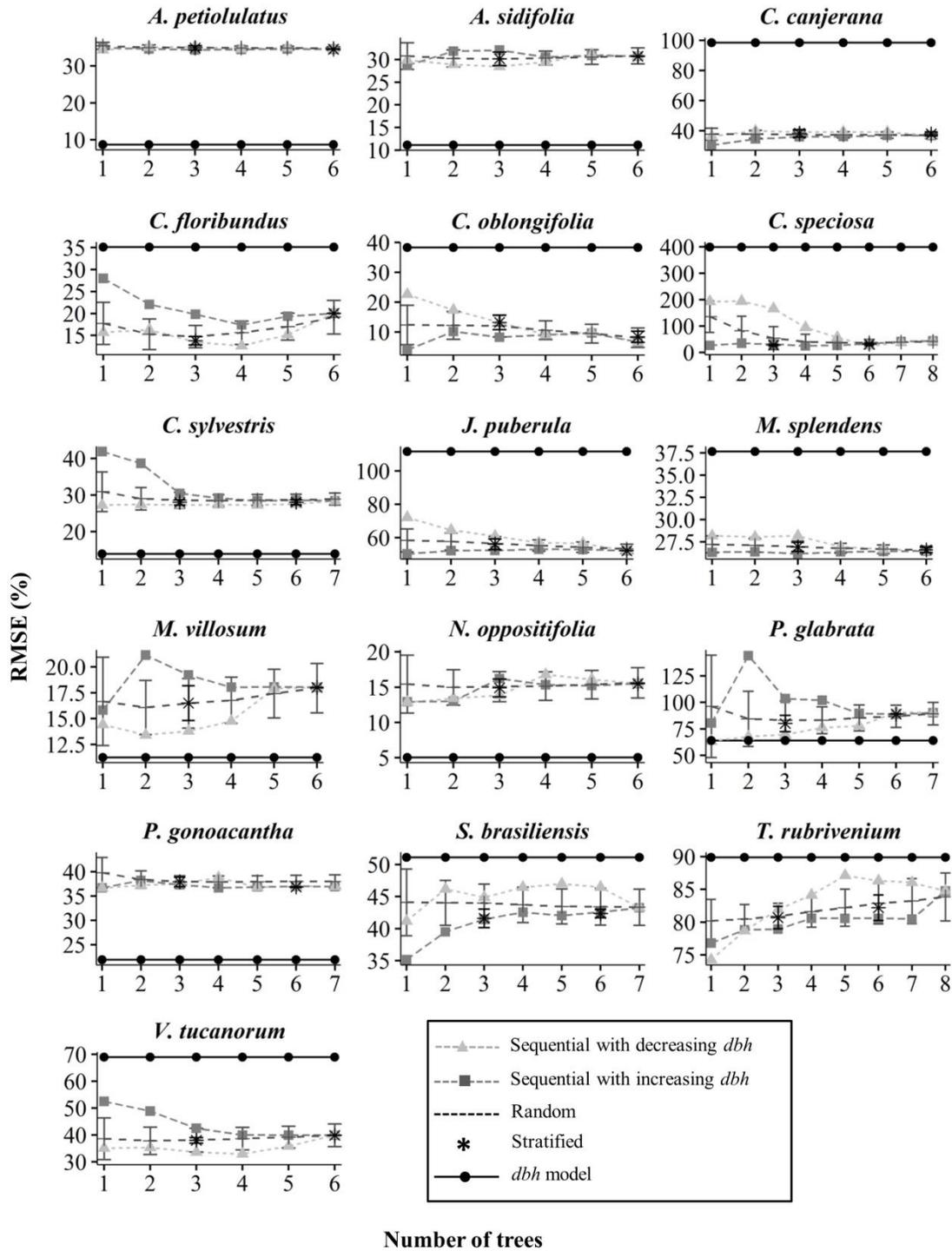


Figure 4. Percent root mean square error (RMSE) for calibration of 16 species for prediction of individual tree aboveground biomass from Atlantic Forest in Brazil using different types of sample tree selection methods for linear mixed-effects (LME) and various sample sizes. The sample selection methods included random, stratified, and sequential (increasing and decreasing) compared to the generalized model with diameter at breast height (*dbh*). The error bars are \pm two standard deviations.

