AN ABSTRACT OF THE DISSERTATION OF

Fabio Guimarães Gonçalves for the degree of Doctor of Philosophy in Forest Science presented on November 21, 2014.

Title: Vertical Structure and Aboveground Biomass of Tropical Forests from Lidar Remote Sensing

Abstract approved:

_____________________________________________________________________

Beverly E. Law

Methods for obtaining accurate, spatially explicit estimates of biomass density in tropical forests are required to reduce uncertainties in the global carbon cycle, and to support international climate agreements and emerging carbon markets. Three-dimensional (3-D) remote sensing techniques sensitive to the vertical structure of vegetation provide a unique opportunity for mapping and monitoring forest carbon stocks across large areas in the tropics. However, approaches to forest biomass estimation from remotely sensed structure have yet to be fully developed to deliver the required biomass accuracy. In this research, we use airborne laser scanning (ALS), space-based lidar observations (ICESat/GLAS), and detailed in situ measurements made at the Tapajós National Forest, Brazil, to advance methods of biomass estimation in tropical regions from remotely sensed structure. The overall objectives were to (1) test and refine methods for the extraction of structural information from lidar data; (2) determine the accuracy of lidar estimates of structure in relation to detailed field measurements of vertical structure; and (3) develop and validate structure-based models for optimal prediction of aboveground biomass. Because remote sensing approaches to biomass estimation begin with field estimates of biomass, field plot data collected at Tapajós were also used to gain a better understanding of the uncertainty associated with plot-level biomass estimates.
obtained specifically for calibration of remote sensing data. This included an evaluation of the error resulting from spatial disagreement between field and remote sensing measurements (i.e., co-location error), and the error introduced when accounting for temporal differences in data acquisition. Results show that a new approach to biomass estimation based on Fourier transforms of lidar profiles significantly improves predictions of aboveground biomass ranging from 2 to 538 Mg ha\(^{-1}\) in primary and secondary forests. Data from two different regions in the Amazon were used to demonstrate and test this method in a range of conditions. The improvement in biomass estimation performance was consistent across sites and the approach was integrated in a multi-stage scaling strategy to biomass estimation to produce a wall-to-wall map of biomass across a large area in the Amazon.
Vertical Structure and Aboveground Biomass of Tropical Forests from Lidar Remote Sensing

by
Fabio Guimarães Gonçalves

A DISSERTATION

submitted to
Oregon State University

in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Presented November 21, 2014
Commencement June 2015
Doctor of Philosophy dissertation of Fabio Guimarães Gonçalves presented on November 21, 2014

APPROVED:

________________________________________________________________________

Major Professor, representing Forest Science

________________________________________________________________________

Head of the Department of Forest Ecosystems and Society

________________________________________________________________________

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

________________________________________________________________________

Fabio Guimarães Gonçalves, Author
ACKNOWLEDGMENTS

This research was funded by NASA’s Terrestrial Ecology Program, with additional support from the Brazilian Ministry of Education (CAPES/Fulbright Doctoral Program, process BEX-2684/06-3), the Jet Propulsion Laboratory (JPL), and the Woods Hole Research Center (WHRC). I was also very fortunate to receive fellowships and tuition support from the OSU Department of Forest Ecosystems & Society (special thanks to Bev Law), the Graduate School, and International Student Advising & Services. Airborne lidar data were acquired with support from the United States Agency for International Development and the U.S. Department of State with the technical assistance of the Brazilian Corporation for Agricultural Research (EMBRAPA) and the U.S. Forest Service Office of International Programs.

I would like to thank the Santarém office of the Large Scale Biosphere-Atmosphere Experiment in Amazonia (LBA) for providing logistical support in the field (especially Louro Lima), and Luciano Dutra and João Roberto dos Santos of INPE for their assistance in obtaining research authorizations from the overly bureaucratic Brazilian government. I would also like to thank the NSIDC User Services staff and Maiza Nara dos Santos (Sustainable Landscapes Program), who were extremely helpful in providing information about ICESat/GLAS data products and the airborne lidar data, respectively.

The field campaigns would not have been possible without the know-how and support of Edilson Oliveira (UFAC), the IFT crew (Rone Brito, Alfredo Viana, Manoel Vitorino, and Amildo Jesus), and the local assistants Jony Oliveira, Raimundo dos Santos, Iracelio Silva, and Emerson Pedroso. I sincerely thank them for their invaluable assistance with the field acquisitions.

It has been an honor to be Robert Treuhaft’s first graduate student and to take advantage of his vast knowledge and unique perspective on the field. I cannot thank Robert enough for his patience and guidance as I ventured into the unfamiliar territory of lidar and InSAR. I was also very fortunate to be adopted by Bev Law, who provided an institutional home for me in the US, guidance, and valuable
feedback; and to work side by side with Alessandro Baccini and Wayne Walker, who hosted me at the WHRC for over a year for an enormously productive and rewarding collaboration.

I would like to thank Dick Waring and Warren Cohen for agreeing to serve on my graduate committee and sharing their valuable experience and knowledge, and also Dan Sullivan, Glen Murphy, and Claire Montgomery for serving as a GCR.

I made it to the finish line thanks to Hyojung Kwon, Piero Mazzini, and Sooie Hoe Loke, whose help and encouragement during the final months of this project were valuable beyond words. I would like to extend a very special thanks to all my colleagues, friends, and family who have supported me in different ways over the course of my PhD.

Finally, I would like to thank my beautiful wife Ive for her incredible patience, support, and love throughout this journey, and my children Davi, Lucas, and Pedro for making our lives fun and fulfilling.
CONTRIBUTION OF AUTHORS

Robert N. Treuhaft assisted with study design, data collection, data analysis, and writing of Chapters 2 and 3. Beverly E. Law assisted with study design, interpretation of results, and writing of Chapters 2 and 3. Alessandro Baccini and Wayne S. Walker provided essential data and assisted with data analysis and writing of Chapters 2 and 3. João Roberto dos Santos and Paulo Maurício L. A. Graça assisted with data collection and provided valuable comments. André Q. Almeida assisted with data collection and contributed to parts of the analysis. Michael Keller provided valuable data and comments.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1. Introduction</th>
<th>................................................................. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 2. Estimating aboveground biomass in tropical forests: Field methods and error analysis for the calibration of remote sensing observations</td>
<td>9</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>................................................................. 10</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>................................................................. 11</td>
</tr>
<tr>
<td>MATERIALS &amp; METHODS</td>
<td>................................................................. 14</td>
</tr>
<tr>
<td>RESULTS</td>
<td>................................................................. 21</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>................................................................. 27</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>................................................................. 38</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>................................................................. 57</td>
</tr>
<tr>
<td>Chapter 3. Improving lidar-based estimates of tropical forest biomass: The performance of Fourier versus traditional structural metrics</td>
<td>59</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>................................................................. 60</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>................................................................. 61</td>
</tr>
<tr>
<td>MATERIALS &amp; METHODS</td>
<td>................................................................. 66</td>
</tr>
<tr>
<td>RESULTS &amp; DISCUSSION</td>
<td>................................................................. 77</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>................................................................. 89</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>................................................................. 112</td>
</tr>
<tr>
<td>Chapter 4. Conclusions</td>
<td>................................................................. 120</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Geographical location of the Tapajós National Forest, PA, Brazil, outlined in yellow.</td>
</tr>
<tr>
<td>2.2</td>
<td>Variation in basal area ($B$), top height ($H_{TOP}$), and aboveground biomass ($AGB$) with stand age ($t$) at Tapajós, as described by the growth model of Neeff &amp; Santos (2005).</td>
</tr>
<tr>
<td>2.3</td>
<td>Schematic diagram illustrating an arbitrary intersection of a GLAS footprint (ellipse) and a field plot (square), portrayed in their correct relative nominal sizes and shapes.</td>
</tr>
<tr>
<td>2.4</td>
<td>Differences between repeated measurements of (a) diameter, (b) height to the base of the live crown, (c) total height, (d) crown depth, and (e) crown radius, ordered by the magnitude of the measurement (estimated as the average of the two measurements).</td>
</tr>
<tr>
<td>2.5</td>
<td>(a) Log-log scatterplot of laser-measured heights versus ocular heights, with points coded by height type ($H_C$ = circles, $H_T$ = triangles) and observer ($O_1$ = filled symbols, $O_2$ = open symbols).</td>
</tr>
<tr>
<td>2.6</td>
<td>Average stem density (bars) and tree height (circles) per diameter class (trees ≥ 10 cm) for secondary (dark gray) and primary (light gray) forests.</td>
</tr>
<tr>
<td>2.7</td>
<td>Relationship between stand biomass at Tapajós (trees ≥ 5 cm diameter) and the fraction of that biomass found in trees 5–10 cm diameter.</td>
</tr>
<tr>
<td>2.8</td>
<td>Dependence of the co-location error estimated with the binomial approach on the spatial overlap between GLAS and field measurements.</td>
</tr>
<tr>
<td>2.9</td>
<td>Dependence of the biomass error introduced by the growth model of Neeff &amp; Santos (2005) on the temporal difference between field and remote sensing acquisitions (i.e., remote sensing data acquired 1, 2, 3, …, 10 years prior to the field measurements).</td>
</tr>
<tr>
<td>2.10</td>
<td>Stand biomass versus measurement ($\sigma_M$), allometric ($\sigma_A$ and $\sigma_S$), and co-location ($\sigma_C$) errors, with the estimated regression lines.</td>
</tr>
<tr>
<td>2.A1</td>
<td>Biomass estimation error at Tapajós due to the use of a site-specific height-diameter allometry versus the number of trees used to develop the allometry.</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.1</td>
<td>Geographical location of the Tapajós National Forest, PA, Brazil, outlined in yellow.</td>
</tr>
<tr>
<td>3.2</td>
<td>Side view of the modeled canopy structure of (a) a primary forest (AGB = 421.0 Mg ha(^{-1})), and (b) an early successional forest (AGB = 12.1 Mg ha(^{-1})) measured at Tapajós, represented in their correct relative sizes.</td>
</tr>
<tr>
<td>3.3</td>
<td>GLAS waveform for the primary forest plot displayed in Figure 3.2a.</td>
</tr>
<tr>
<td>3.4</td>
<td>Relative vegetation density profiles from field (shaded polygon), GLAS (black line), and ALS (gray line) measurements for nine field plots at Tapajós.</td>
</tr>
<tr>
<td>3.5</td>
<td>Relationship between species diversity (as measured by the Shannon index) and (a) mean stand-level wood density (calculated on a per stem basis), (b) basal area, (c) mean height, and (d) aboveground biomass for the 30 plots measured in 2010.</td>
</tr>
<tr>
<td>3.6</td>
<td>Relationship between height-based metrics derived from GLAS (open symbols) and ALS (filled symbols) versus the corresponding field metrics: (a) mean canopy height, (b) standard deviation, (c–e) quartile heights, and (f) maximum canopy height.</td>
</tr>
<tr>
<td>3.7</td>
<td>Field-estimated aboveground biomass versus that predicted by traditional and Fourier metrics derived from field, ALS, and GLAS profiles for Tapajós, as indicated in the panels.</td>
</tr>
<tr>
<td>3.8</td>
<td>Graphical representation of the 500 Fourier-based models with the smallest RMSPE.</td>
</tr>
<tr>
<td>3.9</td>
<td>Location of the Xingu River basin, Brazil.</td>
</tr>
<tr>
<td>3.10</td>
<td>Field-estimated aboveground biomass versus that predicted by traditional (left) and Fourier (right) metrics derived from GLAS profiles acquired over two sites in the Xingu region (São Félix and Upper Xingu).</td>
</tr>
<tr>
<td>3.11</td>
<td>Left: Performance of the regional GLAS-biomass model based on Fourier metrics; Right: Average biomass estimates from GLAS versus “randomForest” MODIS estimates.</td>
</tr>
<tr>
<td>3.12</td>
<td>Aboveground live biomass density map for the Xingu River basin, Brazil, produced from field plot data, GLAS lidar measurements, and MODIS imagery.</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>3.A1</td>
<td>Relationships between crown radius and (a) stem diameter (on the square root scale), (b) crown depth, and (c) total height, with the estimated regression lines for secondary (gray) and old-growth forests (black).</td>
</tr>
<tr>
<td>3.A2</td>
<td>Relative vegetation density profiles derived from field measurements of crown radius ($C_R$), crown depth ($C_D$), and total height ($H_T$) for nine 12.5 m x 50 m subplots sampled in 2010 (shaded polygons).</td>
</tr>
<tr>
<td>3.A3</td>
<td>Dendrogram for an agglomerative clustering of the 2010 dataset based on the Hellinger distance (Kindt &amp; Coe 2005).</td>
</tr>
<tr>
<td>3.A4</td>
<td>Field profiles derived for six plots at Tapajós that were measured in 2010 and remeasured in 2013.</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>53</td>
</tr>
<tr>
<td>2.2</td>
<td>54</td>
</tr>
<tr>
<td>2.3</td>
<td>55</td>
</tr>
<tr>
<td>2.4</td>
<td>56</td>
</tr>
<tr>
<td>2.1A</td>
<td>57</td>
</tr>
<tr>
<td>3.1</td>
<td>109</td>
</tr>
<tr>
<td>3.2</td>
<td>110</td>
</tr>
<tr>
<td>3.3</td>
<td>111</td>
</tr>
<tr>
<td>3.1A</td>
<td>119</td>
</tr>
</tbody>
</table>

2.1 Allometric equations used to calculate individual tree biomass at Tapajós.

2.2 Summary statistics of differences between repeated measurements of diameter (D), height to the base of the live crown (H_C), total height (H_T), crown depth (C_D), and crown radius (C_R) for trees sampled at Tapajós.

2.3 Characteristics of secondary (SF), selectively-logged (PF_L), and primary (PF) forest stands used in this study.

2.4 Uncertainties in field-based estimates of plot biomass at Tapajós.

2.A1 Standard deviations (SD) of differences in repeated measurements of diameter (D), height to the base of the live crown (H_C), total height (H_T), crown depth (C_D), and crown radius (C_R) for trees sampled at Tapajós.

3.1 Root mean square deviation (RMSD), bias (i.e., mean difference), and results of simple regressions between corresponding field and lidar metrics.

3.2 Root mean square prediction error (RMSPE), cross-validation coefficient of determination (RCV2), and confidence for the selected regression models at Tapajós.

3.3 Root mean square prediction error (RMSPE), cross-validation coefficient of determination (RCV2), and confidence for the selected regression models at São Félix (SFX) and the Upper Xingu (UX).

3.1A Results from the regression of crown radius (C_R, m) on stem diameter (√D, cm), crown depth (C_D, m), and total height (H_T, m).
Vertical Structure and Aboveground Biomass of Tropical Forests from Lidar Remote Sensing

by
Fabio Guimarães Gonçalves
Chapter 1. Introduction

Deforestation and forest degradation have been estimated to contribute 6–20% of total anthropogenic CO$_2$ emissions, ranking as the second largest source of anthropogenic carbon to the atmosphere, after fossil fuel burning (Le Quéré et al., 2009; van der Werf et al., 2009). Uncertainties in the net flux of carbon from these activities are however high, ranging from 10 to 100% in many developed countries, and possibly exceeding 100% in most developing regions (NRC, 2010). The standard practice to estimating the carbon flux from land-use change rely heavily on passive optical systems such as MODIS or Landsat, essentially linking detected changes in forest area or condition to average values of biomass density estimated from forest inventories (Goetz et al., 2009; NRC, 2010; Law and Harmon, 2011). Techniques for classification of deforestation, forest degradation, and regrowth have advanced significantly since the 1990s and are now available to support forest monitoring efforts at regional scales, with unprecedented detail and accuracy (e.g. Kennedy et al., 2007; Asner, 2009). Conversely, inventory-based estimates of forest biomass have been regarded as the major source of uncertainty in calculations of land-use emissions, particularly in the tropics, where systematic forest inventories are rare (Houghton et al., 2009). Methods for obtaining accurate, spatially explicit estimates of biomass density are therefore required to reduce uncertainties in the net land flux, and to support international climate agreements and emerging carbon markets.

Our ability to estimate aboveground forest biomass from remote sensing observations has advanced substantially over the past decade, largely due to the increased availability of direct three-dimensional (3-D) measurements of vegetation structure provided by light detection and ranging (lidar; Lefsky et al. 2002) and interferometric synthetic aperture radar (InSAR; Treuhaft et al. 2004). Methods for retrieving relevant structural information from lidar and InSAR have been tested and validated in a number of tropical sites, resulting in errors in canopy height estimates that range from about 2 to 4 m, at plot scales of 0.5 ha and below (Drake et al., 2002; Neeff et al., 2005; Lefsky et al., 2007; Hajnsek et al., 2009; Treuhaft et al., 2009; Simard et al., 2011). On the other hand, methods for transforming remotely sensed measurements of structure into predictions of forest biomass are still under
development. Although lidar and InSAR have shown improved biomass estimation performance relative to nonstructural approaches, it is still unclear what accuracies are achievable from these techniques in forests with high biomass density. Studies that employ empirical, height-based methods have reported errors that range from approximately 40 to 100 Mg ha\(^{-1}\) (Drake et al., 2003; Lefsky et al., 2005; Neeff et al., 2005; Treuhaft et al., 2010; Clark et al., 2011; Asner et al., 2012; Baccini et al., 2012), generally well above the required accuracy of 20 Mg ha\(^{-1}\) or 20% (not to exceed 50 Mg ha\(^{-1}\) in forests ≥ 250 Mg ha\(^{-1}\)), necessary to estimate carbon fluxes from deforestation and forest degradation to within 20%, at spatial resolutions of 100–250 m (Houghton et al., 2009; Hall et al., 2011).

The typical approach to producing spatially explicit estimates of biomass from 3-D remote sensing is characterized by two primary steps. First, field estimates of aboveground biomass density are obtained from sample plot data together with published allometric equations, which allow the estimation of tree-level biomass from more easily measured quantities such as diameter, height, and wood density (Brown, 1997; Chave et al., 2005, 2014). Second, the plot-level estimates of biomass are related to co-located remote sensing estimates of structure (e.g., mean canopy height) using a statistical model. The model is then applied together with remote sensing data to predict biomass in locations where ground measurements are not available (Asner & Mascaro, 2014; Drake et al., 2003; Treuhaft et al., 2010, 2014; Chapter 3). When the 3-D measurements are spatially discontinuous, as is usually the case with lidar, the resulting biomass predictions can be further integrated with radar and/or passive optical imagery (typically using machine learning algorithms) to produce wall-to-wall maps of biomass/carbon (Baccini et al., 2012; Saatchi et al., 2011), although often with poorer resolution and unknown accuracy.