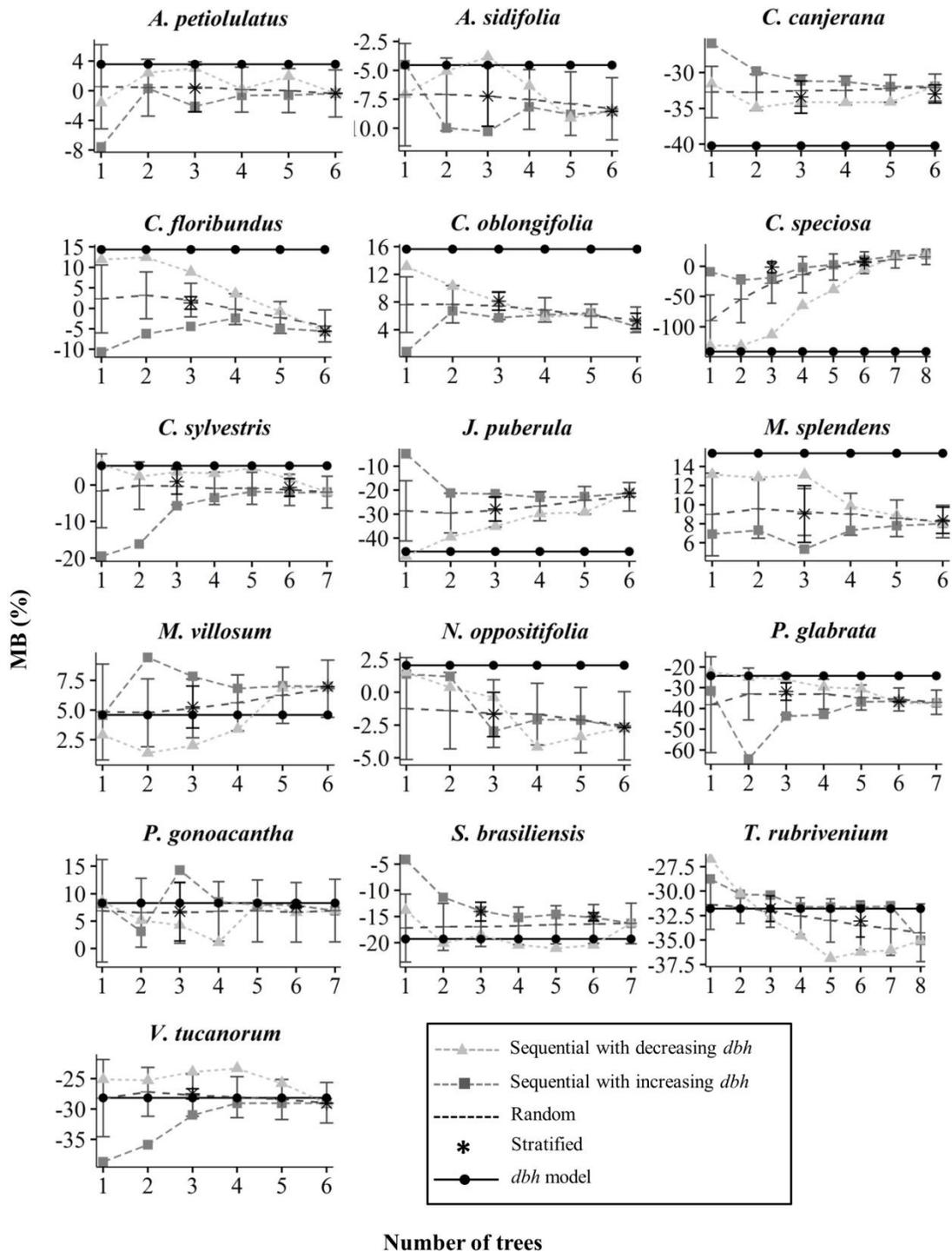


Figure 5. Percent mean bias (MB) for calibration of 16 species for prediction of individual tree aboveground biomass from Atlantic Forest in Brazil using different types of sample tree selection methods for linear mixed-effects (LME) and various sample sizes. The sample selection methods included random, stratified, and sequential (increasing and decreasing) compared to the generalized model with diameter at breast height (*dbh*). The error bars are \pm two standard deviations.

4.4. Discussion

4.4.1. Calibration methods' performance

The generalized models have been extensively used for tropical forests biomass estimation (Brown et al. 1989; Scatena et al. 1993; Overman et al. 1994; Chambers et al. 2001; Chave et al. 2001; Nogueira et al. 2008b). This approach is advantageous and practical in the field, where most commonly only *dbh* measurements are taken. However, biased estimates are obtained when models fitted for a specific site are used in others (Nogueira et al. 2008a, 2008b). This limitation has rarely been considered, and the accuracy on the prediction of the individual tree biomass and the error propagation on estimates at the stand- and forest-level are often ignored.

On the other side, species-specific models have higher accuracy (Nelson et al 1999; van Breugel et al. 2011). However, due to the high diversity in tropical forests, it is impracticable to develop species-specific models for all species. So, this approach seems to be a reasonable alternative only for the most abundant species, while local generalized models for least abundant species can be used. Nevertheless, it is expected that the variation on the most abundant species across the sites will occur at local and regional scales. Based on our results, if a new species in the most abundant species sampled in an inventory, the calibration methods can overcome this limitation.

This study evaluated model calibration for predicting aboveground biomass for new species using both LME and OLS as well as various sample sizes. It was believed that model calibration could help simplify the process for developing species-specific biomass prediction equations. As expected, a general improved model performance was observed using LME calibration, which is a result similar to that obtained by Temesgen et al. (2008) for individual tree height imputation. However, calibration was not efficient for all species examined in this analysis.

According to Vismara et al. (2016), a mixed-effect calibration is efficient only when true values of the random effects are not near zero. In other words, when the sample used in the calibration has a property similar to the population, no benefit of calibration will be observed as the new random effect does not differ from the population. In this case, the fixed effects may be more appropriate, which was observed in this analysis.

In this study, the most important trait for species that varied across them was *wsg*, and we believe that when a species has a *wsg* outside the range used to fit the model calibration can be effective. For instance, *C. speciosa* had the smallest *wsg* and calibration was substantially more effective than using the generalized models. The same was observed for *C. oblongifolia*, which was the species with highest *wsg*. However, extra sampled species with a wider range of *wsg* are needed to confirm this pattern.

4.4.2. Strategies for tree selection and the interference on LME calibration

For random sampling with LME calibration, except for *C. speciosa*, no significant improvement was observed by increasing the number of trees, which was different from prior studies on the topic (Temesgen et al. 2008; Crecente-Campo et al. 2014; Vismara et al. 2016). Effective model calibrations often need a smaller dataset in comparison to existing equations that are based on much larger datasets (de-Miguel et al. 2014). In this study, we believe that one to three trees was sufficient for improved model calibration. In prior studies, 4 to 10 trees were recommended for sampling (Temesgen et al. 2008; Crecente-Campo et al. 2014; Vismara et al. 2016). The lower sample sizes may reflect the effectiveness of using species as a random effect and the general strength of relationship between *dbh* and biomass.

A similar performance between stratified and random sampling with three or six trees was consistent with de-Miguel et al. (2014). However, other studies have found stratification or sequential sampling to be more effective than random sampling. For example Calama and Montero (2004) recommended a subsample of the four largest trees in a stand, while Dorado et al. (2006) suggested a subsample of the three smallest trees. According to Temesgen et al. (2008), model calibration using the largest tree was superior to using random selection. Although a better performance by using small trees for *C. oblongifolia* and *C. speciosa* was observed in our analysis (RMSE; Fig. 4), a general pattern was not present for the other fourteen species. Harvesting large trees is less practical than small ones and more time consuming since they are rarer (e.g., Poorter et al. 2008), and provoke more damage to others nearby trees when felled (e.g., Khai et al. 2017). The finding of using random selection also has practical implications for natural tropical forests, which are often characterized by uneven-aged and multi-specific stands. Although we are proposing model calibration for the most abundant species, occurrence of a species is often not uniform across sites (Webb and Peart 2000; Slik et al. 2003; Réjou-Méchain et al. 2008; Eisenlohr and Oliveira-Filho 2014) so model calibration with random sampling will likely need to be repeated for future data collection efforts.

We expect species-level calibration may provide improved accuracy in predictions and estimates for highly diverse tropical forest, where the variation in species composition is typical. We believe employing this methodology can substantially improve biomass quantification and provide a better understanding of the real role of tropical forests in the global carbon stock.

4.5. Conclusions

The results obtained in this study warrant the future use of model calibration using LME to develop more refined species-specific aboveground biomass models for diverse areas like the Atlantic Forest in Brazil using relatively small data collection efforts. We recommend the following steps in order to do that most effectively:

- Model calibration is most effective for species at either end of the *wsg* spectrum so it is useful to obtain previous information from the *wsg* in order to determine the new species for calibration. Otherwise, the use of the local generalized models is suggested.
- One to three trees are likely sufficient for calibration, and no stratification is needed, which greatly simplifies data collection efforts.
- LME is more effective than OLS for model calibration, so it is recommended for improved performance, particularly with reduced sample sizes.

References

- Alvares, C.A., Stape, J.L., Sentelhas, P.C., Gonçalves, J.L.M., and Sparovek, G. 2013. Köppen's climate classification map for Brazil. *Meteorol. Z.* **22**(6): 711-728. doi: 10.1127/0941-2948/2013/0507.
- Arias-Rodil, M., Castedo-Dorado, F., Cámara-Obregón, A., and Diéguez-Aranda, U. 2015. Fitting and calibrating a multilevel mixed-effects stem taper model for maritime pine in NW Spain. *PLoS ONE.* **10**(12): 1-20. doi: 10.1371/journal.pone.0143521.
- Brown, S., Gillespie, A.J.R., and Lugo, A.E. 1989. Biomass estimation methods for tropical forest with applications to forest inventory data. *For. Sci.* **35**(4): 881-902.

- Calama, R. and Montero, G. 2004. Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain. *Can. J. For. Res.* **34**(1): 150-163. doi: 10.1139/x03-199.
- Cepagri. 2017. Centro De Pesquisas Meteorológicas E Climáticas Aplicadas À Agricultura [online]. Available from <http://www.cepagri.unicamp.br/> [assessed 8 August 2017].
- Chambers, J.Q., Santos, J., Ribeiro, R.J., and Higuchi, N. 2001. Tree damage, allometric relationships, and above-ground net primary production in central Amazon Forest. *For. Ecol. Manage.* **152**: 73-84.
- Chave, J., Riera, B., and Dubois, M. 2001. Estimation of biomass in a Neotropical forest of French Guiana: spatial and temporal variability. *J. Trop. Ecol.* **17**: 79-96. DOI: 10.1017/S0266467401001055.
- Crecente-Campo, F., Corral-Rivas, J.J., Vargas-Larreta, B., and Wehenkel, C. 2014. Can random components explain differences in the height – diameter relationship in mixed uneven-aged stands? *Ann. For. Sci.* **71**:51-70. doi: 10.1007/s13595-013-0332-6.
- de-Miguel, S., Mehtätalo, L., and Durkaya, A. 2014. Developing generalized, calibratable, mixed-effects meta-models for large-scale biomass prediction. *Can. J. For. Res.* **44**: 648–656. doi: [dx.doi.org/10.1139/cjfr-2013-0385](https://doi.org/10.1139/cjfr-2013-0385).
- Dorado, F.C., Diéguez-Aranda, U., Anta, M.B., Rodríguez, M.S., and von Gadow, K. 2006. A generalized height–diameter model including random components for radiata pine plantations in northwestern Spain. *For. Ecol. Manage.* **229**(1): 202-213. doi: 10.1016/j.foreco.2006.04.028.
- Eisenlohr, P. V., and Oliveira-Filho, A.T. 2014. Tree species composition in areas of Atlantic Forest in southeastern Brazil is consistent with a new system for classifying the vegetation of South America. *Acta bot. bras.* **28**(2): 227-233. doi: 10.1590/S0102-33062014000200009.

- Fauset, et al. 2015. Hyperdominance in Amazonian forest carbon cycling. *Nature Communications*. doi: 10.1038/ncomms7857.
- Flora do Brasil. 2015. Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de Janeiro [online]. Available from <http://floradobrasil.jbrj.gov.br/> [accessed 20 June 2016].
- Khai, T.C., Mizoue, N., Kajisa, T., Ota, T., Yoshida, and S. 2017. Using a tree-based approach to evaluate logging damage in a tropical mixed deciduous forest of Myanmar: comparison with cases in Cambodia. *Jour. For. Res.* **22**(3): 185-190. doi: 10.1080/13416979.2017.1302545.
- Lam, T.Y., Kershaw, J.A., Hajar, Z.S.N., Rahman, K.A., Weiskittel, A.R., and Potts, M.D. 2016. Evaluating and modeling genus and species variation in height-to-diameter relationships for Tropical Hill Forests in Peninsular Malaysia. *Forestry*. **90**(2): 268-278. doi: 10.1093/forestry/cpw051.
- Lappi, J. 1991. Calibration of height and volume equations with random parameters. *For. Sci.* **37**: 781–801.
- Lewis, S.L., et al. 2009. Increasing carbon storage in intact African tropical forests. *Nature*. **457**(19): 1003-1007. doi:10.1038/nature07771.
- MacFarlane, W.D., and Weiskittel, A.R. 2016. A new method for capturing stem taper variation for trees of diverse morphological types. *Can. J. For. Res.* **46**: 804–815. doi: dx.doi.org/10.1139/cjfr-2016-0018.
- Mobot. 2016. Missouri Botanical Garden Plant Science [online]. Available from <http://www.tropicos.org> [accessed 20 June 2016].
- Nelson, B.W., Mesquita, R., Pereira, J.L.G., Souza, S.G.A., Batista, G.T. and Couto, L.B. 1999. Allometric regressions for improved estimate of secondary forest biomass in central Amazon. *For. Ecol. Manage.* **117**:149-167.

- Nogueira, E.M., Nelson, B.W., Fearnside, F.M., França, M.B., and Oliveira, A.C.A. 2008a. Tree height in Brazil's 'arc of deforestation': Shorter trees in south and southwest Amazonia imply lower biomass. *For. Ecol. Manage.* **255**(7): 2963–2972. doi: 10.1016/j.foreco.2008.02.002.
- Nogueira, E.M., Fearnside, P.M., Nelson, B.W., Barbosa, R.L., and Keizer, E.W.H. 2008b. Estimates of forest biomass in the Brazilian Amazon: New allometric equations and adjustments to biomass from wood-volume inventories. *For. Ecol. Manage.* **256**: 1853–1867. doi: 10.1016/j.foreco.2008.07.022.
- Overman, J.P., Witte, H.J., and Saldarriaga, J.G. 1994. Evaluation of regression models for aboveground biomass determination in Amazon rainforest. *J. Trop. Ecol.* **10**(2): 207-218. Doi: 10.1017/S0266467400007859.
- Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S.W., Mcguire, A.D., Piao, S., Rautiainen, A., Sitch, S., and Hayes, D. 2011. A large and persistent carbon sink in the world's forests. *Science.* **333**: 988-993. doi: 10.1126/science.1201609.
- Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D., R Core Team. 2016. nlme: Linear and nonlinear mixed effects models. R Package Version 3.1-125. <https://cran.r-project.org/web/packages/nlme/nlme.pdf>. (accessed 12.12.16).
- Poorter, L., Hawthorne, W., Bongers, F., and Sheil, D. 2008. Maximum size distributions in tropical forest communities: relationships with rainfall and disturbance. *J. Ecol.* **96**: 495-504. doi: 10.1111/j.1365-2745.2008.01366.x.
- R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org/>.

- Réjou-Méchain, M, Péliisser,R., Gourlet-Fleury, S., Couteron, P., Nasi, R., and Thompson, J.D. 2008. Regional variation in tropical forest tree species composition in the Central African Republic: an assessment based on inventories by forest companies. *J. Trop. Ecol.* **24**:663–674. doi: 10.1017/S0266467408005506.
- Robinson, G.K. 1991. That BLUP is a Good Thing: The Estimation of Random Effects *Stat. Sci.* **6**(1): 15-32.
- Russell, M.B, Weiskittel, A.R., and Kershaw, J.A. 2014. Comparing strategies for modeling individual-tree height and height-to-crown base increment in mixed-species Acadian forests of northeastern North America. *Eur. J. For. Res.* **133**: 1121-1135. doi: 10.1007/s10342-014-0827-1.
- Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M., and Morel, A. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *PNAS.* **108** (24): 9899–9904. doi: 10.1073/pnas.1019576108.
- Scatena, F.N., Silver, W., Siccama, T., Johnson, A., and Sánchez, M.J. 1993. Biomass and nutrient content of the Bisley Experimental Watersheds, Luquillo Experimental Forest, Puerto Rico, before and after Hurricane Hugo, 1989. *Biotropica.* **25**(1): 15-27. Doi: 10.2307/2388975.
- Scolforo, H.F., Scolforo, J.R.S., Mello, C.R., Mello, J.M., and Ferraz Filho, A.C. 2015. Spatial distribution of aboveground carbon stock of the arboreal vegetation in Brazilian biomes of Savanna, Atlantic Forest and Semi-Arid Woodland. *PLoS ONE.* **10** (6). doi: 10.1371/journal.pone.0128781.
- Scolforo, H.F., Scolforo, J.R.S., Mello, J.M., Mello, C.R., and Morais, V.A. 2016. Spatial interpolators for improving the mapping of the carbon stock of the arboreal vegetation in

Brazilian biomes of Atlantic forest and Savanna. *For. Ecol. Manage.* **376**: 24–35. doi: 10.1016/j.foreco.2016.05.047.