Because remote sensing approaches to biomass estimation begin with field estimates of biomass, a thorough understanding of the sources of error associated with biomass estimates in calibration plots is critical to develop strategies for reducing the overall uncertainty. In Chapter 2, we address this issue by using field plot data collected at the Tapajós National Forest, Brazil, to gain a better understanding of the uncertainty associated with plot-level biomass estimates.
obtained for calibration of lidar data. In addition to accounting for sources of error that would be normally expected in conventional biomass estimates (e.g., measurement and allometric errors; Brown et al., 1995; Chave et al., 2004; Keller et al. 2001; Molto et al. 2013; Phillips et al. 2000), we examine two sources of uncertainty that are specific to the calibration process and should be taken into account in most remote sensing studies: (1) the error resulting from spatial disagreement between field and remote sensing samples (co-location error), and (2) the error introduced when accounting for temporal differences in data acquisition. Because the uncertainty in the biomass estimate depends in part on uncertainties in the field plot data, we use repeated measurements to determine the precision of several tree-level attributes that affect the estimation of biomass and other structural characteristics. Finally, because height measurements are critical to computing accurate biomass estimates (Chave et al., 2005, 2014; Nogueira et al. 2008), but are often time-consuming and difficult to obtain in practice and especially in tall, closed-canopy forests (Hunter et al. 2013; Larjavaara & Muller-Landau, 2013), we present and evaluate a method based on calibration of ocular estimates with lasers, which can be used to optimize the collection of tree height information in the field.

In Chapter 3, we use airborne and space-based lidar observations along with detailed forest inventory data collected at the Tapajós National Forest to advance methods of biomass estimation from remotely sensed structure. More specifically, our objectives were to (1) test and refine methods for the extraction of structural information from lidar data; (2) determine the accuracy of lidar estimates of structure in relation to detailed field measurements of vertical structure; (3) develop and validate structure-based models for optimal prediction of aboveground biomass; and (4) Integrate the resulting models in a multi-stage scaling approach to biomass estimation (Baccini et al., 2012) and produce a wall-to-wall map of biomass for a test site in the Amazon. In addition to evaluating the performance of simple lidar canopy metrics such as mean or maximum canopy height, we explore Fourier transforms of profiles as a means of incorporating information on directional, mechanism-driven change in structure into empirical biomass models. Rather than relying on regional averages of basal area and wood density, or attempting to estimate it from remote
sensing, we look at lidar profiles in the frequency domain and consider the vertical scales at which the canopy is organized explicitly, as it relates to basal area, species composition (i.e. wood density), and a number of other forest attributes. We build on our previous work (Treuhaft et al., 2010, 2013) with the airborne LVIS instrument (Blair et al., 1999) at La Selva, Costa Rica, extending it to a tropical moist forest in the central Amazon, and to the analysis of 3-D data from commercial off-the-shelf airborne lidar and the spaceborne Geoscience Laser Altimeter System (GLAS, Schutz et al. 2005). We also investigate an alternative modeling approach that allows more flexibility in the use of Fourier-based metrics while accounting for the uncertainty in the variable selection process.

References


forest using Fourier transforms of profiles from lidar or interferometric SAR. *Geophysical Research Letters*, 37(23), L23403.
Chapter 2. Estimating aboveground biomass in tropical forests: Field methods and error analysis for the calibration of remote sensing observations

F.G. Gonçalves a,b,*, R.N. Treuhaft c, B.E. Law a, J.R. Santos d, W.S. Walker b, A. Baccini b, A.Q. Almeida e, P.M.L.A. Graça f

a Department of Forest Ecosystems & Society, Oregon State University, Corvallis, OR, USA
b Woods Hole Research Center, Falmouth, MA, USA
c Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
d Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brazil
e Núcleo de Engenharia Agrícola, Universidade Federal de Sergipe, SE, Brazil
f Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, Brazil

Remote Sensing of Environment
Elsevier – Customer Service Department
6277 Sea Harbor Drive
Orlando, FL 32887-4800 USA
Submitted

* Corresponding author: fgoncalves@whrc.org
2.1 ABSTRACT

Mapping and monitoring of forest carbon stocks across large areas in the tropics will necessarily rely on remote sensing approaches, which in turn depend on field estimates of biomass for calibration and validation purposes. Despite recent advances in remote sensing approaches to biomass estimation, much progress remains to be made towards properly characterizing sources of error, which begin with the errors in field estimates, and understanding how these errors propagate to contribute to the overall uncertainty in predicted carbon stocks. Here, we used field plot data collected in a tropical moist forest in the central Amazon to gain a better understanding of the uncertainty associated with plot-level biomass estimates obtained specifically for the calibration of remote sensing measurements. In addition to accounting for sources of error that would be normally expected in conventional biomass estimates (e.g., measurement and allometric errors), we also examined two sources of uncertainty that are specific to the calibration process and should be taken into account in most remote sensing studies: (1) the error resulting from spatial disagreement between field and remote sensing measurements (i.e., co-location error), and (2) the error introduced when accounting for temporal differences in data acquisition. Because uncertainty in the biomass estimate depends in part on uncertainties in the field plot data, we used repeated measurements to determine the precision of several tree-level attributes that affect the estimation of biomass and other structural characteristics. Finally, because height measurements are critical to computing accurate biomass estimates, but are often time-consuming and difficult to obtain in tall, closed-canopy forests, we present and evaluate a method based on calibration of ocular estimates with lasers. A thorough understanding of the sources of error associated with plot-level biomass estimates in tropical forests is critical to determine confidence in remote sensing estimates of carbon stocks and fluxes, and to develop strategies for reducing the overall uncertainty of remote sensing approaches.
2.2 INTRODUCTION

Our ability to estimate aboveground forest biomass from remote sensing observations has advanced substantially over the past decade, largely due to the increased availability of direct three-dimensional (3-D) measurements of vegetation structure provided by light detection and ranging (lidar; Lefsky, Cohen, Parker, & Harding, 2002) and interferometric synthetic aperture radar (InSAR; Treuhaft, Law, & Asner, 2004). Although approaches to forest biomass estimation based on remotely sensed structure have yet to be fully developed and validated (cf. Zolkos, Goetz, & Dubayah, 2013), they are already greatly expanding our knowledge of the amount and spatial distribution of carbon stored in terrestrial ecosystems, particularly in tropical forests (e.g., Asner et al., 2010; Baccini et al., 2012; Saatchi et al., 2011), where large areas have never been inventoried on the ground (Houghton, 2005). Lidar remote sensing, calibrated with field measurements and combined with wall-to-wall observations from InSAR and/or passive optical systems, represents a promising alternative to more traditional approaches to biomass mapping (e.g., Houghton, Lawrence, Hackler, & Brown, 2001; Malhi et al., 2006) and is expected to play a key role in forest monitoring systems being developed in the context of climate change mitigation efforts such as REDD (Reducing Emissions from Deforestation and Forest Degradation), and to improve our understanding of the global carbon balance (Asner, 2009; Goetz & Dubayah, 2011; Hall et al., 2011; Houghton, Hall, & Goetz, 2009).

The typical approach to producing spatially explicit estimates of biomass from 3-D remote sensing is characterized by two primary steps. First, field estimates of aboveground biomass density are obtained from sample plot data together with published allometric equations, which allow the estimation of tree-level biomass from more easily measured quantities such as diameter, height, and wood density (Brown, 1997; Chave et al., 2005, 2014). Second, the plot-level estimates of biomass are related to co-located remote sensing estimates of structure (e.g., mean canopy height) using a statistical model. The model is then applied together with remote sensing data to predict biomass in locations where ground measurements are not available (Asner & Mascaro, 2014; Drake et al., 2003; Treuhaft et al., 2010, 2015; Chapter 3). When the 3-D measurements are spatially discontinuous, as is usually the case with lidar,
the resulting biomass predictions can be further integrated with radar and/or passive optical imagery (typically using machine learning algorithms) to produce wall-to-wall maps of biomass/carbon (Baccini et al., 2012; Saatchi et al., 2011), although often with poorer resolution and unknown accuracy.

One of the main limitations of this scaling approach, as noted by Clark & Kellner (2012), is that biomass is never measured directly (i.e., quantified by harvesting and weighing the leaves, branches, and stems of trees). Because direct measurements are laborious, time-consuming, and ultimately destructive (e.g., Araújo, Higuchi, & Carvalho, 1999), the remotely sensed structure is calibrated against allometrically estimated biomass (a function of diameter and sometimes height and wood density) and the final product is, in essence, “an estimate of an estimate” of biomass.

Despite significant advances in the development of allometric equations for tropical forest trees over the past decade (Chave et al., 2005, 2014), the allometrically-derived biomass is subject to a number of sources of error, including (i) uncertainty in the estimation of the parameters of the allometric equation as a result of sampling error (e.g., resulting from a relatively small number of trees being harvested or bias against the harvest of trees with a “typical” form), natural variability in tree structure (i.e., trees of the same diameter, height, and wood density can display a range of biomass values), and measurement errors on the harvested trees; (ii) uncertainty associated with the choice of a particular equation or application of a given equation beyond the site(s) and/or species for which it was developed (uncertainty driven primarily by biogeographic variation in allometric relations due to soil fertility and climate); and (iii) measurement errors in the diameter, height, and wood density of the trees that the allometric equation is being applied to (Ahmed, Siqueira, Hensley, & Bergen, 2013; Chave et al., 2004; Clark & Kellner, 2012). Combined, these sources of uncertainty have been estimated to represent approximately 50–80% of the estimated biomass at the tree level, and over 20% at the plot scale (Chave et al., 2004).

Because remote sensing algorithms for prediction of forest biomass are typically calibrated with allometrically estimated biomass (see Colgan, Asner, & Swemmer
(2013) for an exception), they incorporate all of the sources of uncertainty described above, in addition to those associated with the remote sensing observations. As a result, while the precision (degree of reproducibility) of remote sensing estimates of biomass can be easily assessed, their accuracy is rarely known. And although precision may be all that is needed for the relative tracking of changes in carbon stocks for REDD-like initiatives, accuracy is ultimately critical for determining the absolute amount of carbon stored in forests, as required for global carbon budgets and climate change science (Clark & Kellner, 2012).

While the accuracy of remote sensing-based estimates of biomass cannot be truly determined without whole-plot harvests, it can nevertheless be optimized. When calibrating remotely sensed structure to allometrically estimated biomass, it is reasonable to expect that the uncertainty in the estimated biomass will vary from plot to plot as a function of differences in, for example, tree size distribution and species composition. If these plot-level uncertainties can be estimated precisely relative to one another, the prediction accuracy of the statistical model relating biomass to remotely sensed structure can be significantly improved by giving sample plots with smaller standard deviations more weight in the parameter estimation. In practice, this can be accomplished using the method of weighted least squares (WLS), where each plot is weighted inversely by its own variance (Bevington, 1969; Ramsey & Schafer, 2002). Although considered a standard statistical technique for dealing with nonconstant variance when responses are estimates, WLS is almost never used in the current context, in part because the uncertainty in the field biomass is rarely quantified.

In this study, we use field plot data collected at the Tapajós National Forest, Brazil, to gain a better understanding of the uncertainty associated with plot-level biomass estimates obtained specifically for calibration of remote sensing measurements in tropical forests. In addition to accounting for sources of error that would be normally expected in conventional biomass estimates (e.g., measurement and allometric errors; Brown et al., 1995; Chave et al., 2004; Keller, Palace, & Hurtt, 2001; Molto, Rossi, & Blanc, 2013; Phillips, Brown, Schroeder, & Birdseye, 2000), we examine two sources of uncertainty that are specific to the calibration process and
should be taken into account in most remote sensing studies: (1) the error resulting from spatial disagreement between field and remote sensing samples (co-location error), and (2) the error introduced when accounting for temporal differences in data acquisition. Because the uncertainty in the biomass estimate depends in part on uncertainties in the field plot data, we use repeated measurements to determine the precision of several tree-level attributes that affect the estimation of biomass and other structural characteristics. Finally, because height measurements are critical to computing accurate biomass estimates (Chave et al., 2005, 2014; Nogueira, Fearnside, Nelson, Barbosa, & Keizer, 2008), but are often time-consuming and difficult to obtain in practice and especially in tall, closed-canopy forests (Hunter, Keller, Victoria, & Morton, 2013; Larjavaara & Muller-Landau, 2013), we present and evaluate a method based on calibration of ocular estimates with lasers, which can be used to optimize the collection of tree height information in the field.

2.3 MATERIALS & METHODS

2.3.1 Study site

The Tapajós National Forest is located along highway BR-163, approximately 50 km south of the city of Santarém, Pará, in the central Brazilian Amazon (Figure 2.1). The climate is tropical monsoon (Köppen Am), with a mean annual temperature of 25.1°C and annual precipitation of 1909 mm, with a 5-month dry season (< 100 mm month⁻¹) between July and November (Vieira et al., 2004). The vegetation is dense, upland, tropical moist forest. Common genera among 193 tree species sampled in this study include *Psychotria* (Rubiaceae), *Protium* (Burseraceae), *Otoba* (Myristicaceae), *Eschweilera* (Lecythidaceae), *Pouteria* (Sapotaceae), and *Rinorea* (Violaceae) in primary forests, and *Cecropia* (Urticaceae), *Banara* (Salicaceae), and *Inga* (Fabaceae) in secondary forests. The soils are nutrient-poor oxisols and ultisols, with low pH, organic matter, and cation exchange capacity, and a high concentration of aluminum oxides (Silver et al., 2000a). Our sample sites were situated on a relatively flat plateau, with the elevation ranging from approximately 80 to 180 m.
2.3.2 Field data

Field data were collected in September 2010 in 30 0.25-ha plots (50 m x 50 m) intended for calibration of lidar data acquired by the Geoscience Laser Altimeter System (GLAS; Chapter 3). Of the 30 plots, 8 were primary forest (PF), 8 were primary forest subject to reduced-impact selective logging (PF_L) between 1999 and 2003 (~3.5 trees harvested per hectare during this period; Gonçalves & Santos, 2008), and 14 were secondary forest (SF) with different age and disturbance histories. Plots were centered on GLAS footprints selected along two sensor acquisition tracks, spanning a wide range in vertical structure and aboveground biomass (Figure 2.1). Individual footprint centers were located on the ground using a total station and the Differential Global Positioning System (DGPS).

Field biometric measurements included diameter at breast height (D) measured with a diameter tape at 1.3 m and recorded to the nearest 0.1 cm; height to the base of the live crown (HC) and total height (HT), estimated visually by experienced members of the field crew to the nearest 0.5 m; and crown depth (CD), calculated as the difference between HT and HC. Measurements were taken for each living tree ≥ 5 cm in diameter in early successional stands and ≥ 10 cm in all other stands. For a 12.5 m x 50 m subplot extending along the major axis of the GLAS footprint, we also measured crown radius (CR) in four orthogonal directions by projecting the edge of the crown to the ground and recording its horizontal distance to the trunk to the nearest 0.1 m using a tape measure. All trees were identified to species or genus (when species was uncertain) level and assigned a wood density value (ρ, oven-dry weight over green volume) derived from the literature (Chave et al., 2006; Reyes, Brown, Chapman, & Lugo, 1992).

To estimate measurement errors associated with the inventory, we conducted a blind remeasurement of 2–4 trees selected at random in each plot, resulting in a total resampling effort of 3%. For a portion of the trees that were remeasured (and additional trees selected in open areas), heights were also obtained with a laser rangefinder using the tangent method (Larjavaara & Muller-Landau, 2013). We used these observations to develop a regression model relating precise laser-measured heights to less precise, however more readily obtained, visually estimated heights (cf.
Gonçalves & Santos, 2008). This model, described in detail in section 2.4.2, was applied to calibrate all visually estimated heights.

2.3.3 Biomass estimation

The oven-dry aboveground mass of each live tree \((M)\) was estimated from its diameter, total height, and wood density using an established allometric equation for tropical moist forests (Chave et al., 2005; Table 2.1). Exceptions were made for \(Cecropia\) spp. and palms, which differ significantly from other species in wood density and allometry (Chave et al., 2005; Nelson et al., 1999) and had their biomass estimated with specific equations, as indicated in Table 2.1. We selected Chave’s equation as the basis of our estimate because it was developed using a large number of harvested trees (1350) covering a wide range in diameter (5–156 cm), and because it included information on tree height and wood density, which greatly improves the accuracy of biomass estimates (Chave et al., 2005, 2014). In addition, approximately 43% of the trees used in the fit were harvested in the Brazilian Amazon, and 11% in the state of Pará, in sites having climatic and edaphic conditions comparable to ours. Nevertheless, we also estimated \(M\) in this study using two additional, widely used equations (Brown, 1997; Chambers, Santos, Ribeiro, & Higuchi, 2001; Table 2.1), to obtain a measure of allometric uncertainty as described in section 2.3.4.2.

The aboveground biomass density (AGB, Mg ha\(^{-1}\)) at the plot level was calculated by adding the masses of all inventoried trees in the plot and dividing by the plot area (i.e., 0.25 ha). In plots where the minimum diameter was 10 cm, we corrected for the AGB in the 5–10 cm class by (i) fitting a negative exponential function to the diameter distribution of the plot (van Laar & Akça, 2007)

\[
N_i = k e^{-ad_i}
\]

where \(N_i\) was the number of trees per hectare in the \(i\)th diameter class with midpoint \(d_i\), and \(k\) and \(a\) were model parameters estimated by nonlinear least squares; (ii) estimating the number of trees per hectare in the 5–10 cm class; and (iii) multiplying the resulting number by the biomass of a tree with diameter of 7.5 cm (midpoint) and...
wood density equal to the plot mean. To avoid errors due to the estimation of a mean height for the 5–10 cm class, we used an alternative equation of Chave et al. (2005) based on diameter and wood density only (Table 2.1).