Scolforo, H.F., Scolforo, J.R.S., Thiersch, C.R., Thiersch, M.F., McTague, J.P., Burkhardt, H.E., Ferraz-Filho, A.C., Mello, J.M., and Roise, J. 2017. A new model of tropical tree diameter growth rate and its application to identify fast-growing native tree species. *For. Ecol. Manage.* **400**: 578-586. doi: 10.1016/j.foreco.2017.06.048.

Scolforo, H.F., Raimundo, M.R., McTague, J.P., Weiskittel, A.R., Scolforo, J.R.S., and Carrero, O. 2018. Comparison of taper functions applied to eucalypts of varying genetics in Brazil: Application and evaluation of the penalized mixed spline approach. *Can. J. For. Res.* doi: 10.1139/cjfr-2017-0366.

Slik, J.W.F., Poulsen, A.D., Ashton, P.S., Cannon, C.H., Eichhorn, K.A.O., Kartawinata, K., Lanniari, I., Nagamasu, H., Nakagawa, M., van Nieuwstadt, M.G.L., Payne, J., Saridan, A., Sidiyasa, K., Verburg, R.W., Webb, C.O., and Wilkie, P. 2003. A floristic analysis of the lowland dipterocarp forests of Borneo. *J Biogeogr.* **30**: 1517–1531. doi: 10.1046/j.1365-2699.2003.00967.x.

Soil Survey Staff., 2014. *Keys to Soil Taxonomy*. 12th ed. Washington, DC: US Department of Agriculture–Natural Resources Conservation Service, US. pp. 1-372.

Sotomayor, J.F.M. 2013 *Determinação de biomassa aérea em florestas nativas num ambiente agrícola do Estado de São Paulo*. Ph.D. thesis, Department of Forest Science, University of São Paulo, Piracicaba, SP.

Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations. *Ecology*. **64**(1): 209-210.

- Temesgen, H., Affleck, D., Poudel, K., Gray, A., and Sessions, J. 2015. A review of the challenges and opportunities in estimating above ground forest biomass using tree-level models. *Scand. J. For. Res.* **30**(4): 326-335. doi:10.1080/02827581.2015.1012114.
- Temesgen, H., Monleon, V.J. and Hann, D.W. 2008. Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. *Can. J. For. Res.* **38**: 553-565. doi: 10.1139/X07-104.
- van Breugel, M., Ransijn, J., Craven, D., Bongers, F., and Hall, J.S., 2011. Estimating carbon stock in secondary forests: Decisions and uncertainties associated with allometric biomass models. *For. Ecol. Manage.* **262**: 1648-1657. doi: 10.1016/j.foreco.2011.07.018.
- Vicent, J.B., Henning, B., Saulei, S., Sosanika, G., and Weiblen, G.D. 2015. Forest carbon in lowland Papua New Guinea: Local variation and the importance of small trees. *Austral Ecol.* **40**: 151–159. doi:10.1111/aec.12187.
- Vieira, S.A., Alves, L.F., Duarte-Neto, P.J., Martins, S.C., Veiga, L.G., Scaranello, M.A., Picollo, M.C., Camargo, P.B., Carmo, J.B., Sousa Neto, E., Santos, F.A.M., Joly, C.A., and Martinelli, L.A. 2011. Stocks of carbon and nitrogen and partitioning between above- and belowground pools in the Brazilian coastal Atlantic Forest elevation range. *Ecol. Evol.* doi: 10.1002/ece3.41.
- Vismara, E.S. 2013. Avaliação da construção e aplicação de modelos florestais de efeito fixo e efeitos mistos sob o ponto de vista preditivo. Ph.D. thesis, Department of Forest Science, University of São Paulo, Piracicaba, SP.
- Vismara, E.S., Mehtätalo, L., and Batista, J.L.F. 2016. Linear mixed-effects models and calibration applied to volume models in two rotations of *Eucalyptus grandis* plantations. *Can. J. For. Res.* **46**: 132–141. doi: dx.doi.org/10.1139/cjfr-2014-0435.
- Webb, C.O., and Peart, D.R. 2000. Habitat associations of trees and seedlings in a Bornean rain forest. *J. Ecol.* **88**: 464-478. doi: 10.1046/j.1365-2745.2000.00462.x.

5. FINAL REMARKS

The species-specific models using linear mixed-models (LME) had the better performance, so this approach is recommended for individual tree-level prediction. However, the approach may be suitable only for the most abundant species due to the huge amount of species in tropical forests. In this study, it was arbitrarily considered abundant species those that represented about 70% of the total of stems, but this number may vary according to the forest types.

For species with no specific parameters, it is proposed the calibration of previously existing models using LME. Based on the results of this study, no stratification was required, and the diameter at breast height (*dbh*) size did not interfere the calibration performance. However, the calibration is not efficient in all cases. Model calibration is most effective for species at either end of the woody specific gravity (*wsg*) spectrum, so it is useful to obtain previous information from the *wsg* to determine the new species for calibration. It is believed that one to three trees are sufficient for an efficient calibration for this case.

The existing models had lower performance in individual tree-level predictions. Consequently, biased stand- and forest-level estimates were obtained. The Pan-tropical model had positively biased estimates leading to stand- and forest-level biomass overestimation of over 40%. Additionally, the IPCC carbon conversion factor value of 0.47 is higher than the value proposed in this study of 0.45, with associated to Pan-tropical model overestimated over 50% the carbon in stand-level. It is expected that the models proposed in this study contribute to more accurate estimates of the carbon stocked in Atlantic Forest biomass, and consequently improvement in the understanding of the real role of the biome in the global carbon cycle.

It is believed the using species-specific models for abundant species, and a consequent calibration for species with no specific models, and local models which include *dbh*, *wsg* and plant

traits for less abundant species, is an effective strategy for consistent biomass estimation in highly diverse forests, such as the Atlantic Forest. It is suggested for future analysis to investigate others potential covariates, such as other plant traits, may provide an improvement in models performance.

APPENDIX

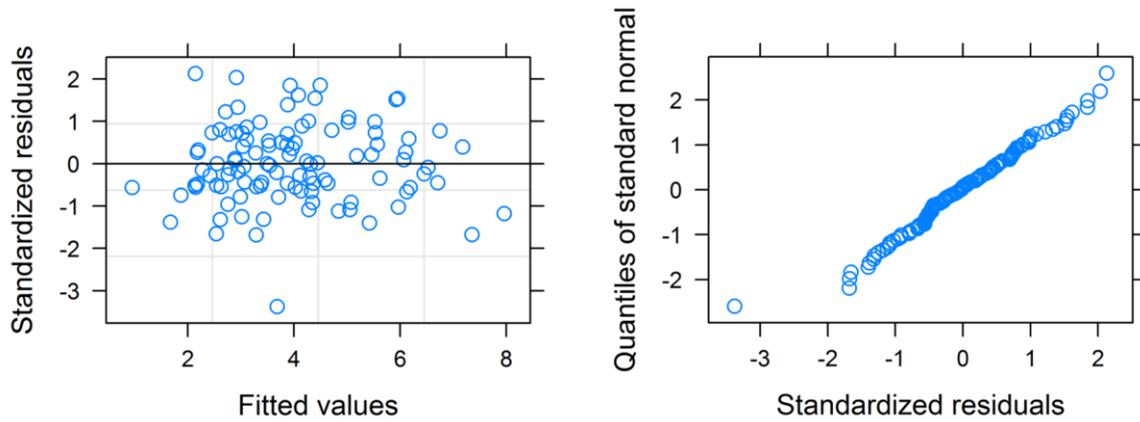
APPENDIX A. Species from the studied site in Atlantic Forest at Serra da Cantareira-SP/Brazil.

Family	Genus	Specie
Annonaceae	<i>Annona</i>	<i>Annona cacans</i> Warm.
Annonaceae	<i>Annona</i>	<i>Annona neosericea</i> H.Rainer
Annonaceae	<i>Annona</i>	<i>Annona sylvatica</i> A.St.-Hil.
Asteraceae	<i>Gochnatia</i>	<i>Gochnatia polymorpha</i> (Less.) Cabrera
Asteraceae	<i>Piptocarpha</i>	<i>Piptocarpha axillaris</i> (Less.) Baker
Asteraceae	<i>Piptocarpha</i>	<i>Piptocarpha macropoda</i> (DC.) Baker
Bignoniaceae	<i>Jacaranda</i>	<i>Jacaranda puberula</i> Cham.
Boraginaceae	<i>Cordia</i>	<i>Cordia sellowiana</i> Cham.
Canellaceae	<i>Cinnamodendr</i>	<i>Cinnamodendron</i> sp.
Cannabaceae	<i>Trema</i>	<i>Trema micrantha</i> (L.) Blume
Celastraceae	<i>Maytenus</i>	<i>Maytenus evonymoides</i> Reissek
Clethraceae	<i>Clethra</i>	<i>Clethra scabra</i> Pers.
Clethraceae	<i>Clethra</i>	<i>Cletra</i> sp.
Cunoniaceae	<i>Lamanonia</i>	<i>Lamanonia ternata</i> Vell.
Elaeocarpaceae	<i>Sloanea</i>	<i>Sloanea guianensis</i> (Aubl.) Benth.
Euphorbiaceae	<i>Alchornea</i>	<i>Alchornea glandulosa</i> Poepp. & Endl.
Euphorbiaceae	<i>Alchornea</i>	<i>Alchornea sidifolia</i> Müll. Arg.
Euphorbiaceae	<i>Alchornea</i>	<i>Alchornea triplinervia</i> (Spreng.) Müll.Arg.
Euphorbiaceae	<i>Croton</i>	<i>Croton floribundus</i> Spreng.
Euphorbiaceae	<i>Pera</i>	<i>Pera glabrata</i> (Schott) Poepp. ex Baill.
Euphorbiaceae	<i>Sapium</i>	<i>Sapium glandulosum</i> (L.) Morong.
Euphorbiaceae	<i>Tetrorchidiu</i>	<i>Tetrorchidium rubrivenium</i> Poepp.
Fabaceae	<i>Andira</i>	<i>Andira anthelmia</i> (Vell.) Benth.
Fabaceae	<i>Dalbergia</i>	<i>Dalbergia brasiliensis</i> Vogel
Fabaceae	<i>Inga</i>	<i>Inga sessilis</i> (Vell.) Mart.
Fabaceae	<i>Machaerium</i>	<i>Machaerium nyctitans</i> (Vell.) Benth.
Fabaceae	<i>Machaerium</i>	<i>Machaerium stipitatum</i> Vogel
Fabaceae	<i>Machaerium</i>	<i>Machaerium villosum</i> Vogel
Fabaceae	<i>Piptadenia</i>	<i>Piptadenia gonoacantha</i> (Mart.) J. F. Macbr.
Fabaceae	<i>Piptadenia</i>	<i>Piptadenia paniculata</i> Benth.
Fabaceae	<i>Platymiscium</i>	<i>Platymiscium floribundum</i> Vogel
Fabaceae	<i>Schizolobium</i>	<i>Schizolobium parahyba</i> (Vell.) Blake
Fabaceae	<i>Senna</i>	<i>Senna macranthera</i> (DC. ex Collad.) H. S.
Fabaceae	<i>Senna</i>	<i>Senna multijuga</i> (Rich.) H. S. Irwin & Barneb
Hypericaceae	<i>Vismia</i>	<i>Vismia brasiliensis</i> Choisy
Hypericaceae	<i>Vismia</i>	<i>Vismia guianensis</i> (Aubl.) Choisy
Lauraceae	<i>Endlicheria</i>	<i>Endlicheria paniculata</i> (Spreng.) J. F. Macbr