Because the GLAS lidar data were acquired in 2007, three years prior to our field measurements, we applied the site-specific, stand-level growth model of Neeff & Santos (2005)

\[
\begin{align*}
AGB_t &= 0.397 B_t H_{TOP_t}, \text{ with} \\
B_t &= 21.057 \left(1 - e^{-0.109 t}\right)^{1.894} \\
H_{TOP_t} &= 23.067 \left(1 - e^{-0.074 t}\right)^{0.946}
\end{align*}
\]

to all SF plots to correct for the AGB change between observation epochs, where \(AGB_t\) is the plot biomass (Mg ha\(^{-1}\)), estimated with the equation of Brown, Gillespie, & Lugo (1989) (Table 2.1); \(B_t\) is the basal area (m\(^2\) ha\(^{-1}\)); \(H_{TOP_t}\) is the top height (m), defined as the mean total height of the tallest 20% of the trees; and \(t\) is the stand age (years since stand initiation). Because this growth model is based on a biomass equation that is different from the ones used in this study, we calculated the biomass change as a proportion of the plot biomass. This was done by (i) inverting the first line of (2) to estimate the plot age in 2010 \(t_{2010}\) from its measured biomass \((AGB_{2010})\); (ii) estimating the plot biomass at the time of the GLAS acquisition \((AGB_{2007})\), making \(t = t_{2010} - 3\); and (iii) calculating the ratio \((AGB_{2010} - AGB_{2007})/AGB_{2010}\). For primary forests, we assumed no biomass change in the 3-year period. This is supported by Neeff & Santos (2005), who found that biomass tends to increase rapidly in early stand development at Tapajós, reaching near-asymptote as early as 40–50 years after clear-cutting (Figure 2.2).

2.3.4 Error Analysis

2.3.4.1 Individual tree measurements
Measurement errors in D, H_C, H_T, C_D, and C_R were described in terms of total error, systematic error (bias), and random error. The total error for each attribute was quantified with the root mean square deviation (RMSD)

\[
\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2}, \quad \text{with} \quad e_i = m_{1i} - m_{2i} \quad (3)
\]

where \( n \) is the number of pairs of repeated measurements for the attribute, and \( e_i \) is the measurement difference for the \( i \)th pair, with \( m_{1i} \) and \( m_{2i} \) representing the original measurement and remeasurement, respectively. The systematic and random errors were quantified, respectively, as the mean and sample standard deviation (SD) of the measurement differences \( e_i \)

\[
\text{Mean} = \frac{1}{n} \sum_{i=1}^{n} e_i \quad (4)
\]

\[
\text{SD} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (e_i - \text{Mean})^2} \quad (5)
\]

We also calculated all of the above in relative terms, expressing \( e_i \) as a fraction of the average of the two measurements. For the wood density values, the standard deviation was either taken or estimated from the supplementary material provided by Chave et al. (2006).

To test the hypothesis of no systematic difference between the first and second measurements of a given attribute, we used either a paired t-test or the alternative Wilcoxon signed-rank test, depending on the assessment of distributional assumptions and the presence of outliers (Ramsey & Schafer, 2002). To determine whether measurement variation increased with the magnitude of the measurement, we (i) divided the measurements for a given attribute into four classes with an equal number of samples; (ii) regressed the SD calculated for each size class on the average value of the measurement for that class; and (iii) tested whether the slope was significantly
different from zero. Finally, we tested if measurement differences varied with forest type by including forest type as a factor in the regression of the absolute measurement difference on the magnitude of the measurement—i.e., incorporating different intercepts and slopes for SF, PF_L and PF—and testing for the equality of the coefficients using the extra-sum-of-squares F-test (Ramsey & Schafer, 2002). Similar methods were applied to analyze differences between ocular and laser-measured heights, as well as differences between observed and predicted laser heights, as discussed in section 2.4.2.

2.3.4.2 Biomass

Measurement errors in diameter (σ_D), height (σ_H), and wood density (σ_ρ) were propagated to the biomass estimate by expanding the allometric equations in Table 2.1 to a Taylor series and retaining only first-order terms. For a model like M_{T2} (Table 2.1), of the form M = aD^kHρ, with ρ uncorrelated with both D and H, we expressed the uncertainty in the mass of a tree (σ_M) in terms of measurement errors as (Chave et al., 2004)

\[
\sigma_M = \left[ \sigma_D^2 \left( \frac{\partial M}{\partial D} \right)^2 + \sigma_H^2 \left( \frac{\partial M}{\partial H} \right)^2 + \sigma_\rho^2 \left( \frac{\partial M}{\partial \rho} \right)^2 + 2\sigma_{DH}^2 \left( \frac{\partial M}{\partial D} \right) \left( \frac{\partial M}{\partial H} \right) \right]^{1/2}
\]

\[
= M \left( k^2 \frac{\sigma_D^2}{D^2} + \frac{\sigma_H^2}{H^2} + \frac{\sigma_\rho^2}{\rho^2} + 2k \frac{\sigma_{DH}^2}{DH} \right)^{1/2} \quad (6)
\]

where the terms in parentheses in the upper equation are the partial derivatives of M with respect to each of the dendrometric quantities, D, H, and ρ; and σ_{DH} is the covariance between D and H.

We accounted for two sources of allometric uncertainty: (i) the uncertainty induced by errors in the determination of the allometric parameters, σ_A, estimated as (Baskerville, 1972; Chave et al., 2004)
\begin{equation} \sigma_A = \left[ e^{(2\hat{\sigma}^2 + 2\ln M)} - e^{(\hat{\sigma}^2 + 2\ln M)} \right]^{1/2} = (e^{\hat{\sigma}^2} - 1)^{1/2} \langle M \rangle \quad (7) \end{equation}

where \( \hat{\sigma} \) is the standard error of the regression on log-transformed data (see Table 2.1), and \( \langle M \rangle = M \times \exp(\hat{\sigma}^2/2) \) is an unbiased estimate of the back-transformed biomass prediction \( M^1 \); and (ii) the uncertainty involved in the selection of the allometric equation, \( \sigma_S \), estimated by calculating the mass of each tree with three independent equations (\( M_{T_2}, M_{A_1}, \) and \( M_{A_2} \) in Table 2.1) and obtaining the standard deviation of the resulting values.

Spatial disagreement between the field plots and the GLAS footprints introduced additional uncertainty in the biomass estimates. This co-location error, \( \sigma_C \), occurred because of positional errors associated with both data sets, and because the size, shape, and orientation of the field and the GLAS samples did not exactly coincide. We took a Monte Carlo approach based on counting statistics (Treuhaft et al., 2010) to estimate the difference in biomass between what was measured in each 50 m x 50 m field plot and what was actually present in the area covered by the GLAS footprint (Figure 2.3). Each of the \( n_p \) trees in a field plot was considered to represent a “mass bin” which, on average, was populated by one tree. To simulate biomass for each of the three zones depicted in Figure 2.3 (F, I, and G), we used the binomial distribution

\begin{equation} P_B(x|n_z, p) = \binom{n_z}{x} p^x (1-p)^{n_z-x} \quad (8) \end{equation}

which describes the probability \( P_B(x|n_z, p) \) of observing \( x \) trees per mass bin out of \( n_z \) possible trees in a zone from Figure 2.3, each of which has a probability \( p = 1/n_p \) of occurring. We populated the bins for each zone at random using the probability weights derived from (8). After populating the bins for each zone and summing over all bins to obtain the simulated biomass, we calculated the difference (\( \text{AGB}_{G\cup I}/\text{Area}_{G\cup I} \) – \( \text{AGB}_{F\cup I}/\text{Area}_{F\cup I} \)). This procedure was repeated automatically.

\footnote{We note that some allometries (e.g., \( M_{T_2} \), Table 2.1) already include the correction factor \( \exp(\hat{\sigma}^2/2) \) in their estimates.}
10^4 times for each plot and the standard deviation of the biomass difference was taken as the co-location error.

In addition to the error sources discussed above, we included in our error budget the uncertainty associated with the corrections based on (1) and (2), described in section 2.3.3. The error of estimating biomass for the 5–10 cm diameter class, $\sigma_{5-10}$, was assumed to be the same as for the 10–15 cm class, where all trees were actually measured and the error could be determined. This assumption was tested and verified on plots where the minimum diameter was 5 cm. The uncertainty associated with the application of the growth model, $\sigma_G$, was estimated by propagating the uncertainties in the parameters of (2) to the determination of the biomass change, in a framework similar to (6).

Errors $\sigma_M$, $\sigma_A$, and $\sigma_S$ were calculated at the tree level and added in quadrature to obtain plot-level estimates on a per-hectare basis. These errors were in turn combined in quadrature with $\sigma_C$, $\sigma_{5-10}$, and $\sigma_G$, calculated directly at the plot level, to obtain an estimate of the overall uncertainty under the assumption of additivity and statistical independence.

### 2.4 RESULTS

#### 2.4.1 Tree measurement errors

Uncertainties resulting from differences in repeated measurements of $D$, $H_C$, $H_T$, $C_D$, and $C_R$ are summarized in Table 2.2. $D$ was the most precisely measured quantity, with a RMSD of less than 2%. Repeated measurements of height were typically within 1 m of each other (RMSD = 15–18%), with approximately half of the $H_C$, and a quarter of the $H_T$ observations showing identical repeated measurements. $C_D$ and $C_R$ measurements showed considerably less agreement (RMSD of 31 and 26%, respectively). However, with the exception of $D$, there was no evidence of a systematic difference between first and second measurements (Table 2.2 and Figure 2.4). For $D$, the data suggested that the second survey produced values that were
lower to a statistically significant degree, although the estimated median difference of less than 0.1 cm has no practical significance.\(^2\)

Measurement variation increased significantly with the magnitude of the measurement across all attributes (Figure 2.4). The estimated rates of increase in the standard deviation of the measurement differences were 4, 10, 7, 18, and 26% for \(D\), \(H_C\), \(H_T\), \(C_D\), and \(C_R\), respectively. When differences were expressed as a percentage of the measurement, \(H_C\) and \(C_R\) showed no significant trend. The relative differences in \(D\) also increased with \(D\) (at a rate of 0.04% cm\(^{-1}\)), although the evidence was only suggestive, and the differences in \(H_T\) and \(C_D\) actually decreased with the magnitude of the measurements, at rates of 0.4% m\(^{-1}\) and 1.9% m\(^{-1}\), respectively. There was no evidence that absolute differences between repeated measurements (both the mean and the rate of change) varied with forest type, after accounting for differences in the magnitude of the measurements.

In terms of wood density, about 90% of the inventoried trees showed a coefficient of variation (CV) of less than 20%. The CV was 15% on average (median of 14%), but ranged from 0 to as high as 64%, depending on the method used to assign the wood density value (e.g., species- vs. genus-level database estimates).

**2.4.2 Height calibration model**

Figure 2.5a shows a scatterplot of ocular (\(H_{ocular}\)) versus laser-measured (\(H_{laser}\)) heights for 200 observations, with different symbols for height to crown base (circles), total height (triangles), and the two observers who performed ocular measurements (filled vs. open symbols). We found that \(H_C\) was estimated with no bias between methods, but that laser measurements of \(H_T\) exceeded ocular estimates from both observers by an average of 1.9 m (9.9%). The differences between ocular and laser measurements were also found to be slightly higher (0.5–0.8 m) for trees measured in PF relative to those measured in PF\(_L\), SF, and open areas, after accounting for tree height. The regression analysis was performed on log-transformed

\(^2\) Statistical significance is observed for such a small difference because the test used (Wilcoxon signed-rank test) is based on the ranks of the magnitudes of the differences.
data to correct for the increasing variability of $H_{laser}$ with height, as observed on the original scale. The lines in Figure 2.5a represent the fit of the parallel lines model

$$
\ln(H_{laser}) = \beta_0 + \beta_1 I_{HC_{O_1}} + \beta_2 I_{HC_{O_2}} + \beta_3 \ln(H_{ocular})
$$

which incorporates the indicator variables $I_{HC_{O_1}}$ and $I_{HC_{O_2}}$ to allow the intercept to vary arbitrarily with different combinations of height type ($HC$ or $HT$) and observer ($O_1$ or $O_2$). For $HT$, we found no evidence of an intercept difference between the two observers. $HT$ observations were therefore grouped together and treated as the reference level in (9). A more general, separate lines model (different intercepts and slopes) was also tested, but there was no evidence of a difference in the slopes of the resulting regression lines.

From Figure 2.5a, a straight-line regression of $\ln(H_{laser})$ on $\ln(H_{ocular})$ seems appropriate for each of the three height type-observer combinations (i.e., $HC_{O_1}$, $HC_{O_2}$, and $HT_{O_1}/HT_{O_2}$). The bias introduced by taking the antilog of the model estimates was < 1%, but still corrected for as in Baskerville (1972). The ocular measurements explained 93% of the variation in the laser measurements, with a root mean square error (RMSE) about the regression of 2 m (14.9%) (Figure 2.5b). Despite the interobserver variability, the ocular method was accurate for $HC$, resulting in a model error of only 1.4 m (13.5%). The error for $HT$ was higher, estimated at 2.5 m (15.0%). In fact, we found that the absolute model error increases significantly with height (Figure 2.5b), at a rate of about 0.1 m/m, although no clear trend was observed for the relative error. We also found no evidence that the model error is associated with the tree location (i.e., PF, PF_L, SF, or open areas), after accounting for tree height.

2.4.3 Field biomass

Plot-level aboveground biomass ranged from 1.9 to 130.1 Mg ha$^{-1}$ in secondary forests, and from 162.6 to 423.6 Mg ha$^{-1}$ in primary forests, with no apparent

---

$^3$ i.e., $\beta_0$ is the intercept used for the calibration of any $H_t$, regardless the observer. For $HC$, $\beta_0 + \beta_1$ is the intercept for observer 1, and $\beta_0 + \beta_2$ is the intercept for observer 2.
difference between PF and PFL plots. The overall mean was 174.8 ± 134 (SD) Mg ha\(^{-1}\), and the median was 172.8 Mg ha\(^{-1}\). These results are summarized by forest type in Table 2.3, along with a number of other stand characteristics.

Figure 2.6 shows the average stem density (bars) and tree height (circles) per diameter class for all secondary (dark gray) and primary (light gray) forest plots. The solid lines show the fit of (1) to the average diameter distributions, and the dashed lines show the fit of a similar exponential decay model to the tree height data. The diameter distributions followed an inverted J-shaped curve typical of tropical forests, with the ratio of the number of trees in successive diameter classes roughly constant (~1.9 for SF and 1.7 for PF/PFL). Secondary forests showed considerably fewer (and generally shorter) trees than primary forests at any given diameter class, except for the smallest classes (< 20 cm). Primary forests showed fairly balanced diameter distributions (both PF and PFL stands), while secondary forests contained virtually no trees above 60 cm diameter.

The estimated frequency of trees 5–10 cm diameter with (1) averaged 475 ha\(^{-1}\) for mid-successional forests and 234 ha\(^{-1}\) for primary forests, representing 0.8 to 20% of the total plot biomass. As shown in Figure 2.7, these results are consistent with those of stands where trees 5–10 cm diameter were actually measured in the field. For young secondary forests, Figure 2.7 suggests that the contribution of trees in this class is a linear function of the stand biomass, decreasing rapidly from about 70 to 15% as biomass increases from near 0 to 50 Mg ha\(^{-1}\). For stands with biomass greater than 50 Mg ha\(^{-1}\), the contribution of trees 5–10 cm diameter declines exponentially from an initial value of 12%, leveling off at about 1.4% after ~280 Mg ha\(^{-1}\).

The estimated biomass change for the 3-year period between field and GLAS measurements ranged from 7% in the oldest SF stand (~27 years) to 97% in the youngest stand (~4 years). As suggested by Figure 2.2, biomass accumulation rates varied nonlinearly with stand age (from a low of 0.6 to a maximum of 6.6 Mg ha\(^{-1}\) yr\(^{-1}\)), with the highest rates observed for stands 10 to 14 years old.
2.4.4 Biomass error

The contribution of the different error sources to the overall uncertainty in the field biomass is summarized in Table 2.4. Figure 2.8 explores the calculated sensitivity of our binomial approach in (8), showing the dependence of the co-location error on the spatial overlap between field and GLAS samples. The gray and black lines represent the average co-location error for secondary and primary forests, respectively, when the overlap is artificially changed from 0 to 100%. When the overlap is zero, the binomial model yields an average co-location error of 29% of the estimated biomass for SF plots and an error of 42% for PF plots. These errors decrease slowly (and almost linearly) as the overlap increases from 0 to about 60% overlap, and then converge rapidly to zero as the overlap approaches 100%. On average, overlaps ≥ 75% are needed in primary forests to attain co-location errors not exceeding 20%. In secondary forests, this same level of co-location error can be achieved with overlaps ≥ 50%. The estimated overlap between GLAS and our field plots ranged between 50 and 91%, except for one secondary stand where the overlap was zero—the plot missed the GLAS footprint by about 26 m. The resulting co-location errors (σ_C) were typically 13–26% and dominated the overall uncertainty in both mid-successional and primary stands (Table 2.4).

Uncertainties in diameter (~2%), height (~15%), and wood density (~14%) resulted in a median error of 25% in the mass of individual trees. Nonetheless, this error dropped to only about 6% when scaled to the plot level (σ_M, Table 2.4). In secondary forests, the alternative allometric equations of Brown (1997) and Chambers et al. (2001) (M_{A1} and M_{A2}, Table 2.1) overestimated the Chave-based plot biomass by an average of 29 and 54%, respectively. In primary forests, the Brown equation showed no systematic bias, whereas the Chambers equation resulted in slight underestimates in high-biomass stands (~11%). The errors associated with the choice of the allometry (σ_S) were typically 5–11%, similar to the errors induced by uncertainties in model parameters (σ_A). The individual contributions of measurement and allometric errors to the final uncertainty were generally below 15%, and slightly lower in secondary forests than in primary forests (Table 2.4).
To gain a better understanding of the uncertainty associated with allometric methods, we also compared our reference biomass estimates\(^4\) with four alternative estimates produced with the Chave equations, but making small changes in the input data to illustrate the variability that can be expected when common, suboptimal field data sets are used: Use of the Chave equation \(M_{T2}\) (Table 2.1) with a regional average wood density of 0.667 g cm\(^{-3}\) (Chave et al., 2006), as opposed to taxon-specific densities, resulted in overestimation of biomass values by about 23% in secondary forests and no bias in primary forests. When \(M_{T2}\) was applied using taxon-specific wood densities, but heights derived from a regional height-diameter relationship (Feldpausch et al., 2011), the plot-level biomass was 11% lower on average due to a negative bias in height of 2.6 m. Use of the Chave equation without the height term (\(M_{A3}\), Table 2.1) resulted in plot-level biomass \(\sim 20\%\) higher, regardless of the successional status. When this equation was applied using the regional average wood density of 0.667 g cm\(^{-3}\), the discrepancies in secondary forests were higher still (48%). We should note that the biomass of cecropias and palms, estimated by the specific equations \(M_{T1}\), \(M_{P1}\), and \(M_{P2}\) (Table 2.1), was held constant across all comparisons. Although this introduced some dependence across biomass estimates, these species typically accounted for only about 3% of the total plot biomass.