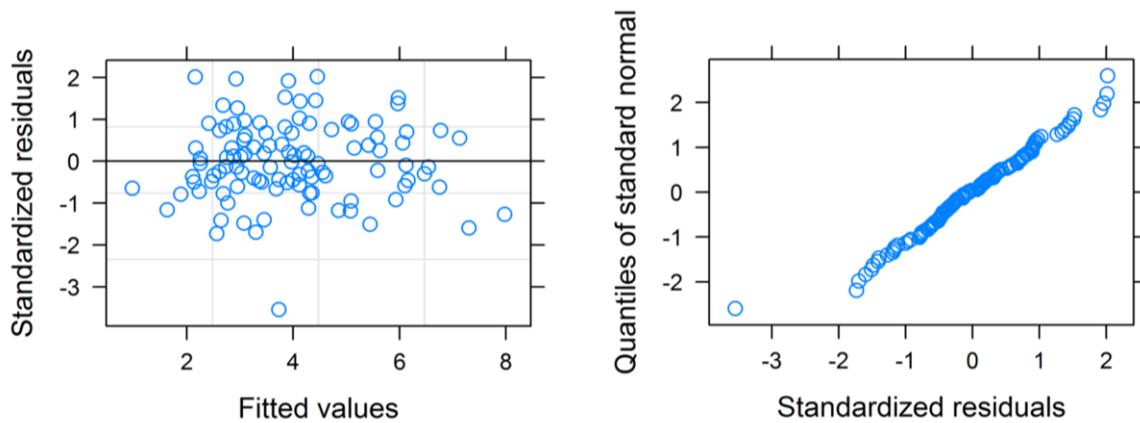
Lauraceae	<i>Nectandra</i>	<i>Nectandra barbellata</i> Coe-Teix.
Lauraceae	<i>Nectandra</i>	<i>Nectandra membranacea</i> (Sw.) Griseb.
Lauraceae	<i>Nectandra</i>	<i>Nectandra oppositifolia</i> Ness
Lauraceae	<i>Ocotea</i>	<i>Ocotea catharinensis</i> Mez
Lauraceae	<i>Ocotea</i>	<i>Ocotea pulchella</i> (Nees & Mart.) Mez
Malvaceae	<i>Ceiba</i>	<i>Ceiba speciosa</i> (A. St.-Hil.) Ravenna
Malvaceae	<i>Luehea</i>	<i>Luehea grandiflora</i> Mart. & Zucc.
Melastomatace	<i>Miconia</i>	<i>Miconia budlejoides</i> Triana
Melastomatace	<i>Miconia</i>	<i>Miconia cabucu</i> Hoehne
Melastomatace	<i>Miconia</i>	<i>Miconia cinnamomifolia</i> (DC.) Naudin
Melastomatace	<i>Miconia</i>	<i>Miconia inconspicua</i> Miq.
Melastomatace	<i>Miconia</i>	<i>Miconia sellowiana</i> Naudin
Meliaceae	<i>Cabrlea</i>	<i>Cabrlea canjerana</i> (Vell.) Mart.
Meliaceae	<i>Cedrela</i>	<i>Cedrela fissilis</i> Vell.
Meliaceae	<i>Cedrela</i>	<i>Cedrela odorata</i> L.
Meliaceae	<i>Guarea</i>	<i>Guarea macrophylla</i> Vahl
Monimiaceae	<i>Mollinedia</i>	<i>Mollinedia schottiana</i> (Spreng.) Perkins
Moraceae	<i>Ficus</i>	<i>Ficus insipida</i> Willd.
Myrtaceae	<i>Campomanesia</i>	<i>Campomanesia guaviroba</i> (DC.) Kiaersk.
Myrtaceae	<i>Campomanesia</i>	<i>Campomanesia neriiflora</i> (O. Berg) Nied.
Myrtaceae	<i>Campomanesia</i>	<i>Campomanesia xanthocarpa</i> (Mart.) O. Berg
Myrtaceae	<i>Campomanesia</i>	<i>Campomanesia guazumifolia</i> (Cambess.) O. Berg
Myrtaceae	<i>Eugenia</i>	<i>Eugenia involucrata</i> DC.
Myrtaceae	<i>Eugenia</i>	<i>Eugenia</i> sp.
Myrtaceae	<i>Myrceugenia</i>	<i>Myrceugenia myrcioides</i> (Cambess.) O. Berg
Myrtaceae	<i>Myrcia</i>	<i>Myrcia hebeptala</i> DC.
Myrtaceae	<i>Myrcia</i>	<i>Myrcia multiflora</i> (Lam.) DC.
Myrtaceae	<i>Myrcia</i>	<i>Myrcia splendens</i> (Sw.) DC.
Myrtaceae	<i>Myrcia</i>	<i>Myrcia tomentosa</i> (Aubl.) DC.
Myrtaceae	<i>Myrciaria</i>	<i>Myrciaria floribunda</i> (H. West. ex Willd.) O. B.
Nyctaginaceae	<i>Pisonia</i>	<i>Pisonia ambigua</i> Heimerl
Olacaceae	<i>Heisteria</i>	<i>Heisteria silvianii</i> Schwacke
Phyllanthaceae	<i>Hieronyma</i>	<i>Hieronyma alchorneoides</i> Allemão
Primulaceae	<i>Myrsine</i>	<i>Myrsine coriacea</i> (Sw.) R. Br. ex Roem. & Schu
Primulaceae	<i>Myrsine</i>	<i>Myrsine umbellata</i> Mart.
Rubiaceae	<i>Psychotria</i>	<i>Psychotria suterella</i> Müll. Arg.
Rutaceae	<i>Dictyoloma</i>	<i>Dictyoloma vandellianum</i> A. Juss.
Rutaceae	<i>Zanthoxylum</i>	<i>Zanthoxylum rhoifolium</i> Lam.
Salicaceae	<i>Casearia</i>	<i>Casearia obliqua</i> Spreng.
Salicaceae	<i>Casearia</i>	<i>Casearia sylvestris</i> Sw.

Sapindaceae	<i>Allophylus</i>	<i>Allophylus petiolulatus</i> Radlk.
Sapindaceae	<i>Cupania</i>	<i>Cupania emarginata</i> Cambess.
Sapindaceae	<i>Cupania</i>	<i>Cupania oblongifolia</i> Mart.
Sapindaceae	<i>Matayba</i>	<i>Matayba elaeagnoides</i> Radlk.
Solanaceae	<i>Sessea</i>	<i>Sessea brasiliensis</i> Toledo
Solanaceae	<i>Solanum</i>	<i>Solanum bullatum</i> Vell.
Solanaceae	<i>Solanum</i>	<i>Solanum pseudoquina</i> A. St.-Hil.
Symplocaceae	<i>Symplocos</i>	<i>Symplocos</i> sp.
Urticaceae	<i>Cecropia</i>	<i>Cecropia glaziovii</i> Snethl.
Urticaceae	<i>Cecropia</i>	<i>Cecropia pachystachya</i> Trécul
Urticaceae	<i>Urera</i>	<i>Urera baccifera</i> (L.) Gaudich. ex Wedd.
Verbenaceae	<i>Citharexylum</i>	<i>Citharexylum myrianthum</i> Cham.
Vochysiaceae	<i>Vochysia</i>	<i>Vochysia tucanorum</i> Mart.

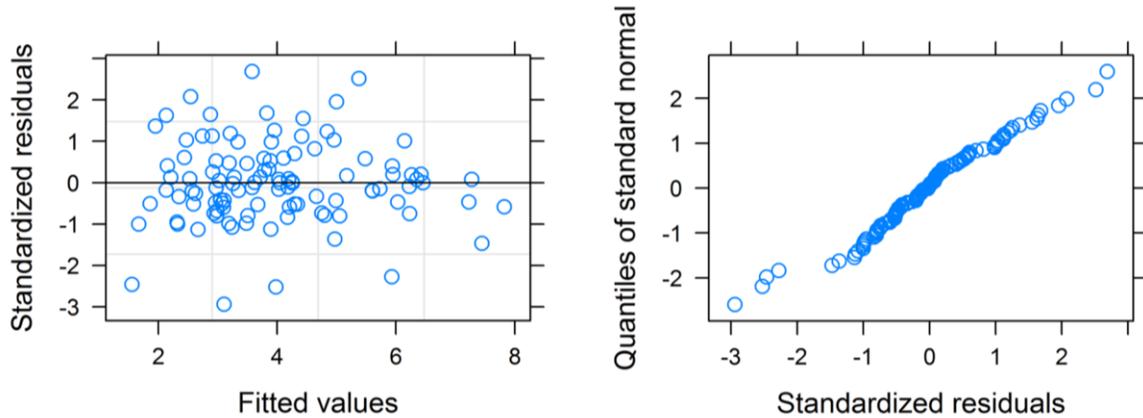
APPENDIX B. Scatter plot of standardized residuals versus fitted values, and normal probability plot of the standardized residuals for fitted models.