In primary forests, where the minimum measured diameter was 10 cm, the error of estimating biomass for the 5–10 cm diameter class (\(\sigma_{5-10}\)) contributed less than 1% to the total variance and could safely be neglected. However, this error was about seven times larger in mid-successional forests, being comparable to other sources in magnitude (Table 2.4). In secondary forests, the projection of biomass values backward in time three years induced errors (\(\sigma_G\)) of the order of 7–19%. This term dominated the uncertainties in early successional stands, accounting for about half of the total variance on average, and represented the second largest component in mid-successional stands (Table 2.4). The dependence of \(\sigma_G\) on the temporal difference between field and remote sensing acquisitions is illustrated in Figure 2.9 for SF plots of different ages. As expected, \(\sigma_G\) increases significantly with the time gap in data.

\(^4\) Obtained with the Chave equation \(M_{T2}\) (Table 2.1) using field-measured \(D\) and \(H_T\), and taxon-specific \(\rho\) derived from the literature.
acquisition. The increase is faster for younger forests, which display higher values of $\sigma_G$ than older forests at any given temporal interval (the greater the relative change in biomass, the greater the uncertainty in the model estimate).

The overall uncertainty in the field biomass was typically 25% (for both secondary and primary forests), but ranged from 16 to 53%. Measurement ($\sigma_M$), allometric (both $\sigma_A$ and $\sigma_S$), and co-location ($\sigma_C$) errors increased significantly with plot biomass, at rates of 8, 10, 11, and 22%, respectively (Figure 2.10). The error of estimating biomass for the 5–10 cm diameter class (not shown in the figure) increased significantly with biomass in secondary forests (at a rate of 10%), but showed no trend in old-growth forests. The growth model error showed no linear trend with biomass, but was a straight-line function of the biomass accumulation rate, increasing by about 0.7 Mg ha$^{-1}$ for each unit increase in the growth rate. As a result of the above trends, the overall uncertainty also increased with biomass, at a combined rate of 28%. However, there was no evidence that the mean errors or rates of error increase differed among forest types, after accounting for differences in biomass.

2.5 DISCUSSION

2.5.1 Precision of individual tree measurements

Tree diameter measurements are not difficult to obtain, involve limited subjectivity, and can usually be made with a high degree of precision (e.g., Barker, Bollman, Ringold, Sackinger, & Cline, 2002; Elzinga, Shearer, & Elzinga, 2005; Kitahara, Mizoue, & Yoshida, 2009; McRoberts, Hahn, Hefty, & Cleve, 1994). The small variation observed in this study (RMSD = 0.8 cm or 1.8%) is consistent with previous findings and likely resulted from divergences in tape placement, with repeated measurements taken at slightly different heights or angles. Other potential sources of variation include mistakes reading the tape, recording error, and data entry error, all of which are difficult to detect if the resulting values are not particularly unusual.

Despite the obvious subjectivity associated with ocular height estimates (both $H_C$ and $H_T$), they were surprisingly precise, with a combined RMSD of only 2 m. For
comparison, Kitahara et al. (2009) reported nearly the same level of precision (1.8 m) for repeated height measurements made with a modern ultrasonic hypsometer (Haglöf Vertex) in less dense temperate forests with relatively lower structural complexity. In a recent study also conducted at Tapajós, Hunter et al. (2013) obtained a precision of 4.7 m for heights obtained with a clinometer and a measuring tape, keeping angles below 50° and correcting for slope to minimize measurement error. Factors contributing to variation in our height measurements include difficulty in determining the location of the treetop due to occlusion by surrounding vegetation, and disparity in the perception of where the base of the crown is located. In the following section we discuss the ocular height estimates in more detail and provide a possible explanation for the unexpected high precision.

Variation was considerably higher for crown measurements. Crown depth, calculated as the difference between $H_T$ and $H_C$, was very similar in absolute precision to the ocular height estimates. However, while a RMSD of 2 m typically represents a relatively small percentage of a tree height, it corresponds to a large fraction of a typical crown depth measurement (6 m on average for the trees sampled in this study). The precision of crown radius measurements was better than 1 m, but represented ~26% in relative terms. These measurements required some level of personal judgment and were affected by visibility restrictions in ways similar to the height estimates. In addition, crown spread was typically 25–45% of the tree height and the resulting high levels of crown overlap among trees made it often challenging to identify the correct branches for the measurement. We note that the horizontal position of the crown edge is somewhat difficult to determine from directly below and suggest that the precision of crown radius measurements would likely be improved by sighting the edges along a clinometer held at a 90 degree angle. In terms of height measurements (and the derived crown depth), uncertainties may be reduced with the aid of a telescoping height measuring pole. Although not necessarily practical, the pole could be used to obtain direct height measurements for small trees.

$$^5$$ Note that the errors in $H_T$ and $H_C$ did not simply add in the calculation of $C_D$. Errors in $H_T$ were somewhat compensated by similar errors in $H_C$ as a result of the correlation between these two variables.
(up to 10–15 m), and serve as a height reference for the ocular estimation of taller
trees.

Not surprisingly, measurements were more precise for small trees than for large
trees (most sources of measurement variation become more pronounced as tree size
increases). Although measurement variation generally increased with the dimension
of the measurement, the magnitude of this effect differed substantially among
attributes, with stem diameter showing the lowest rate of increase, followed by
height, and crown dimensions. For tree height, Hunter et al. (2013) observed an
eightfold increase in measurement variation (from 1.1 to 8.2 m) after dividing the
data into four diameter classes with an equal number of trees. This contrasts sharply
with the less than twofold increase observed in this study (from 1.8 to 2.9 m),
indicating that the precision of the ocular height estimates was not only high, but also
displayed a relatively low, yet statistically significant, dependence on tree height.

Because precision is not constant across the range of diameters and heights, it is
important to account for this variation when propagating measurement errors to
determine the uncertainty in biomass. The standard deviation of the differences
between repeated measurements, calculated by quartiles of the ranked set of
measurements, is provided in Table 2.A1 for reference.

While measurement uncertainty was generally not negligible, with precision
clearly declining with increasing tree size, we found no systematic errors. In addition,
we found no differences in precision (or rates of decline in precision with increasing
tree size) between primary and secondary forests, after accounting for tree size. This
suggests that measurement precision was fairly robust to changes in measurement
conditions (induced by changes in stem density, species composition, leaf area index,
etc.), with divergences in overall precision being largely attributable to differences in
tree size distribution (see Figure 2.6). We stress that the results presented here refer
strictly to reproducibility of measurements, and that no reference is made to the
agreement of those measurements with the true, unknown values (i.e., accuracy).
2.5.2 Height calibration model

In addition to being precise, ocular height estimates agreed closely with laser measurements. While $H_C$ was visually estimated with no detectable bias, we found a small bias in the estimation of $H_T$, likely reflecting the reduced visibility conditions associated with this measurement. We observed a slightly lower agreement between ocular and laser measurements for trees measured in primary forests compared to trees of the same size located elsewhere (i.e., in open areas, secondary forests, or selectively logged forests). Because the precision of ocular estimates did not vary with forest type, this effect is probably attributed to variation in laser measurements, which become increasingly more difficult to obtain as canopy density and complexity increase (Hunter et al., 2013; Larjavaara & Muller-Landau, 2013).

The model parameters were very similar for the two observers, despite their different levels of experience and professional background. While observer 1 is a senior field technician who works full time performing forest inventories, observer 2 is a researcher (J. R. Santos) who typically conducts one forest inventory per year. Both individuals have over 25 years of experience. Interestingly, there were systematic differences between observers in the ocular estimation of $H_C$, but not $H_T$. The parallel lines model in (9) accounted for such differences and produced unbiased estimates of laser height, with an error of about 2 m. This figure matched closely the measurement uncertainty associated with the ocular heights (Table 2.2), indicating that the model error was essentially determined by the precision of the ocular estimates. We note that although methods are available to account for measurement errors in the explanatory variable in models like (9) (Frost & Thompson, 2000), they are neither necessary nor appropriate when the purpose of the model is prediction, and predictions are based on observations that have the same imprecision as the ones used in the model fit (Ramsey & Schafer, 2002).

In agreement with a previous study (Gonçalves & Santos, 2008), the results suggest that ocular heights can be statistically corrected to obtain accurate estimates consistent with laser heights. The caveat is that laser heights are themselves only estimates of the actual heights. In three recent studies conducted in moist tropical forests, measurements from handheld instruments based on the tangent method were
compared to direct height measurements obtained by climbing, and were shown to produce unbiased (or minimally biased) height estimates (Curto, Fernandes, Soares, Martins, & David, 2013; Larjavaara & Muller-Landau, 2013; Silva, Curto, Soares, & Piassi, 2012). However, there was little agreement among and within studies with respect to the random error. While Curto et al. (2013) reported scatters of 0.2–0.8 m for measurements obtained with either an ultrasonic hypsometer (Vertex) or an electronic clinometer in an Atlantic forest in southeastern Brazil, Silva et al. (2012) found errors of ~4 m for a site in the same region, using measurements taken by the same technician as Curto et al. (2013), with the same hypsometer (R. A. Curto, personal communication). Larjavaara & Muller-Landau (2013) observed errors with a laser rangefinder in central Panama that ranged from 2.4 to 11.3 m, depending on the technician who performed the measurements. Taken together, these studies suggest that the degree of accuracy obtained with the tangent method in tropical forests has little dependence on the type of instrument used, but can vary substantially with both operator and site. Because a trained operator took the laser measurements in this study in a careful and unhurried way, we expect errors to be relatively small, perhaps falling at the low end of the range described above. In fact, we made a few direct height measurements in the field for a quick assessment of the laser accuracy and observed a total error of only 0.2 m. Nevertheless, the sample was nonrepresentative and prohibitively small to allow any generalization.

In addition to determining the accuracy of instrument measurements of tree height, Curto et al. (2013) and Silva et al. (2012) examined the accuracy of ocular height estimates. Surprisingly, they found ocular estimates to perform similarly or only slightly worse than instrument measurements, although some bias was detected in the study of Curto et al. (2013). When observers had their judgment “recalibrated” by being informed of the actual height of the trees after providing their estimates, the random error was reduced substantially (becoming smaller even than that of instrument measurements) and bias was not an issue. In the present study, ocular estimates were checked periodically in the field with the laser rangefinder and the observers most likely used this information to recalibrate their estimates for subsequent trees. This contributed not only to the close agreement between ocular and
laser measurements, but also to the high precision associated with the ocular estimates.

Ideally, accurate instrument measurements should be taken for all sampled trees. In practice, however, this approach is time-consuming and not usually cost effective as it involves measuring the horizontal distance to each tree from a position that allows a sufficiently clear view of both the base and the top of the tree, while standing at a distance of at least one tree height away (Larjavaara & Muller-Landau, 2013)—a challenge in dense, structurally complex tropical forests. Whereas regional equations for estimating tree height from measured diameter are available for tropical forests (Feldpausch et al., 2011), height-diameter relationships can vary significantly from site to site and the resulting biomass estimates can be severely biased (see section 2.4.4 and Hunter et al., 2013). Although site-specific equations based on a relatively small number of samples can yield better results (Hunter et al., 2013), the errors may still be prohibitively large, particularly for small plots. For the 0.25-ha plots used in this study, for example, we found that height-diameter allometries developed using up to 20% of the data can result in biomass estimation errors of more than 16% on average for primary forests (Figure 2.A1).

The ocular method presented here offers a convenient alternative that can produce useful results. Instead of obtaining instrument measurements for all sampled trees, heights can be quickly estimated by eye by experienced observers—over 10 times faster\(^6\), as measured by Curto et al. (2013)—and calibrated against either actual heights or unbiased instrument measurements obtained for a small subset of the trees. The ability to estimate tree height by eye can be developed with practice, being fairly common among seasoned field technicians. While accepted as a common practice in forest inventories when precision is not critical (Husch, Beers, & Kershaw, 2002), it has also been used in established, large-scale inventory programs such as the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service, resulting in acceptable levels of uncertainty (Bechtold, Zarnoch, & Burkman, 1998). Although the reliability of ocular height estimates will obviously depend on the observer, even

---

\(^6\) Because ocular estimation does not involve measurement of horizontal distances, it increases the freedom of the observer to locate a clear view of the tree top (often from shorter distances to the tree than one tree height), resulting in significant savings of time.
relatively inexperienced field technicians should be able to produce useful estimates if they are systematically provided with a reference height (cf. Curto et al., 2013; Silva et al., 2012). We recommend that systematic checks be made at each plot by instrument measurements, regardless of observer experience. These checks can be used as a reference by observers in the field, and also for the calibration of equations like (9).

2.5.3 Biomass estimation and its error

Our results show that co-location error, defined in this study as the uncertainty in the biomass estimate resulting from the spatial disagreement between field and lidar samples (i.e., field plots including trees not captured by GLAS and/or excluding trees that were actually captured), accounts for a substantial portion of the total error. In agreement with our findings for stands at La Selva Biological Station, Costa Rica (Treuhaft et al., 2010), co-location error dominated the overall uncertainty in the field biomass, except in early-successional forests where the application of the growth model resulted in larger errors on average (Table 2.4). The results illustrated in Figure 2.8 are consistent with the expectation of lower co-location error with increasing lidar/field overlap, and also of lower errors for secondary forests compared to primary forests, given their lower species diversity and more homogeneous canopy structure (cf. Figure 2.6 and Chapter 3). We should note that the binomial approach in (8) assumes that the tree size-frequency distribution of each individual zone depicted in Figure 2.3 (i.e., F, I, and G) is similar to that observed for the full 50 m x 50 m field plot (F ∪ I). This explains the relatively low maximum values of co-location error in Figure 2.8 when the overlap is zero. We should also note that although the binomial approach is presented here using GLAS as an example, it depends exclusively on the field data and on the amount of overlap between the field and the remote sensing samples, and thus could be applied regardless of the remote sensing data type.

In a previous study in a tropical rainforest in Hawaii, Asner, Hughes, Varga, Knapp, & Kennedy-Bowdoin (2009) found that misalignment of lidar and field data introduced errors in biomass estimates of only 0–10 Mg ha\(^{-1}\). While differences in floristic composition, vegetation structure, and plot size make direct comparisons
between studies difficult, differences in methodology most likely account for much of the observed discrepancy. In their study, Asner et al. (2009) estimated the co-location error by varying the location and size of the lidar “plots” by small amounts (10%); regressing the lidar metrics obtained for each new location/size against the (fixed) field-measured biomass; and determining the variation in the biomass predictions resulting from variation in the lidar metrics. This is conceptually different from the approach used in this study, where the field-based estimate of biomass was varied instead. This distinction is important because inclusion/exclusion of trees in the calibration plots (particularly big trees) can cause large changes in the field estimate of biomass that may not be proportionally reflected in the vertical structure captured by the lidar (and in turn, in the lidar estimate of biomass). As shown by Mascaro, Detto, Asner, & Muller-Landau (2011), even relatively small changes in field-based estimates of biomass in tropical forests (as a result of accounting for portions of trees that fall outside the plot boundary) can have a significant impact on the relationship with lidar metrics, accounting for as much as 55% of the error associated with lidar-biomass models.

In comparison to the co-location error, measurement and allometric errors were relatively small. Uncertainties in height and wood density values were large relative to the uncertainty in diameter and contributed the bulk of the uncertainty in biomass due to measurement variation. This measurement error was fairly large at the tree level, but decreased significantly at the plot scale because measurement variation was unbiased (Table 2.2) and tree-level errors were added in quadrature to produce realistic plot-level estimates.

The overestimation of the reference, Chave-based biomass in secondary forests by the equations of Brown (1997) and Chambers et al. (2001) ($M_{A1}$ and $M_{A2}$, Table 2.1) was largely explained by the omission of wood density information in the models. These alternative mixed-species equations were derived from primary forest trees, which tend to have much denser wood than the secondary forest trees to which they were applied (cf. Table 2.3 and Nelson et al., 1999). When we corrected $M_{A1}$ and $M_{A2}$ by including a dependence on wood density as in Chave et al. (2004), the overestimation of the reference biomass in secondary forests decreased by a factor of
3 and 2, respectively. From our tests with the Chave equations with and without height, we would expect these differences to decrease by an additional ~20% if $M_{A1}$ and $M_{A2}$ also included a dependence on tree height, and if tree allometry is somewhat conserved across moist tropical sites as indicated by Chave et al. (2014). Thus, most of the variation captured by $\sigma_S$ was apparently due to the use of allometric equations, which differed with respect to the inclusion of height and wood density information.

As with the measurement error, allometric errors were assumed to be uncorrelated and decreased significantly (by a factor of 3–4) when scaled to the plot level. While this assumption seems reasonable for $\sigma_A$ (cf. Ahmed et al., 2013; Chave et al., 2004; Phillips et al., 2000) given the random nature of the regression errors (assumed to be normally distributed with mean zero), one could argue that the error due to the choice of the allometric equation ($\sigma_S$) is systematic and unlikely to be independent (trees with similar diameter, for example, can have nearly the same $\sigma_S$). One way of testing if the sum in quadrature is appropriate is to calculate $\sigma_S$ directly at the plot level by taking the standard deviation of the plot-level biomass estimates obtained with the alternative equations. This resulted in a median error of 13% for our 30 plots, which is only slightly higher than that obtained by adding tree-level errors in quadrature (Table 2.4). The error for primary forests alone was virtually the same (9%), confirming the tendency of individual tree errors to offset each other when combined to generate plot-level estimates. We note that Keller et al. (2001) observed the same level of error (10%) when estimating the biomass density (trees $\geq$ 15 cm diameter) of an area of 392 ha at Tapajós using four alternative equations (one of which is $M_{A1}$). A similar error (13%) was also reported by Chave et al. (2004) for estimates obtained with eight different equations in a 50 ha plot in Panama, after correcting for variation in wood density.