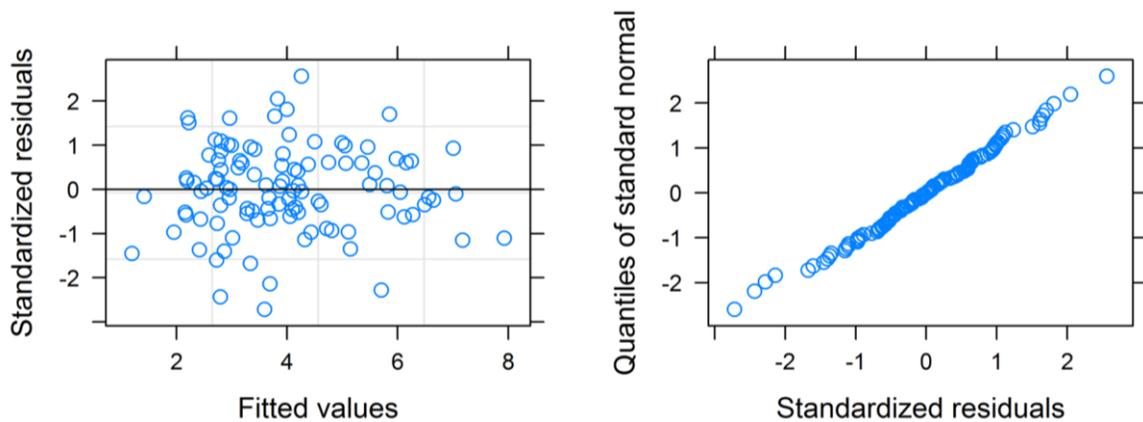
APPENDIX B.1 On the right is scatter plot of standardized residuals versus fitted values, and on the left is a normal probability plot of the standardized residuals for diameter at breast height (*dbh*) only model fitted by linear mixed-effect (LME).



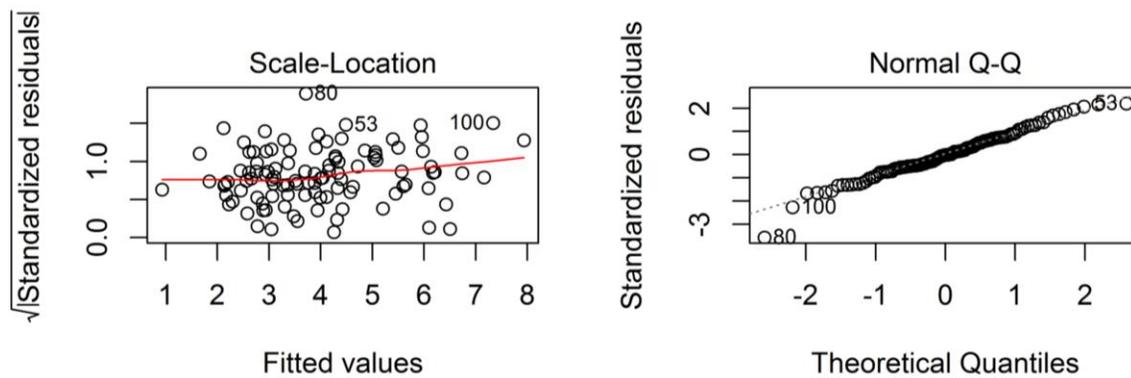
APPENDIX B.2 On the right is scatter plot of standardized residuals versus fitted values, and on the left is a normal probability plot of the standardized residuals for diameter at breast height (*dbh*) and woody specific gravity (*wsg*) model fitted by linear mixed-effect (LME).



APPENDIX B.3 On the right is scatter plot of standardized residuals versus fitted values, and on the left is a normal probability plot of the standardized residuals for functional plant trait based on diameter at breast height (*dbh*) model fitted by linear mixed-effect (LME).



APPENDIX B.4 On the right is scatter plot of standardized residuals versus fitted values, and on the left is a normal probability plot of the standardized residuals for functional plant trait based on diameter at breast height (*dbh*) and woody specific gravity (*wsg*) model fitted by linear mixed-effect (LME).



APPENDIX B.5 On the right is scatter plot of standardized residuals versus fitted values, and on the left is a normal probability plot of the standardized residuals for diameter at breast height (*dbh*) model fitted by ordinary least square (OLS).

APPENDIX C. Species in study site was the Cantareira State Park (Parque Estadual da Cantareira, in portuguese).

Family	Genus	Specie
Annonaceae	<i>Annona</i>	<i>Annona neosericea</i> H.Rainer
Annonaceae	<i>Xylopia</i>	<i>Xylopia brasiliensis</i> Spreng
Apocynaceae	<i>Aspidosperma</i>	<i>Aspidosperma</i> sp.
Aquifoliaceae	<i>Ilex</i>	<i>Ilex</i> sp.
Araliaceae	<i>Dendropanax</i>	<i>Dendropanax cuneatus</i> (DC.) Decne. & Planch.
Araliaceae	<i>Schefflera</i>	<i>Schefflera morototoni</i> (Aubl.) Maguire et al.
Asteraceae	<i>Piptocarpha</i>	<i>Piptocarpha</i> sp.
Boraginaceae	<i>Cordia</i>	<i>Cordia sellowiana</i> Cham.
Celastraceae	<i>Maytenus</i>	<i>Maytenus</i> sp.
Clethraceae	<i>Clethra</i>	<i>Clethra scabra</i> Pers.
Clusiaceae	<i>Clusia</i>	<i>Clusia</i> sp.
Clusiaceae	<i>Garcinia</i>	<i>Garcinia gardneriana</i> (Planch. & Triana) Zappi
Clusiaceae	<i>Tovomitopsis</i>	<i>Tovomitopsis paniculata</i> (Spreng.) Planch. & Triana
Combretaceae	<i>Terminalia</i>	<i>Terminalia</i> sp.
Euphorbiaceae	<i>Alchornea</i>	<i>Alchornea sidifolia</i> Müll.Arg.
Euphorbiaceae	<i>Alchornea</i>	<i>Alchornea</i> sp.
Euphorbiaceae	<i>Alchornea</i>	<i>Alchornea triplinervia</i> (Spreng.) Müll.Arg.
Euphorbiaceae	<i>Aparisthium</i>	<i>Aparisthium cordatum</i> (A.Juss.) Baill.
Euphorbiaceae	<i>Croton</i>	<i>Croton</i> sp.
Euphorbiaceae	<i>Sapium</i>	<i>Sapium</i> sp.
Euphorbiaceae	<i>Tetrorchidium</i>	<i>Tetrorchidium</i> sp.
Fabaceae	<i>Hymenaea</i>	<i>Hymenaea</i> sp.
Fabaceae	<i>Machaerium</i>	<i>Machaerium</i> sp.
Fabaceae	<i>Machaerium</i>	<i>Machaerium villosum</i> Vogel
Fabaceae	<i>Ormosia</i>	<i>Ormosia</i> sp.
Fabaceae	<i>Piptadenia</i>	<i>Piptadenia gonoacantha</i> (Mart.) J.F.Macbr.
Fabaceae	<i>Pterodon</i>	<i>Pterodon emarginatus</i> Vogel
Hypericaceae	<i>Vismia</i>	<i>Vismia</i> sp.
Lauraceae	<i>Cryptocarya</i>	<i>Cryptocarya</i> sp.
Lauraceae	<i>Nectandra</i>	<i>Nectandra megapotamica</i> (Spreng.) Mez
Lauraceae	<i>Ocotea</i>	<i>Ocotea</i> sp.
Lauraceae	<i>Persea</i>	<i>Persea</i> sp.
Lecythidaceae	<i>Cariniana</i>	<i>Cariniana</i> sp.
Malvaceae	<i>Luehea</i>	<i>Luehea</i> sp.
Melastomataceae	<i>Miconia</i>	<i>Miconia cinnamomifolia</i> (DC.) Naudin
Melastomataceae	<i>Mouriri</i>	<i>Mouriri</i> sp.
Meliaceae	<i>Cabrlea</i>	<i>Cabrlea canjerana</i> (Vell.) Mart.
Meliaceae	<i>Guarea</i>	<i>Guarea macrophylla</i> subsp. <i>tuberculata</i> (Vell.) T.D.Penn.