The estimation of biomass for the modeled diameter class of 5–10 cm made a negligible contribution to the final uncertainty in primary forests, but represented a significant source of error in mid-successional forests. This is not surprising when one considers the rapid decrease in the contribution of trees 5–10 cm diameter to biomass with increasing biomass, as observed in Figure 2.7. Because in young secondary forests ($< 50$ Mg ha$^{-1}$) trees 5–10 cm diameter can account for as much as
70% of the aboveground biomass, modeling the stem frequency of this class with (1) and estimating biomass for the modeled class has the potential to introduce large errors. For these young forests, we recommend that trees 5–10 cm diameter be directly measured (or tallied) in the field.

The biomass accumulation rates of secondary forests estimated with the site-specific growth model of Neeff & Santos (2005) agreed well with rates from a long-term study of SF regrowth in the central Amazon (Feldpausch, Prates-Clark, Fernandes, & Riha, 2007), as well as with a more general estimate for tropical moist forests based on data from a number of sites (Silver, Ostertag, & Lugo, 2000b). Yet, the error associated with this model dominated the overall uncertainty in the biomass of the youngest stands and represented the second largest contribution in older secondary forests. In (2), stand biomass is defined as the product of basal area and top height (the two state variables that are explicitly modeled), and most of its uncertainty results from the relatively poor fit of the basal area model (Figure 2.2). Although site-specific, the model is based on growth rates inferred from a chronosequence, which may differ substantially from actual growth rates due to stand differences in land-use history (e.g., old agricultural fields vs. abandoned pastures) (Feldpausch et al., 2007). This probably explains most of the variability observed by Neeff & Santos (2005) in basal area for a given age, and the relatively large uncertainty associated with the model parameters.

The estimated biomass change of over 30% per year for the youngest secondary forests emphasizes the importance of accounting for temporal differences between field and remote sensing data. Here, we assumed no biomass change for primary forests, but these changes have been determined to be small at Tapajós (~1% per year) relative to the average biomass (Pyle et al., 2008; Rice et al., 2004; Vieira et al., 2004). We would expect the resulting errors to be negligible, unless some disturbance occurred between observation epochs (e.g., logging, fire, windstorms, etc.). This is generally detectable in the field and/or in remote sensing data (Frolking et al., 2009), and neither data source revealed the occurrence of a significant disturbance event during the period of this study.
The overall uncertainty in the plot-level biomass observed in this study (typically 20–34%) was similar in relative size to that observed by Treuhaft et al. (2010) for a tropical wet forest in Costa Rica (26–31%) where measurement, allometric, and co-location errors were also considered. It is also in line with the total plot-level error observed by Chave et al. (2004) (~23–27%, as calculated from their Table 2.3), although different methods were used for the assessment of uncertainty. Recent estimates of biomass from lidar (Chapter 3) and InSAR (Treuhaft et al., 2015) observations for the plots used in this study showed RMS errors about the field-estimated biomass of 20–35%. The uncertainty in the field biomass can therefore represent a dominant term in the overall error budget for biomass maps derived from remote sensing, and should be taken into account in calibration and validation efforts.

Because co-location and temporal errors have the potential to account for a large fraction of the total variance (~70% on average for the plots in this study, but as much as 94% in individual plots), they emerge as obvious targets for reducing uncertainty in studies relating tropical forest biomass to remotely sensed data. Temporal errors can be minimized by conducting field campaigns as close as possible in time to the remote sensing data acquisition. In cases where this is not possible (e.g., when using historical data), our results underscore the importance of selecting an appropriate growth model to account for biomass change in secondary forests, and of quantifying the uncertainty associated with this model. Reducing co-location errors requires not only the acquisition of high-precision differential GPS measurements for plot locations (often complemented with topographic surveying in closed-canopy forests), but also that field and remote sensing samples agree as much as possible in size, shape, and orientation. A simple statistical approach such as the one presented in (8) can be used to account for errors due to partial overlap.

Finally, we note that although measurement and allometric errors were relatively unimportant when considered alone, combined they accounted for roughly 30% of the total variance on average (as much as 64% in individual plots) and therefore should not be ignored. Steps can be taken to reduce uncertainties in height and wood density, as well as in allometric equations. However, reducing co-location and temporal errors may be a more cost-effective solution for reducing the overall
uncertainty when resources are limited. For instance, the total error in Table 2.4 would drop by nearly half if co-location and temporal errors were zero, but only by about 20% if we disregard measurement and allometric errors instead.

2.6 ACKNOWLEDGEMENTS

The research described in this paper was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration, under the Terrestrial Ecology program element. F.G. Gonçalves was partially funded by the CAPES Foundation, Brazilian Ministry of Education, through the CAPES/Fulbright Doctoral Program (process BEX-2684/06-3). The authors would like to thank Brazil’s Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/MCTI, Scientific Expedition process 010301/2009-7) and Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio/MMA, SISBIO process 20591-2) for research authorizations, and the Santarém office of the Large Scale Biosphere-Atmosphere Experiment in Amazonia (LBA) for providing logistical support. They would also like to thank E. Oliveira (UFAC) and the local assistants J. Oliveira, R. Santos, I. Silva, and E. Pedroso for the invaluable help with the field acquisitions.

References


Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., … Vieilledent, G. (2014). Improved allometric models to estimate the


Hunter, M. O., Keller, M., Victoria, D., & Morton, D. C. (2013). Tree height and
Wiley & Sons.
National Forest, Brazil examination of sampling and allometric uncertainties.
Japanese National Forest Inventory. *Environmental Monitoring and Assessment,
159*(1-4), 331–40.
comparison of two common field methods in a moist tropical forest. *Methods in
Ecology and Evolution, 4*(9), 793–801.
uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sensing of
Environment, 115*(12), 3770–3774.
forest inventory field measurements. *Canadian Journal of Forest Research,
24*(9), 1766–1770.
Molto, Q., Rossi, V., & Blanc, L. (2013). Error propagation in biomass estimation in
Neeff, T., & Santos, J. R. Dos. (2005). A growth model for secondary forest in
Nelson, B. W., Mesquita, R., Pereira, J. L. , Garcia Aquino de Souza, S., Teixeira
estimate of secondary forest biomass in the central Amazon. *Forest Ecology and
Nogueira, E. M., Fearnside, P. M., Nelson, B. W., Barbosa, R. I., & Keizer, E. W. H.
(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*(11), 1853–1867.
analysis of large-scale forest carbon budgets. *Global Ecology and Biogeography,
9*(4), 305–313.
Pyle, E. H., Santoni, G. W., Nascimento, H. E. M., Hutyra, L. R., Vieira, S., Curran,
two Amazonian forests. *Journal of Geophysical Research, 113*, G00B08.
Learning.


Figure 2.1. Geographical location of the Tapajós National Forest, PA, Brazil, outlined in yellow. The gray lines are GLAS tracks from 2003–2009, and the blue targets are the locations of the plots sampled in this study. The pictures on the right illustrate three of the stands where plots were located, with aboveground biomass ranging from near zero (bottom) to over 400 Mg ha⁻¹ (top).
Figure 2.2. Variation in basal area ($B$), top height ($H_{\text{TOP}}$), and aboveground biomass (AGB) with stand age ($t$) at Tapajós, as described by the growth model of Neeff & Santos (2005). The solid lines represent structural values at age $t$, and the dashed line shows the expected AGB at the time of the GLAS acquisition (3 years prior) for a stand of age $t$. The gray bands for $B$ and $H_{\text{TOP}}$ are 95% confidence intervals from a Monte Carlo simulation (Neeff & Santos, 2005).
Figure 2.3. Schematic diagram illustrating an arbitrary intersection of a GLAS footprint (ellipse) and a field plot (square), portrayed in their correct relative nominal sizes and shapes. Co-location error was introduced due to differences in the areas sampled by each technique (G U I for GLAS vs. F U I for field). The gray zone “I” represents the area covered effectively by both techniques and averaged 75% of the footprint area for the 30 plots used in this study.
Figure 2.4. Differences between repeated measurements of (a) diameter, (b) height to the base of the live crown, (c) total height, (d) crown depth, and (e) crown radius, ordered by the magnitude of the measurement (estimated as the average of the two measurements). The side-by-side boxes in panel (f) represent the middle 50% of the distributions, with medians marked by a thick black line. The whiskers extend to the smallest and largest differences not more than 1.5 box-lengths away from the box, and the dots represent extreme values.
Figure 2.5. (a) Log-log scatterplot of laser-measured heights versus ocular heights, with points coded by height type ($H_C =$ circles, $H_T =$ triangles) and observer ($O_1 =$ filled symbols, $O_2 =$ open symbols). The parallel lines represent the fit of the regression model in (9), which includes indicator variables to account for three height type-observer combinations ($H_CO_1 =$ gray solid line, $H_CO_2 =$ black solid line, and $H.TO_1/H.TO_2 =$ dotted line). (b) Observed versus predicted laser heights, with 95% confidence intervals.
Figure 2.6. Average stem density (bars) and tree height (circles) per diameter class (trees ≥ 10 cm) for secondary (dark gray) and primary (light gray) forests. The solid lines show the fit of (1) to the average diameter distributions, and the dashed lines show the fit of a similar exponential decay model to the tree height data.
Figure 2.7. Relationship between stand biomass at Tapajós (trees $\geq 5$ cm diameter) and the fraction of that biomass found in trees 5–10 cm diameter. The symbols in black represent plots measured in 2000, as described by Santos et al. (2003).
Figure 2.8. Dependence of the co-location error estimated with the binomial approach on the spatial overlap between GLAS and field measurements. The lines represent the mean error for secondary (gray) and primary (black) stands when the overlap is artificially changed from 0 to 100%. The gray band and the error bars represent ±1 SD.
Figure 2.9. Dependence of the biomass error introduced by the growth model of Neeff & Santos (2005) on the temporal difference between field and remote sensing acquisitions (i.e., remote sensing data acquired 1, 2, 3, …, 10 years prior to the field measurements). The dependence is illustrated for secondary forests of different ages, as indicated by the labels to the right of the lines.
Figure 2.10. Stand biomass versus measurement ($\sigma_M$), allometric ($\sigma_A$ and $\sigma_S$), and co-location ($\sigma_C$) errors, with the estimated regression lines. Sources of error that are not common to all plots (i.e., $\sigma_{5-10}$ and $\sigma_G$) were omitted from the figure for clarity. By looking at points intersected by an imaginary vertical line at any level of biomass in the x-axis, one can see the relative contribution of the different error sources for the plot represented by that biomass.
Table 2.1. Allometric equations used to calculate individual tree biomass at Tapajós.

<table>
<thead>
<tr>
<th>Category</th>
<th>Equation*</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trees</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cecropia spp.</strong></td>
<td>$M_{T1} = \exp(-2.5118 + 2.4257 \ln(D))$</td>
<td>Nelson et al. (1999)</td>
</tr>
<tr>
<td>All others</td>
<td>$M_{T2} = \exp(-2.977 + \ln(\rho D^2 H_T))$</td>
<td>Chave et al. (2005)</td>
</tr>
<tr>
<td><strong>Palms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Attalea spp.</strong></td>
<td>$M_{P1} = 63.3875 H_C - 112.8875$</td>
<td>Anderson (1983)</td>
</tr>
<tr>
<td>All others</td>
<td>$M_{P2} = \exp(-6.379 + 1.754 \ln(D) + 2.151 \ln(H_T))$</td>
<td>Saldarriaga et al. (1988)</td>
</tr>
<tr>
<td><strong>Alternative equations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M_{A1} = \exp(-2.134 + 2.530 \ln(D))$</td>
<td>Brown (1997)</td>
<td></td>
</tr>
<tr>
<td>$M_{A2} = \exp(-0.370 + 0.333 \ln(D) + 0.933 \ln(D)^2 - 0.122 \ln(D)^3)$</td>
<td>Chambers et al. (2001)</td>
<td></td>
</tr>
<tr>
<td>$M_{A3} = \rho \exp(-1.499 + 2.148 \ln(D) + 0.207 \ln(D)^2 - 0.028 \ln(D)^3)$</td>
<td>Chave et al. (2005)</td>
<td></td>
</tr>
<tr>
<td>$M_{A4} = \exp(-3.1141 + 0.9719 \ln(D^2 H_T))$</td>
<td>Brown et al. (1989)</td>
<td></td>
</tr>
</tbody>
</table>

* $M$ (kg) is the oven-dry aboveground tree biomass, $D$ (cm) is the diameter at breast height (1.3 m), $H_T$ (m) is the total height, $H_C$ (m) is the height to the base of the live crown, and $\rho$ (g cm$^{-3}$) is the wood density measured as oven-dry weight over green volume.
Table 2.2. Summary statistics of differences between repeated measurements of diameter (D), height to the base of the live crown (H_C), total height (H_T), crown depth (C_D), and crown radius (C_R) for trees sampled at Tapajós. Statistics include total error (RMSD), systematic error (mean), and random error (SD), in both absolute and relative terms, as described in section 2.3.4.1. The number of observations was 104, except for C_R (n = 144).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Range</th>
<th>Differences</th>
<th>% that is:</th>
<th>0</th>
<th>≤ 10%</th>
<th>≤ 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSD</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D (cm)</td>
<td>5.5–110.5</td>
<td>0.8</td>
<td>(1.8%)</td>
<td>0.1</td>
<td>(0.5%)</td>
<td>0.8</td>
</tr>
<tr>
<td>H_C (m)</td>
<td>1.5–31.0</td>
<td>1.8</td>
<td>(17.7%)</td>
<td>0.1</td>
<td>(0.6%)</td>
<td>1.8</td>
</tr>
<tr>
<td>H_T (m)</td>
<td>5.0–40.0</td>
<td>2.3</td>
<td>(15.2%)</td>
<td>-0.2</td>
<td>(-1.7%)</td>
<td>2.3</td>
</tr>
<tr>
<td>C_D (m)</td>
<td>1.0–20.0</td>
<td>1.8</td>
<td>(30.7%)</td>
<td>-0.3</td>
<td>(-4.8%)</td>
<td>1.8</td>
</tr>
<tr>
<td>C_R (m)</td>
<td>0.7–8.0</td>
<td>0.8</td>
<td>(25.7%)</td>
<td>0.0</td>
<td>(0.0%)</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Table 2.3. Characteristics of secondary (SF), selectively-logged (PF_L), and primary (PF) forest stands used in this study. Field plots were 0.25 ha in size (50 m x 50 m). The minimum diameter was 10 cm, except for seven early successional stands where the minimum was 5 cm. Values are median and range (in parentheses).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>SF (14 plots)</th>
<th>PF_L (8 plots)</th>
<th>PF (8 plots)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of species (0.25 ha⁻¹)</td>
<td>28 (7–45)</td>
<td>41 (36–49)</td>
<td>42 (31–45)</td>
</tr>
<tr>
<td>Stem density (trees ha⁻¹)</td>
<td>488 (132–1052)</td>
<td>380 (340–456)</td>
<td>354 (304–424)</td>
</tr>
<tr>
<td>Basal area (m² ha⁻¹)</td>
<td>8.7 (1.5–16.9)</td>
<td>22.8 (18.9–31.3)</td>
<td>24.8 (15.9–30.4)</td>
</tr>
<tr>
<td>Mean height (m)</td>
<td>12.9 (7.1–19.9)</td>
<td>19.7 (17.2–21.8)</td>
<td>18.0 (13.5–22.1)</td>
</tr>
<tr>
<td>Mean wood density (g cm⁻³)</td>
<td>0.50 (0.37–0.54)</td>
<td>0.65 (0.54–0.70)</td>
<td>0.62 (0.59–0.63)</td>
</tr>
<tr>
<td>Biomass (Mg ha⁻¹)</td>
<td>37.3 (1.9–130.1)</td>
<td>285.8 (183.0–423.6)</td>
<td>293.0 (162.6–417.5)</td>
</tr>
<tr>
<td>Mean crown depth (m)</td>
<td>5.2 (2.9–8.3)</td>
<td>7.0 (6.4–8.2)</td>
<td>7.0 (5.5–9.5)</td>
</tr>
<tr>
<td>Mean crown radius (m)*</td>
<td>2.3 (0.8–3.0)</td>
<td>3.0 (2.2–3.5)</td>
<td>3.1 (2.2–4.1)</td>
</tr>
</tbody>
</table>

* Estimated from trees located in a central 12.5 m x 50 m subplot.
Table 2.4. Uncertainties in field-based estimates of plot biomass at Tapajós. The error is presented in terms of the median value and the interquartile range (in parentheses) of the relative errors of all applicable plots. The last three columns give the percentage of variance in the biomass estimate which is due to each error source. Values are means for early-successional (SF\textsubscript{early}), mid-successional (SF\textsubscript{mid}), and primary (PF/PFL) forest plots.