Moraceae	<i>Ficus</i>	<i>Ficus</i> sp.
Moraceae	<i>Sorocea</i>	<i>Sorocea bonplandii</i> (Baill.) W.C.Burger et al.
Myristicaceae	<i>Virola</i>	<i>Virola bicuhyba</i> (Schott ex Spreng.) Warb.
Nyctaginaceae	<i>Guapira</i>	<i>Guapira opposita</i> (Vell.) Reitz
Peraceae	<i>Pera</i>	<i>Pera glabrata</i> (Schott) Poepp. ex Baill.
Phytolaccaceae	<i>Seguiera</i>	<i>Seguiera</i> sp.
Pittosporaceae	<i>Pittosporum</i>	<i>Pittosporum undulatum</i> Vent.
Primulaceae	<i>Myrsine</i>	<i>Myrsine</i> sp.
Proteaceae	<i>Roupala</i>	<i>Roupala</i> sp.
Rubiaceae	<i>Bathysa</i>	<i>Bathysa australis</i> (A.St.-Hil.) K.Schum.
Salicaceae	<i>Casearia</i>	<i>Casearia</i> sp.
Sapindaceae	<i>Allophylus</i>	<i>Allophylus</i> sp.
Sapindaceae	<i>Cupania</i>	<i>Cupania oblongifolia</i> Mart.
Sapindaceae	<i>Cupania</i>	<i>Cupania</i> sp.
Sapindaceae	<i>Matayba</i>	<i>Matayba juglandifolia</i> (Cambess.) Radlk.
Sapindaceae	<i>Matayba</i>	<i>Matayba</i> sp.
Sapotaceae	<i>Ecclinusa</i>	<i>Ecclinusa</i> sp.
Styracaceae	<i>Styrax</i>	<i>Styrax</i> sp.
Urticaceae	<i>Cecropia</i>	<i>Cecropia hololeuca</i> Miq.
Vochysiaceae	<i>Vochysia</i>	<i>Vochysia</i> sp.

APPENDIX D. This appendix shows the information necessary to calibrate the fixed-effects portion of a species-specific biomass model (eq. 13 in Colmanetti et al. *In review*), for a new species in the Atlantic Forest of Brazil. The variance-covariance matrix (Table A1) is required for the calibration procedure. This model varies on Intercept from Colmanetti et al. (*In review*) since it is not corrected according to Sprugel (1983). The variance of model is 0.02323.

$$D.1 \quad agb = \exp(-2.333 + 2.388 \cdot \ln(dbh))$$

Table A1. Variance-Covariance Matrix for linear mixed-effect model fitted for all sixteen species from Atlantic forest at Serra da Cantareira-SP/Brazil.

	Intercept	<i>dbh</i>
Intercept	0.11413	0
<i>dbh</i>	0	$7.8581 \cdot e^{-7}$

Appendix E

A numerical example for prediction of random effects. Suppose we intend to predict a new random effect for a new specie sample: *C. speciosa*. A previous mixed-model with species as a random effect using the remains 15 species are:

$$E.1 \quad \ln(a|ij) = (2.178 + b_{0i}) + (2.356 + b_{1i}) \cdot \ln(dbh_{ij}) + \epsilon_{ij}$$

where: a is the aboveground for specie i and tree j ; dbh is the diameter at breast height for specie i and tree j ; b_{0i} is the random effect on intercept for specie i ; b_{1i} is the random effect on slope for specie i .

Let us assume three trees with dbh in $\begin{bmatrix} 6.00 \\ 25.85 \\ 40.65 \end{bmatrix}$ cm and aboveground biomasses $\begin{bmatrix} 2.3 \\ 85.77 \\ 342.42 \end{bmatrix}$ kg were randomly felled. All procedures for

biomass determination were done according to section 2.2 and 2.3. According to eq.10, the BLUP can be obtained:

$$\begin{aligned} & \left(\begin{bmatrix} 1 & 6.00 \\ 1 & 25.85 \\ 1 & 40.65 \end{bmatrix}^t \cdot \begin{bmatrix} 0.0476 & 0 & 0 \\ 0 & 0.0476 & 0 \\ 0 & 0 & 0.0476 \end{bmatrix} \cdot \begin{bmatrix} 1 & 6.00 \\ 1 & 25.85 \\ 1 & 40.65 \end{bmatrix} \cdot \begin{bmatrix} 0.03318 & 0 \\ 0 & 2.3 e^{-13} \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 1 & 6.00 \\ 1 & 25.85 \\ 1 & 40.65 \end{bmatrix}^t \cdot \begin{bmatrix} 0.0476 & 0 & 0 \\ 0 & 0.0476 & 0 \\ 0 & 0 & 0.0476 \end{bmatrix} \cdot \begin{bmatrix} \log(2.3) - 2.053 \\ \log(85.77) - 5.479 \\ \log(342.42) - 6.541 \end{bmatrix} \\ & = \begin{bmatrix} -0.6998 \\ -0.0050 \end{bmatrix} \end{aligned}$$

The E.1 can be calibrated for *C. speciosa* as:

$$E.2 \quad \ln(a|j) = (2.178 - 0.6998) + (2.356 - 0.0050) \cdot \ln(dbh_j) + \epsilon_j$$

Predicted values must be back-transformed to original scale using the correction proposed by Sprugel (1983).

Appendix F

A numerical example for ordinary least square (OLS) calibrating method. The same three trees used in Appendix E were used for the OLS calibrating:

$$F1. \quad k^* = \frac{(2.053 \cdot \log(2.3)) + (5.479 \cdot \log(85.77)) + (6.541 \cdot \log(342.42))}{(2.053)^2 + (5.479)^2 + (6.541)^2} = 0.835$$

The fixed-effect from Appendix D. (eq. D.1) can be calibrated by OLS for *C. speciosa* as:

$$F2. \quad \ln(a|j) = 0.835 \cdot (2.178 + 2.356 \cdot \ln(dbh_j)) + \epsilon_j$$

Predicted values must be back-transformed to original scale using the correction proposed by Sprugel (1983).

Appendix reference list

Colmanetti, M.A.C., Weiskittel, A., Barbosa, L.M., Shirasuna, R.T., de Lima, F.C., Ortiz, P.R.T., Catharino, E.L.M., Barbosa, T.C., and do Couto, H.T.Z. 2018. Aboveground biomass and carbon of the highly diverse Atlantic Forest in Brazil: Comparison of alternative individual tree modeling and prediction strategies. *Carbon Management In review*.