<table>
<thead>
<tr>
<th>Error source</th>
<th>Error (%)</th>
<th>% of total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement ($\sigma_M$)</td>
<td>6.4 (4.3–9.0)</td>
<td>6.4 4.8 7.6</td>
</tr>
<tr>
<td>Allometry (model parameters, $\sigma_A$)</td>
<td>7.5 (5.0–10.2)</td>
<td>3.7 8.2 12.9</td>
</tr>
<tr>
<td>Allometry (model selection, $\sigma_S$)</td>
<td>7.1 (4.8–10.6)</td>
<td>9.4 10.6 14.1</td>
</tr>
<tr>
<td>Co-location ($\sigma_C$)</td>
<td>19.1 (13.0–25.6)</td>
<td>28.3 45.8 65.2</td>
</tr>
<tr>
<td>Trees 5–10 cm diameter ($\sigma_{5-10}$)</td>
<td>1.1 (0.7–3.5)</td>
<td>NA 11.6 0.2</td>
</tr>
<tr>
<td>Growth model ($\sigma_G$)</td>
<td>12.0 (7.0–18.6)</td>
<td>52.2 19.0 NA</td>
</tr>
<tr>
<td>Total</td>
<td>25.4 (20.2–33.9)</td>
<td>100 100 100</td>
</tr>
</tbody>
</table>
2.7 APPENDIX

Table 2.A1. Standard deviations (SD) of differences in repeated measurements of diameter (D), height to the base of the live crown (H_C), total height (H_T), crown depth (C_D), and crown radius (C_R) for trees sampled at Tapajós. Values are shown by quartile of the ranked set of measurements, in both absolute and relative terms (see section 2.3.4.1 for details). The total number of observations was 104, except for C_R (n = 144).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0–25%</td>
</tr>
<tr>
<td><strong>D (cm)</strong></td>
<td></td>
</tr>
<tr>
<td>Quantiles</td>
<td>5.5–12.3</td>
</tr>
<tr>
<td>SD</td>
<td>0.1 (1.4%)</td>
</tr>
<tr>
<td><strong>H_C (m)</strong></td>
<td></td>
</tr>
<tr>
<td>Quantiles</td>
<td>1.5–6</td>
</tr>
<tr>
<td>SD</td>
<td>1 (22.9%)</td>
</tr>
<tr>
<td><strong>H_T (m)</strong></td>
<td></td>
</tr>
<tr>
<td>Quantiles</td>
<td>5–11.4</td>
</tr>
<tr>
<td>SD</td>
<td>1.5 (17.9%)</td>
</tr>
<tr>
<td><strong>C_D (m)</strong></td>
<td></td>
</tr>
<tr>
<td>Quantiles</td>
<td>1–4</td>
</tr>
<tr>
<td>SD</td>
<td>0.9 (36.8%)</td>
</tr>
<tr>
<td><strong>C_R (m)</strong></td>
<td></td>
</tr>
<tr>
<td>Quantiles</td>
<td>0.7–1.6</td>
</tr>
<tr>
<td>SD</td>
<td>0.3 (26.5%)</td>
</tr>
</tbody>
</table>
Figure 2.A1. Biomass estimation error at Tapajós due to the use of a site-specific height-diameter allometry versus the number of trees used to develop the allometry. The height-diameter allometry (log-log form) was fit using an equal number of trees randomly selected from each of four diameter classes, as defined in Table 2.A1. For each sample size displayed on the x-axis, the random selection was repeated 1000 times, resulting in 1000 different allometries and 1000 different biomass estimates for each plot. A RMS error was calculated for each plot using the biomass estimated with field-measured heights as the reference. The lines represent the mean RMSE for secondary (gray) and primary (black) forests, and the gray band and the error bars represent ±1 SD.
Chapter 3. Improving lidar-based estimates of tropical forest biomass: The performance of Fourier versus traditional structural metrics

F.G. Gonçalves a,b,* R.N. Treuhaft c, B.E. Law a, A. Baccini b, W. Walker b, A.Q. Almeida d, J.R. Santos e, M. Keller f, P.M.L.A. Graça g

a Department of Forest Ecosystems & Society, Oregon State University, Corvallis, OR, USA
b Woods Hole Research Center, Falmouth, MA, USA
c Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
d Núcleo de Engenharia Agrícola, Universidade Federal de Sergipe, SE, Brazil
e Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brazil
f USDA Forest Service, International Institute of Tropical Forestry, San Juan, PR, USA
g Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, Brazil

* Corresponding author: fgoncalves@whrc.org
3.1 ABSTRACT

Methods for obtaining accurate, spatially explicit estimates of biomass density in tropical forests are required to reduce uncertainties in the global carbon cycle, and to support international climate agreements and emerging carbon markets. In this study, we use airborne laser scanning (ALS), space-based lidar observations (ICESat/GLAS), and detailed \textit{in situ} measurements made at the Tapajós National Forest, Brazil, to advance methods of aboveground biomass estimation in tropical regions from remotely sensed structure. Lidar and field-based vegetation density profiles obtained for stands spanning a wide range in structure and biomass (2–538 Mg ha\(^{-1}\)) showed good qualitative and quantitative agreement. Simple regressions between corresponding field and lidar metrics showed that GLAS and ALS were similarly sensitive to structural differences at Tapajós, explaining 74–87\% of the variation in field measurements. The RMS differences between field and remote sensing metrics ranged between 11 and 20\%, and were typically smaller than the uncertainty in field-based measurements of individual tree height. A new approach to biomass estimation based on Fourier transforms of lidar profiles resulted in models that explained 93–94\% of the variation in biomass, with a cross-validation RMS error of 35–38 Mg ha\(^{-1}\) (22–24\%). This represented an improvement over the prediction performance of traditional models of 25\% for GLAS and 27\% for ALS. Test of the Fourier approach using data from two different sites in the Amazon shows that the improvement in biomass estimation performance is consistent across sites. We incorporate the Fourier approach into a multi-stage scaling strategy to biomass estimation and demonstrate its performance by generating a wall-to-wall map of biomass for a large area in the Amazon.
3.2 INTRODUCTION

Deforestation and forest degradation have been estimated to contribute 6–20% of total anthropogenic CO₂ emissions, ranking as the second largest source of anthropogenic carbon to the atmosphere, after fossil fuel burning (Le Quéré et al., 2009; van der Werf et al., 2009). However, uncertainties in the net flux of carbon from these activities are high, ranging from 10 to 100% in many developed countries, and possibly exceeding 100% in most developing regions (NRC, 2010). The standard practice to estimating the carbon flux from land-use change rely heavily on passive optical systems such as MODIS or Landsat, essentially linking detected changes in forest area or condition to average values of biomass density estimated from forest inventories (Goetz et al., 2009; NRC, 2010; Law and Harmon, 2011). Techniques for classification of deforestation, forest degradation, and regrowth have advanced significantly since the 1990s and are now available to support forest monitoring efforts at regional scales, with unprecedented detail and accuracy (e.g. Kennedy et al., 2007; Asner, 2009). Conversely, inventory-based estimates of forest biomass have been regarded as the major source of uncertainty in calculations of land-use emissions, particularly in areas with low plot densities and the tropics, where systematic forest inventories are rare (Houghton et al., 2009).

Repeated measurements of biomass on permanent plots (e.g. Baker et al., 2004b) and direct eddy covariance measurements of CO₂ fluxes (e.g. Law et al., 2001; Saleska et al., 2003) can provide detailed estimates of carbon balance at local scales, but are difficult to scale up using simple extrapolation (Fisher et al., 2008). This is especially true for tropical forests because they show substantial floristic and structural variation across environmental and disturbance gradients, and a corresponding variation in carbon stocks and productivity (e.g. Clark et al., 2001; Vieira et al., 2004; Asner et al., 2009). Process-based ecosystem models have proven a valuable tool for regional assessment of carbon fluxes (e.g. Law et al., 2004; Turner et al., 2007), but are dependent on structure or biomass knowledge for initialization and/or validation purposes (Turner et al., 2004; Hurtt et al., 2010; Waring et al., 2010). In addition, model predictions are often dependent on a number of climate, soil, and vegetation parameters for which appropriate values may be difficult to
obtain (Landsberg, 2003; Nightingale et al., 2004). In the Brazilian Amazon, for example, biomass estimates based on interpolation of ground measurements, modeling, and/or percent tree cover maps have varied by more than a factor of two, showing no clear pattern as to the spatial distribution (Houghton et al., 2001; Malhi et al., 2006). Methods for obtaining accurate, spatially explicit estimates of biomass density are therefore required to reduce uncertainties in the net land flux, and to support international climate agreements and emerging carbon markets. We note that the requirement for self-reporting of emissions for non-Annex I (developing) countries is to produce periodic inventory of CO₂ (and other greenhouse gases) at the forest sector level. However, recommendations are for all countries to report to improve national inventories and report forestry emissions at the Tier 3 level, which incorporates more complex approaches than change in biomass, and models of emission sources (e.g., net ecosystem production; NRC, 2010).

Over the last two decades, considerable effort has been directed toward estimating aboveground biomass from remote sensing data (Rosenqvist et al., 2003; Lu, 2006; Goetz and Dubayah, 2011). Multispectral sensors, although used with success in a number of ecological applications (Cohen and Goward, 2004), have been of limited value for biomass estimation in tropical regions as spectral differences are not easily discernible for forests older than ~15 years (Sader et al., 1989; Steininger, 2000; Lu, 2005). Approaches based on L- or P-band radar power have demonstrated sensitivities from 60 to 270 Mg ha⁻¹ (Imhoff, 1995; Luckman et al., 1997; Hoekman and Quiriones, 2000; Santos et al., 2002, 2003; Mitchard et al., 2009; Lucas et al., 2010; Saatchi et al., 2011a), but these numbers are still well below those needed for the tropics, where the average biomass could be as high as 380 Mg ha⁻¹ (Grace et al., 2001) and individual sites can reach more than 800 Mg ha⁻¹ (Keith et al., 2009). While some recent studies suggest that both optical and radar sensitivities can be increased, either by exploring a combination of new metrics (Gonçalves et al., 2011; Eckert, 2012) or using novel scaling approaches along with sophisticated statistical algorithms (Saatchi et al., 2011b; Baccini et al., 2012), the prediction accuracy may continue to be insufficient beyond current saturation limits (cf. Saatchi et al., 2011b, Fig. S4). Furthermore, there has been no consistent support for an improved
optical/radar performance in the literature (Goetz and Dubayah, 2011), with recent results having yet to be reproduced and validated with appropriate ground measurements.

More recent three-dimensional (3-D) remote sensing techniques sensitive to the vertical structure of vegetation—namely light detection and ranging (lidar; Lefsky et al., 2002) and interferometric synthetic aperture radar (InSAR; Treuhaft et al., 2004)—are much less affected by saturation in tropical forests and are expected to be a major component of any carbon-monitoring program. Methods for retrieving structural information from lidar and InSAR have been tested and validated in a number of tropical sites, resulting in errors in canopy height estimates that range from about 2 to 4 m, at plot scales of 0.5 ha and below (Drake et al., 2002; Neeff et al., 2005; Lefsky et al., 2007; Hajnsek et al., 2009; Treuhaft et al., 2009; Simard et al., 2011). On the other hand, methods for transforming remotely sensed measurements of structure into predictions of forest biomass are still under development. Although lidar and InSAR have shown improved biomass estimation performance relative to nonstructural approaches, it is still unclear what accuracies are achievable from these techniques in forests with high biomass density. Studies that employ empirical, height-based methods have reported errors that range from approximately 40 to 100 Mg ha\(^{-1}\) (Drake et al., 2003; Lefsky et al., 2005; Neeff et al., 2005; Treuhaft et al., 2010; Clark et al., 2011; Asner et al., 2012; Baccini et al., 2012), generally well above the required accuracy of 20 Mg ha\(^{-1}\) or 20% (not to exceed 50 Mg ha\(^{-1}\) in forests \(\geq 250\) Mg ha\(^{-1}\)), necessary to estimate carbon fluxes from deforestation and forest degradation to within 20%, at spatial resolutions of 100–250 m (Houghton et al., 2009; Hall et al., 2011).

In fact, by propagating the errors of recently published continental and pantropical-level models that relate plot-level aboveground biomass to a type of field-measured canopy height (Saatchi et al., 2011b), one can easily verify that current 2 to 4 m errors in lidar- or InSAR-derived heights will make it difficult to meet a biomass accuracy requirement of 20%. If in practice 2 m is about the best achievable accuracy in dense tropical forests, which seems a reasonable assumption given the various sources of error associated with both lidar and InSAR (Treuhaft et al., 2009; Dubayah
et al., 2010), then remote sensing methods for estimating biomass will need to include more than just height. This argument is fairly intuitive as allometrically-derived biomass is a function of not only tree height, but also diameter and wood density (Chave et al., 2005). Although tree diameter is generally highly correlated with height, the relationship can vary significantly as a function of climate and basal area (Feldpausch et al., 2011). Similarly, wood density is known to vary substantially among species and across regions (Chave et al., 2006), determining large-scale patterns in the distribution of forest biomass (Baker et al., 2004a; Malhi et al., 2006).

In a recent attempt to address this issue, Asner et al. (2012) extended the tree allometry theory to the plot scale and proposed a generalized biomass model (in units of Mg C ha\(^{-1}\)) that explicitly incorporates information on tree diameter (in the form of basal area) and wood density, in addition to lidar height. When calibrated with plot-level data, the model captured nearly all of the variation in biomass across four tropical regions, with an error of less than 10%. However, when regional estimates of basal area-to-height ratio and wood density were used instead—simulating a more 'real-world' scenario where remote sensing is the only source of information at the plot scale—the model error increased over twofold, becoming comparable to that of models calibrated only with lidar height. While these findings provide evidence for the potential of approaches that incorporate more than just height, they reinforce the need for methods that are capable of delivering the required biomass accuracy.

Besides providing canopy height, lidar and multibaseline InSAR techniques can be used to effectively characterize the vertical distribution of vegetation within the canopy of tropical forests (Drake et al., 2002; Treuhaft et al., 2009). Following stand-replacing disturbances, successional changes driven by dynamics of colonization and competition for resources in these forests induce a gradual, and to some extent, predictable pattern of species replacement and canopy organization, and a corresponding pattern of biomass accumulation (Saldarriaga et al., 1988; Brown and Lugo, 1990; Guariguata and Ostertag, 2001; Neef and Santos, 2005). Although stand development is largely dependent on environmental conditions and disturbance regime, general trends in vegetation growth and dynamics allow the establishment of strong predictive relationships between different measures of forest structure, and
projections of future states of forests from knowledge of their current state (Spies, 1998). The close connection between vertical structure and biomass, for example, has been represented with success in mechanistic models of forest dynamics (Hurtt et al., 2010). Ecological theories that describe how trees use resources, fill space, and grow, establish solid quantitative links between size-frequency distribution, canopy structure (i.e. dimensions and spacing of tree crowns), and biomass (Enquist et al., 2009; West et al., 2009), suggesting that vertical profiles provided by lidar or InSAR contain a lot more information than what has been exploited in standard height-based approaches.

In this study, we use airborne and space-based lidar observations along with detailed forest inventory data collected at the Tapajós National Forest, Brazil, to advance methods of biomass estimation in tropical regions from remotely sensed structure. More specifically, our objectives were to (1) test and refine methods for the extraction of structural information from lidar data; (2) determine the accuracy of lidar estimates of structure in relation to detailed field measurements of vertical structure; (3) develop and validate structure-based models for optimal prediction of aboveground biomass; and (4) integrate the resulting models in a multi-stage scaling approach to biomass estimation (Baccini et al., 2012) and produce a wall-to-wall map of biomass for a test site in the Amazon. In addition to evaluating the performance of simple summary metrics such as lidar mean or maximum canopy height, we explore Fourier transforms of profiles as a means of incorporating information on directional, mechanism-driven change in structure into empirical biomass models. Rather than relying on regional averages of basal area and wood density, or attempting to estimate it from remote sensing, we look at lidar profiles in the frequency domain and consider the vertical scales at which the canopy is organized explicitly, as it relates to basal area, species composition (i.e. wood density), and a number of other forest attributes. We build on our previous work (Treuhaft et al., 2010) with the airborne LVIS instrument (Blair et al., 1999) at La Selva, Costa Rica, extending it to a tropical moist forest in the central Amazon, and to the analysis of 3-D data from commercial off-the-shelf airborne lidar and the spaceborne Geoscience Laser Altimeter System (GLAS, Schutz et al. 2005). We also investigate an alternative modeling approach
that allows more flexibility in the use of Fourier-based metrics while accounting for the uncertainty in the variable selection process.

3.3 MATERIALS & METHODS

3.3.1 Study site

The Tapajós National Forest is a 545,000 ha federal conservation unit in the Brazilian Amazon, situated about 50 km south of Santarém, PA, between the Tapajós River and the Cuiabá-Santarém highway (Figure 3.1). The climate is tropical humid (Köppen Am), with a mean annual temperature of 25.1°C and mean annual precipitation of 1909 mm, with a 5-month dry season (< 100 mm month⁻¹) between July and November (Vieira et al., 2004). The vegetation is classified as dense, upland, tropical moist forest, and included over 200 tree species in the plots sampled in this study. Common families included Fabaceae, Sapotaceae, Burseraceae, Lecythidaceae, Salicaceae, and Urticaceae. The soils are nutrient-poor oxisols and ultisols, with low pH, organic matter, and cation exchange capacity, and a high concentration of aluminum oxides (Silver et al., 2000). Our sample sites were situated on a relatively flat plateau, with the elevation ranging from approximately 80 to 180 m.

3.3.2 Field data

We collected field data in September 2010 (30 plots) and August 2013 (20 plots) in a total of 50 0.25-ha plots (50 m x 50 m) co-located with lidar data acquired at Tapajós (Figure 3.1). Of the 50 plots, 15 were primary forest (PF), 9 were primary forest subject to reduced-impact selective logging (PF₁), and 26 were secondary forest (SF) with different ages and disturbance histories. The 30 plots measured in 2010 were centered on GLAS footprints selected along two ground tracks spanning a wide range in vertical structure and aboveground biomass. Of those, 19 were also covered by airborne laser scanning (ALS). The 20 plots measured in 2013 were all located within the ALS area. Plot centers were geolocated using differential Global Positioning System (GPS). In 2010, we also used a total station to locate the center of the GLAS footprints.
Field measurements included diameter at breast height (1.3 m) or above trunk irregularities ($D$), height to the base of the live crown ($H_C$), total height ($H_T$), and location of each living tree $\geq 5$ cm in diameter in early successional stands, and $\geq 10$ cm in all other stands. For the 2010 plots, we also measured crown radius ($C_R$) in four orthogonal directions in a 12.5 m x 50 m subplot extended along the major axis of the GLAS footprint. In the 2013 campaign, $C_R$ measurements were taken for all trees in the 50 m x 50 m plot. All trees were identified to species or genus level (when species was uncertain) and assigned a wood density value ($\rho$, oven-dry weight over green volume) derived from the literature (Reyes et al., 1992; Chave et al., 2006).

In the 2010 campaign, heights were estimated by eye by experienced members of the field crew. For a small subset of the trees, heights were also obtained with a laser rangefinder. We used these observations to develop a regression model relating precise laser-measured heights to the less precise, however more readily obtained, visually estimated heights. This model, described in detail in Chapter 2, was then applied to calibrate all visually estimated heights. In the 2013 campaign, all heights were obtained with a clinometer and a measuring tape, keeping angles below 50° to minimize measurement error.

3.3.2.1 Biomass estimation and its error

The oven-dry aboveground mass of each live tree ($M$) was estimated from its diameter, total height, and wood density using an established allometric equation for tropical moist forests (Chave et al., 2005). Exceptions were made for *Cecropia* spp. and palms, which differ significantly from other species in wood density and allometry (Chave et al., 2005; Nelson et al., 1999) and had their biomass estimated with specific equations. The aboveground biomass density (AGB, Mg ha$^{-1}$) at the plot level was calculated by adding the masses of all inventoried trees in the plot and dividing by the plot area (i.e., 0.25 ha). In plots where the minimum diameter was 10 cm, we corrected for the AGB in the 5–10 cm class by estimating the number of stems in that class using the diameter distribution of the plot. Because the field and lidar data were acquired with a temporal difference of up to 3 years, we also applied a site-specific, stand-level growth model (Neeff & Santos 2005) to all SF plots to
correct for the AGB change between observation epochs. For primary forests, we assumed no significant change in the 3-year period, based on results of previous studies conducted at Tapajós (Pyle et al., 2008; Rice et al., 2004; Vieira et al., 2004).

We quantified the uncertainty in the field biomass accounting for a number of sources of error. This included sources that would be normally expected in conventional biomass estimates (e.g., measurement and allometric errors) and those that are specific to remote sensing studies, such as the error resulting from spatial disagreement between field and lidar samples (co-location error) and the error introduced when accounting for temporal differences in data acquisition. The methods discussed in this section are described in detail in Chapter 2.

3.3.2.2 Derivation of vegetation density profiles

We characterized the canopy vertical structure by modeling individual tree crowns as ellipsoids uniformly filled with leaf area (Nelson 1997, Treuhaft et al. 2009)

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

where $x$, $y$, and $z$ are the axes of a 3-D Cartesian coordinate system, and $a$, $b$, and $c$ are the semi-principal axes of the ellipsoid, with $a$ and $b$ representing the mean crown radius in the directions parallel and perpendicular to the plot orientation, and $c$ calculated as half the crown depth (i.e. $(H_T - H_C)/2$, Figure 3.2a). As described in section 3.3.2, for the 2010 plots we only measured crown radius for trees located in a 12.5 m x 50 m subplot covering 25% of the plot area. To characterize the vertical structure of the full 50 m x 50 m plot, making it comparable to the GLAS-derived structure, we used the subplot data to develop an allometric equation relating crown radius to diameter and height measurements. After applying this equation to estimate mean crown radius for all trees in the remaining 75% of the plot, we generated a vegetation density profile by summing the volume of all ellipsoids in the plot.
overlapping each 1-m height interval (Figure 3.2b). For a height bin centered at $z$, the relative density $w_F$ was modeled as (Treuhaft et al. 2009)

$$w_F(z) \propto \sum_{i=1}^{n_z} \int_{z_{lo_i}}^{z_{hi_i}} V_i(z) \, dx \, dy \, dz = \sum_{i=1}^{n_z} \int_{z_{lo_i}}^{z_{hi_i}} \pi a_i b_i \left[ 1 - \left( \frac{z - z_i}{c_i} \right)^2 \right] \, dz$$  \hspace{1cm} (2)

where $n_z$ is the number of trees with any part of the crown inside the height bin at $z \pm 0.5$ m; $V_i(z)$ is the volume of the ellipsoid associated with the $i^{th}$ tree in the bin, calculated from the lowest ($z_{lo_i}$) to the highest ($z_{hi_i}$) point for that tree within the bin; $z_i$ is the height of the center of the ellipsoid representing the $i^{th}$ tree; and $a$, $b$, $c$, $x$, $y$, and $z$ are defined as in (1).

Using an approach similar to that described for biomass, we corrected for the change in mean height ($H_{MEAN}$) between observation epochs by applying the sitespecific growth model of Neeff & Santos (2005)

$$H_{MEAN_t} = 0.982 \, H_{TOP_t} - 0.025 \, H_{TOP_t}^2 + 0.001 \, H_{TOP_t}^3$$  \hspace{1cm} (3)

$$H_{TOP_t} = 23.067 \, (1 - e^{-0.074 \, t})^{0.946}$$

where $H_{TOP_t}$ is the top height, defined as the mean total height of the tallest 20% of the trees; and $t$ is the stand age (years since stand initiation). This was done by (i) inverting (3) to estimate the plot age at the time of the inventory ($t_1$) from its measured height ($H_{MEAN_{t1}}$); (ii) estimating the mean height of the plot at the time of the lidar acquisition ($H_{MEAN_{t2}}$), making $t = t_1 \pm \Delta_t$, with $\Delta_t =$ difference between acquisitions in years; and (iii) calculating the difference between $H_{MEAN_{t1}}$ and $H_{MEAN_{t2}}$. For secondary forests, the change in mean height derived from this plot-level model was applied as a shift to the profile, assuming a uniform height increment for all trees, independent of the tree height—a reasonable assumption, particularly during early stand development.
3.3.3 Lidar data

3.3.3.1 ICESat/GLAS

We used unsaturated, cloud-free lidar profiles acquired by the Geoscience Laser Altimeter System (GLAS, Schutz et al. 2005) at Tapajós during October 2007 (Laser 3I campaign). GLAS is a large-footprint, waveform-recording lidar that measures the timing and intensity of the 1064 nm laser energy returned from illuminated surfaces, quantifying the vertical distribution of vegetation structural components (e.g., leaves, branches, and stems) and the underlying ground within each footprint (Harding & Carabajal 2005). Each GLAS footprint is elliptical in shape, spaced at ~172-m intervals along-track, and averaged 0.22 ha in the ground area covered for the 30 GLAS shots used in this study (average major and minor semi-axes of 58.4 and 48.4, respectively). For the Laser 3I campaign, the horizontal geolocation accuracy was estimated at 0.65 ± 2.22 m, and the vertical accuracy at 0.5 ± 1.55 cm per 0.4° incidence angle (NSIDC 2011). The mean off-nadir pointing angle for the 30 shots was below 5°, and terrain slope ranged from 0.2 to 6.0°, as measured by SRTM (Rabus et al. 2003).

**Waveform processing**

Using parameters reported in the GLA01, GLA05, and GLA14 data products, we performed a number of processing steps in advance of retrieving structural metrics. Following methods of Sun et al. (2008), the waveforms were first filtered by convolution with a discrete Gaussian kernel with the same standard deviation as the transmitted laser pulse (kernel width = 6σ, with σ ≈ 2.5 ns). This procedure reduced background noise, enhancing detection of the ground return, while preserving an adequate level of detail for characterization of the canopy (Figure 3.3).

The GLAS waveforms used in this study were calibrated and digitized into 1000 discrete bins at a time resolution of 1 ns (~15 cm), allowing the detection of a maximum elevation change within a single footprint of 150 vertical meters. The locations of the highest (signal start) and lowest (signal end) detected surfaces within the 150-m waveform were determined, respectively, as the first and last elevations at
which the amplitude exceeded a threshold level, for a minimum of $n$ consecutive bins (Figure 3.3). We set $n$ to half and to the full width of the transmitted laser pulse for detection of signal start and signal end, respectively (roughly $3\sigma$ and $6\sigma$ bins). The threshold was calculated as $(\text{mean noise, } \mu_{\text{noise}}) + k (\text{standard deviation of noise, } \sigma_{\text{noise}})$, where $k = 4$ was found to be the multiplier that minimized the root mean square error (RMSE) between GLAS and field estimates of maximum canopy height. Similar values were used by Lefsky et al. (2005, 2007), supported by comparisons of waveform extent derived from GLAS and coincident airborne lidar. As noted by Sun et al. (2008), the noise level in vegetated terrain before the signal start is usually lower than that after the signal end. As a result, we established two separate estimates of noise mean and standard deviation using the initial and final portions of each waveform, before and beyond any potential canopy or ground return (Figure 3.3).

The peak of the ground return was identified for each waveform using a hybrid approach based on automated detection of plausible peaks followed by manual inspection and correction as necessary. We considered as plausible peaks the lowest peaks in the smoothed waveforms with at least the same width as the transmitted laser pulse, after taking into account the mean noise level (Figure 3.3). The manual inspection involved checking neighboring footprints, as well as field-based profiles, to verify that the elevation of the selected peak was in agreement with the local elevation. It also involved an inspection of the unfiltered waveforms to make sure no important ground information was removed by the filtering step.

3.3.3.2 Airborne Laser Scanning (ALS)

The ALS data used in this study were acquired by GEOID Ltda. in August 2012 and 2013 using an Optech 3100 EA instrument flown at approximately 600 m with a $10^\circ$ field of view, a small footprint (~15 cm), and a 30% overlap between adjacent swaths. For each lidar pulse, the first and the last return were recorded, resulting in an average of 36 returns per m$^2$ for our plots. The primary-forest ALS data (northern red rectangle in Figure 3.1) were acquired in 2012. ALS data over secondary stands (two southern red rectangles) were acquired in 2012 and again in 2013. For comparison with GLAS, the discrete-return data were aggregated to produce pseudo-waveforms
from which similar metrics could be extracted. Waveforms were synthesized by subsetting the lidar point returns co-located with each field plot and counting the number of returns observed in vertical bins of 50 cm. In these waveforms, the ground peak was taken as the maximum amplitude observed in the vertical bins classified as "sub-canopy" by the lidar vendor.

3.3.4 Structure metrics from vertical profiles

Methods are available to account for attenuation of the lidar energy as it travels through the forest canopy, adjusting the waveform to better represent the relative vertical distribution of plant area. A common approach is to use a modified MacArthur-Horn transformation (MacArthur & Horn 1969) to produce the so-called canopy height profile (Lefsky et al. 1999, Harding et al. 2001). An alternative is to estimate the canopy light extinction coefficient and use the Beer-Lambert law (Norman & Campbell 1989). In this study, we tested both techniques and compared the results against field-based profiles. In agreement with results of Drake et al. (2002) and Treuhaft et al. (2009) for another tropical site, we found that transformation of the lidar profiles did not necessarily improve the representation of the canopy structure. Overall, the untransformed lidar profiles showed better qualitative agreement with field profiles than their transformed counterparts. Therefore, we conducted the analyses using the untransformed profiles.

Using the thresholds described above, and assuming the peak of the ground return as the reference level (height = 0 m), we extracted a number of canopy structural metrics to characterize the vertical profiles for use as independent variables in the biomass estimation.

Traditional metrics

Metrics extracted that are commonly referenced in the literature (e.g., Drake et al. 2002, Harding & Carabajal 2005, Nelson et al. 2009, Treuhaft et al. 2009) included (i) the maximum canopy height, H100, defined as the vertical distance
between the ground peak and the signal start (Figure 3.3); (ii) the mean (MCH) and standard deviation (SD) of the height profile, calculated as

\[
MCH = \frac{\int_{0}^{H_{100}} z w(z) \, dz}{\int_{0}^{H_{100}} w(z) \, dz}
\]

\[
SD = \sqrt{\frac{\int_{0}^{H_{100}} z^2 w(z) \, dz}{\int_{0}^{H_{100}} w(z) \, dz} - MCH^2}
\]  

where \( w(z) \) is the laser power received from the 15-cm bin centered at height \( z \); (iii) the canopy height at the three power quartiles, obtained by accumulating the waveform energy and identifying the heights above the ground elevation which cut off the lowest 25% (H25), 50% (H50), and 75% (H75) of the data (Figure 3.3); (iv) the height entropy, \( S_z \), defined as

\[
S_z = -\sum_{i=1}^{n_b} p(w_i) \ln(p(w_i)), \text{ with}
\]

\[
p(w_i) = \frac{w_i(z)}{\int_{0}^{H_{100}} w_i(z) \, dz}
\]  

where \( n_b \) is the number of vertical bins from the ground peak to the signal start, and \( w(z) \) and \( z \) are defined as in (4), but with a vertical resolution of \( \sim 1 \text{ m} \); (v) the front slope, \( m_f \), defined as the rate at which the amplitude changes with height in the outer canopy—the vertical distance between the signal start and the height associated with the maximum canopy return; (vi) the ratio of ground-to-total power, \( R_{gt} \), calculated as

\[
R_{gt} = \frac{2 \int_{0}^{0} w(z) \, dz}{\int_{H_{SE}}^{H_{100}} w(z) \, dz}
\]
where $H_{SE}$ is the height associated with the signal end (i.e. the lowest detected surface) referenced to the centroid of the ground return at zero; and (vii) the number of peaks in the canopy return, $N_{pks}$.

**Fourier-based metrics**

As presented in recent research by Treuhaft et al. (2010), traditionally used height-based metrics tend to utilize only a fraction of the information contained in a lidar profile (i.e., the spatial frequencies near zero or, equivalently, the high vertical scales present in the signal). Given this limitation, we sought to explore use of the information contained in the vertical profiles not only in the space, but also in the frequency domain using

$$
\gamma(f) = \frac{\int_0^\infty w(z) e^{-i2\pi f z} dz}{\int_0^\infty w(z) dz}
$$

where $\gamma(f)$ is the normalized Fourier transform (Bracewell 1986) of the profile at spatial frequency $f$, and $w(z)$ and $z$ are defined as in (4). The waveforms were decomposed into frequencies ranging from 0.3 to 0.01 cyc/m (vertical wavelengths of 3.3–100 m), with a step of 0.01 cyc/m, resulting in a total of 30 additional complex valued metrics (each described by an amplitude and phase). We limited the frequencies to 0.3 cyc/m because the width of the transmitted laser pulse was estimated at 2–3 m for the 30 shots, and Fourier transforms involving frequencies higher than that would be difficult to interpret. The step was set to 0.01 cyc/m because lower steps were tested and the resulting extra metrics were highly correlated and did not improve biomass estimation significantly.

3.3.5 Structure-based biomass estimation

3.3.5.1 Regression analysis

7 The vertical distance over which the phase of $\gamma(f)$ winds through one cycle (i.e. $1/f$).
We used regression analysis to model the relationship between biomass density, AGB, and the set of candidate structural metrics, \( X = \{x_1, x_2, \ldots, x_P\} \)

\[
AGB = \beta_0 + \sum_{j=1}^{P} \beta_j x_j + \epsilon \tag{8}
\]

where \( \beta_0, \beta_1, \ldots, \beta_p \) are model parameters estimated by weighted least squares, with weights \( w_i \) inversely proportional to the overall uncertainty in AGB for the \( i^{th} \) plot, \( x_1, x_2, \ldots, x_p \) is a subset of \( X \), and \( \epsilon \) is a random error term assumed to follow a normal distribution with mean zero and variance \( \sigma^2/w_i \).

Several different subsets of \( X \) can be considered in (8). In this study, we conducted a data-driven search of subset models that work well by fitting all possible models with up to 6 predictors (i.e., structural metrics), and selecting the one that minimized the root mean square prediction error (RMSPE). For each model, the RMSPE was obtained by (i) randomly splitting the dataset into separate model training (80% of the observations) and validation (20%) subsets; (ii) fitting the model to the training data; (iii) applying the fit to predict AGB for the observations in the validation subset; and (iv) calculating

\[
\text{RMSPE} = \sqrt{\frac{1}{n_V} \sum_{i=1}^{n_V} \left( AGB_{\text{obs}_i} - AGB_{\text{pred}_i} \right)^2} \tag{9}
\]

where \( n_V \) is the number of observations in the validation subset, and \( AGB_{\text{obs}} \) and \( AGB_{\text{pred}} \) are the observed and predicted values, respectively. To reduce variability, the four steps above were repeated \( 10^3 \) times for each model and the results were averaged. When multiple models showed similar average predictive ability, we selected the one with the smallest Bayesian Information Criterion (BIC)

\[
BIC = n \log(\hat{\sigma}^2) + \log(n) (p + 1) \tag{10}
\]
where \( n \) is the total number of observations, \( \hat{\sigma} \) is the standard error of the regression, and \( p \) is the number of predictors included in the model. From (10), BIC takes a measure of goodness of fit and adds a penalty for the number of terms, improving for lower \( \hat{\sigma} \) and \( p \). We used BIC as the “tie breaker” because it captures the trade-off between model precision and complexity, and because it is an appropriate statistic when the correct model is expected to involve some, but not all, available predictors (Ramsey & Schafer 2002). We stress that different strategies for model selection exist and have shown similar performance in discriminating between informative and noninformative predictors, in both simulation (Murtaugh 1998) and empirical (Murtaugh 2009) studies.

3.3.5.2 Bayesian model averaging

While variable selection techniques such as BIC can be used to identify useful models for biomass prediction, predictions derived from any single selected model ignore the uncertainty involved in the selection process. In this study, we accounted for model uncertainty by averaging over the predictions of many candidate models adopting a Bayesian approach (Raftery et al. 1997, Hoeting et al. 1999) (see Appendix A for details).

3.3.5.3 Validation

In addition to calculating the RMSPE for the selected models, as described in section 3.3.5.1, we obtained the cross-validation coefficient of determination \((R_{CV}^2)\), calculated as the square of the correlation coefficient between \( AGB_{obs} \) and \( AGB_{pred} \). Following Treuhaft et al. (2010), we also tested the significance of each model by estimating the probability (the “confidence”) that the resulting RMS error between observed and predicted values originated from a real correlation between AGB and the set of structural metrics being considered. This was calculated as \( 1 - P(H_0) \), where \( P(H_0) \) is the probability that the RMSE occurred by chance alone (i.e., the null hypothesis of no relation). To estimate \( P(H_0) \), we randomized the biomasses among plots \( 10^3 \) times, while keeping the predictors fixed, and obtained RMS errors based on
data that were deliberately uncorrelated. \( P(H_0) \) was the number of permutations which produced a RMSE less than or equal to the RMSE that was actually observed, divided by the total number of permutations.

3.4 RESULTS & DISCUSSION
3.4.1 Crown allometry

Crown radius \((C_R)\) measurements in the 12.5 m x 50 m subplots (2010 campaign) were strongly correlated with stem diameter (on the square root scale; \( r = 0.77 \)), crown depth (\( r = 0.71 \)), and total height (\( r = 0.60 \)) (Figure 3.A1). Crown diameter (i.e. \(2C_R\)) was typically one-third of the tree height, and \(~20\%\) smaller than the crown depth, although crown shape varied considerably within plots. Crown depth-to-diameter ratios ranged between 0.5 and 3.0 on average, displaying a significant increase with tree height in half of the plots. A multiple linear regression model including \(\sqrt{D}, C_D,\) and \(H_T\) explained 70\% of the variation in \(C_R\), with a RMS error of 0.8 m (Figure 3.A1d). The regression coefficients were highly significant and differed for primary and secondary forests. The resulting \(C_R\) equations are provided in Table 3.A1, along with equations for predicting height and crown dimensions from stem diameter alone.

The effect of the use of these equations for the estimation of vertical profiles is illustrated in Figure 3.A2. It shows the qualitative agreement between actual profiles of nine subplots measured in 2010 (shaded polygons) and profiles estimated using (i) modeled \(C_R\) and field-measured \(C_D\) and \(H_T\) (dotted lines), and (ii) modeled \(C_R, C_D,\) and \(H_T\) (solid lines). The results indicate that field profiles can be reliably estimated when allometrically-derived \(C_R\) is used in (2) instead of actual \(C_R\) measurements. On the other hand, Figure 3.A2 suggests that the use of height-diameter relationships in addition to crown allometries has the potential to introduce large errors in the estimation of vertical profiles.
3.4.2 General pattern of structural development

Figure 3.4 shows vertical vegetation density profiles derived from field (shaded polygon), GLAS (black line), and ALS (gray line) measurements for nine different representative sample plots spanning nearly the entire range of biomass observed at Tapajós. The plots are arranged alphabetically in order of increasing biomass (a proxy for age) and illustrate patterns typical of structural development at Tapajós following stand-replacing disturbance (although not strictly a chronosequence). Early stages of ecological succession (Figure 3.4a) are characterized by the establishment of a few fast-growing pioneer species with low wood density (e.g., *Cecropia* spp., *Banara nitida*, *Inga cylindrica*), which quickly form a dense, homogenous canopy with low stature and biomass, and minimal structural complexity (although stands sometimes include taller remnant trees; see Figure 3.2b). The canopy grows rapidly in height and depth as trees compete for light during early stand development, but shows relatively little change in structural complexity (Figure 3.4a–d). This period is accompanied by a rapid increase in aboveground biomass, with stands typically accumulating about 120 Mg ha\(^{-1}\) in less than 20 years of regrowth (Neeff & Santos 2005).

As succession proceeds, competition for space and resources imposed by the high stem density in these closed-canopy stands leads to tree mortality and canopy gap formation. Early colonizers are gradually replaced by slower growing, shade-tolerant species with higher wood density, in a process that increases the species diversity and the canopy complexity (through development of distinct under- and mid-story layers), while governing changes in stand-level wood density, basal area, mean height, and therefore biomass (Figure 3.5). As shown in Figure 3.4e–i, late-successional forests are characterized by small differences in maximum canopy height, but substantial variation in vertical organization. As biomass increases from ~190 to 540 Mg ha\(^{-1}\), the initially “bottom-heavy” vegetation density distribution transitions to a “top-heavy” structure, leading to an increase in the profile-averaged mean height which scales with biomass. Structural differences in these old-growth stands are largely driven by disturbances such as selective logging (Figueira et al. 2008) and drought-induced tree mortality (Rice et al. 2004, Nepstad et al. 2007).
interacting with edaphic (Silver et al. 2000) and topographic (Bispo et al. 2014) gradients.

Although the successional patterns described above are commonly observed in tropical forests (e.g., Saldarriaga et al. 1988), the identity of species that dominate at each stage and the rate at which this dominance changes can differ substantially among sites, depending on factors such as past land use and the severity, frequency, and type of disturbance (Cook 1996, Guariguata & Ostertag 2001). Nevertheless, the sequence of events that lead to structural development is expected to occur regardless of the species composition (Guariguata & Ostertag 2001), resulting in patterns of biomass accumulation that can be predicted by gap models (e.g., Moorcroft et al. 2001). Some of the stands depicted in Figures 3.4 and 3.5, for example, had different land use and disturbance histories and showed clear differences in species composition, after accounting for differences in development stage (see cluster analysis in Figure 3.A3). Yet they appear to support the notion of a general and fairly predictable pattern of structural development and biomass accumulation, and thus the basic premise explored in this study that biomass can be reliably estimated from vertical profiles alone. By expressing biomass as a function of both height and vertical distribution, this premise should hold even for less common developmental pathways in which structural complexity arises early in stand development, as a result of disturbances such as wildfire (Donato et al. 2012).

3.4.3 Qualitative agreement between field and lidar profiles

Figure 3.4 shows that in early and mid-successional stands, the vertical position and extent of the short, unimodal canopy agree well between field and lidar profiles. In the taller, multilayered old-growth forests, there is also a good qualitative agreement for large-scale features. However, the lidar profiles include small-scale details that are missed by the field profiles due to the simplifications made in the canopy model in (2) (e.g., that crowns are filled ellipsoids as opposed to highly irregular in shape with clumped foliage). We note that most of these small-scale features were similarly identified in the GLAS and ALS profiles, although with some variation in complexity and relative amplitude. We also note that the amplitude of the
ground return was consistently higher in the ALS profiles, probably due to the use of lower incidence angles by ALS (GLAS requires off-nadir pointing of up to 5° to reduce saturation effects). Another point to bear in mind is that unlike the GLAS waveforms, the pseudo-waveforms generated from ALS counts characterize only the presence or absence of reflecting surfaces and thus are not affected by spectral differences between canopy and ground surfaces (Popescu et al. 2011).

The observed differences between profiles may be explained by a number of factors. The discrepancy between field and lidar near the ground (heights up to 5–10 m) occurs, in part, because the lidar “sees” the understory vegetation and the ground, in addition to the trees that compose the field profiles. Temporal differences in data acquisition of up to 3 years between field and lidar, and 6 years between GLAS and ALS, represent another important source of variation (Figure 3.A4). While the site-specific growth model of Neeff & Santos (2005) was used to correct for changes in mean height in secondary forests (i.e., to translate the profiles on the vertical axis), no correction was applied to account for changes in the shape of the profiles as a result of stand development and/or small-scale disturbances (e.g. treefalls). Other factors underlying the observed differences include errors in the location of the ground peak and noise thresholds in the lidar profiles (e.g., Rosette et al. 2008, Treuhaft et al. 2009); errors due to spatial disagreement between field and lidar measurements (i.e., co-location error; see Figure 3.A4); uncertainty of field measurements of tree height (Hunter et al. 2013, Chapter 2); limitations of the crown shape model used to derive the field profiles (Drake et al. 2002, Treuhaft et al. 2009); and limitations of the waveform in representing the correct relative vertical distribution of vegetation due to attenuation of the lidar energy as the pulse travels through the canopy (Lefsky et al. 1999, Harding et al. 2001). Some of these factors are discussed in more detail in the following section.

3.4.4 Comparison of field and lidar-based measurements of canopy structure

Height-based metrics calculated from both GLAS and ALS profiles were highly correlated with the corresponding field metrics corrected for growth (Figure 3.6). The root-mean-square deviation (RMSD) between field and lidar measurements of mean
canopy height (MCH) was 2.8 m for GLAS and 3.3 m for ALS, representing 17% and 21.6% of the average mean height, respectively (Table 3.1). The RMSD was similarly low for the quartile heights H25 and H50, ranging between 2.4–3.5 m, but increased to 4–4.6 m for H75 and roughly doubled (~6 m) for the maximum canopy height (H100), indicating less agreement between field and lidar in the upper half of the profiles. However, little difference was observed between H100 and MCH when the RMSD was expressed in relative terms (Table 3.1). The agreement for the standard deviation of the profiles was at the 2.2–2.6 m level, or approximately 40%. We note that RMSD for the MCH matched closely that observed with LVIS (RMSD = 3.2 m) in our previous experiment at La Selva (Treuhaft et al. 2009). The RMSD for H100 was also fairly close to that reported by Hunter et al. (2013) for estimates of maximum tree height obtained at Tapajós and Reserva Ducke (north of Manaus, Brazil) using a different ALS system (RMSD = 7.3 m). Taken together, these studies suggest that the accuracy of lidar estimates of canopy height in tropical forests may have little dependence on the type of instrument used (i.e., small or large footprint, discrete-return or waveform-recording).

Although lidar-based metrics were on average close to the corresponding field metrics, showing a relatively low scatter, they were not centered around the 1:1 line. As evidenced by the box plots in Figure 3.6g, and quantified in Table 3.1, there was a tendency for both GLAS and ALS heights to be shorter than field heights in secondary forests, and taller in primary forests. For the profile-averaged mean height and the relative heights associated with the three energy quartiles (i.e., H25, H50, and H75), this could be explained by the assumption in the calculation of field profiles that leaf area density (LAD) is uniformly distributed within crowns. In fact, theory and observations suggest that trees in closed-canopy forests optimize leaf and nitrogen distribution to maximize carbon gain (Hirose 2005). Kitajima et al. (2005), for example, showed that the distribution of foliage in tree crowns of a tropical forest in Panama is a function of crown depth. They found that leaf area accumulation rates tend to decrease exponentially down through the canopy profile, in parallel with the exponential decline in light availability (trees accumulated LAI rapidly at the top 1-2 m from the crown surface, and then more slowly at lower crown depths). A crown
model simulating an exponential decrease in LAD from the top of the crown would result in higher field-based heights at Tapajós (except for H100, discussed below), reducing the observed differences in primary forests. However, species can differ significantly in rates of leaf area accumulation with depth (Kitajima et al. 2005), and no information was available for the species found in this study (with the exception for Cecropia, as described below). In future experiments, rather than relying on species-specific traits, it may be possible to develop models that explore general quantitative links between LAD and simple measures of structure, such as tree height and canopy openness (cf. Sterck and Bongers 2001). Nonetheless, while assumptions concerning the distribution of LAD have the potential to introduce bias, they were found in our previous experiment to have a relatively small effect (< 10%) on the RMSD between field and lidar estimates of canopy height (Treuhaft et al. 2009).

Kitajima et al. (2005) further showed that fast-growing, shade-intolerant Cecropia trees, the dominant canopy species in our SF plots, have sparse, vertically oriented branches with non-overlapping leaves at the tip, which result in very low LAI and light extinction. Since species-specific LAD is not taken into account in the calculation of field profiles, this effect would be neglected, resulting in increased LAD at higher altitudes relative to lidar—due to the low light extinction, the majority of the lidar energy returned would come from shrubs and slower growing tree species positioned lower in the canopy. The overestimation of field heights in secondary forests could also be explained by the observed high density of short trees with DBH smaller than the cutoff limit used in the inventory (i.e., 5 or 10 cm). In our three youngest sites, for example, the stem density in the 3–5 cm DBH class was 4 to 7 times as large as the total density of trees ≥ 5 cm DBH. And while these small trees contributed energy to lower altitudes in the lidar profiles, they were excluded from the field profiles. The contribution of energy from understory trees and shrubs, as well as the ground underneath the canopy, can also explain the systematic overestimation of the standard deviation by lidar (Figure 3.6), particularly at later stages of stand development, when gap-forming processes begin to operate.

In primary forests, field estimates of maximum canopy height (H100), which are independent of assumptions of foliage distribution, were lower than GLAS and
ALS estimates by an average of 4.5 and 5 m, respectively (Table 3.1). An evaluation of the accuracy of 200 tree height measurements from the 2010 field campaign (i.e., ocular estimates calibrated with laser) indicated no evidence of systematic errors for trees up to ~35 m tall, but revealed that the total heights of the only three individuals in the sample taller than that were systematically underestimated by an average of 5.7 m (see Figure 2.5). Although the number of observations is small to allow generalizations, the observed bias is consistent with that of PF plots in Figure 3.6f, all of which have a maximum canopy height greater than 35 m according to both GLAS and ALS. It is also in line with the results of Hunter et al. (2013), who found that heights of emergent trees taller than 34 m are notoriously difficult to measure at Tapajós. In the comparison with ALS, we note that only 4 out of the 13 PF plots were measured in 2010—the remainder is from the 2013 campaign, with heights obtained using the clinometer approach. If these four plots are removed from the analysis, the bias drops by half (from 5 to 2.5 m) and becomes not statistically significant, providing further evidence that the observed differences in $H_{100}$ in primary forests are due to underestimation of field heights rather than overestimation by lidar. In secondary forests, the lower differences in $H_{100}$ (Table 3.1) reinforce the accuracy of our field measurements for trees shorter than 35 m and suggest that the application of the growth model in (3) did not introduce systematic errors. Sources of uncertainty such as co-location error, which can lead to either underestimation or overestimation of heights, are also unlikely to have contributed to the observed biases in height-based metrics.

Simple regressions between corresponding field and lidar metrics showed that GLAS and ALS were similarly sensitive to structural differences at Tapajós, explaining 74–87% of the variation in field measurements (Table 3.1). With the exception of the standard deviation, we found no evidence of a difference between the slopes of the regressions for GLAS and ALS (Figure 3.6). When the comparison was restricted to the same plots, even the slopes for the standard deviation were equal. The RMS errors ranged between 11 and 20% (Table 3.1), and were typically smaller than the uncertainty in individual tree height measurements (Chapter 2, Hunter et al. 2013). In addition, the variability around the regression lines was roughly constant for
all metrics. We note that the regression results for the quartile heights in Table 3.1 (i.e., 2–2.7 m) were comparable to those obtained by Drake et al (2002) for LVIS data collected at La Selva (R² = 0.78–0.81; RMSE = 1.1–2.8 m). In contrast, the regression errors reported by Lefsky et al. (2005, 2007) for GLAS estimates of mean (RMSE = 4.3 m) and maximum (RMSE = 9.9 m) canopy height in the Amazon were more than twice as large as those reported here. This discrepancy reflects differences in waveform processing strategies for extracting height metrics, but also important differences in field protocols and definitions of height.

3.4.5 Structure-based biomass estimation

Traditional height-based metrics, measures of canopy complexity (e.g., height entropy and number of peaks in the canopy return), and Fourier transforms of profiles were generally highly correlated with biomass, despite the observed differences between field and lidar profiles (Figure 3.4). Metrics from field profiles followed similar trends with biomass. The relationship between biomass and most of the structure metrics was not linear. This is because biomass is calculated as a function of the product of wood density, stem diameter squared, and total height (i.e., the compound variable $\rho D^2 H_T$), and both wood density and diameter scale with height. Transforming biomass to the square root scale, as suggested by graphical analysis and the Box-Cox procedure (Box and Cox 1964), yielded a straight line relationship for most of the metrics. The regression analysis was therefore conducted with biomass on the square root scale and the predictions were back-transformed using an unbiased estimator (Gregoire et al. 2008).

Figure 3.7 shows field-estimated biomass versus that predicted by models based on traditional (left) and Fourier (right) metrics derived from field (top), ALS (middle), and GLAS (bottom) profiles. The selected models are given in Table 3.2 along with measures of predictive ability. For the field profiles, the best traditional model included only MCH. This model explained 92% of the variation in biomass (as measured by the $R^2_{CV}$), with a RMSPE of 45 Mg ha⁻¹ (25% of the mean biomass). For the lidar profiles, the best traditional models included metrics indicating canopy
openness and/or heterogeneity ($m_p$, $R_g$, SD), in addition to height (Table 3.2). These models explained a similar proportion of the variation in biomass (88–91%), but exhibited higher RMSPE (GLAS = 50 Mg ha$^{-1}$ or 29%; ALS = 48 Mg ha$^{-1}$ or 33%). The better performance of the field model may be in part attributed to the fact that the biomass and the field-based MCH are not entirely independent as the field profile is calculated using the same tree height measurements that are used to calculate biomass.

The best Fourier models for field, GLAS, and ALS included sets of five Fourier frequencies (vertical wavelengths), which have some overlap. All models included information at 0.01 cyc/m (100 m) and a combination of Fourier transforms including vertical wavelengths between ~4 and 6 m. Although the selected frequencies differed between models, there was surprisingly little difference in model performance. The models explained between 93–94% of the variation in biomass, with a RMSPE of 22–24% (Table 3.2). This represents an improvement over the prediction performance of the traditional model of ~14% for field, 25% for GLAS, and 27% for ALS, with the improvement occurring primarily in high-biomass forests (Figure 3.7). As observed in Table 3.2, the RMSPEs for traditional and corresponding Fourier models differ so markedly that neither value lie within the confidence interval based on the other.

Residual plots and formal tests for the models in Table 3.2 indicated no serious violations of the assumptions for linear regression. Scatterplots of residuals versus fitted values (not shown) indicated that the variability tends to increase with biomass, but the least squares estimates are still unbiased even if the variance is nonconstant (Ramsey & Schafer 2002) and the model predictions were typically centered around the 1:1 line (Figure 3.7). All coefficients for the models in Table 3.2 were significantly different from zero and we found no evidence that the relationship between the selected structure metrics and biomass depends on forest type (i.e., primary or secondary). We note that although some Fourier models included frequencies that differ by less than 25 cm in vertical wavelength, the low RMSPE values in Table 3.2 suggest that no redundant or unnecessary terms are included in the models. This is further supported by the high confidence of the models (Table 3.2),
which indicates that the correlation of biomass with the set of predictors is highly significant.

We emphasize that a particular model shown in Table 3.2 is only one of many possible formulations that can yield similar results. Formally, no interpretation of a particular set of predictors should be made as the selection of individual predictors is heavily influenced by the correlation between them (Ramsey & Schafer 2002). To investigate if there are preferred Fourier frequencies for biomass estimation at Tapajós, we show in Figure 3.8, for each data set (i.e., field, ALS, and GLAS), the 500 models with the smallest RMSPE. In this figure, each row represents a model (ranked by its RMSPE), with the dashes indicating the frequencies that compose the model, and the colors indicating whether the frequency is represented by the amplitude (blue) or phase (orange) of the Fourier transform (green means represented by both). Preferred frequencies are clearly identified for each data set. The most effective models for biomass estimation include the combination of a near zero Fourier frequency (equivalent to the average height) with groups of higher frequencies that vary with data set. Preferred frequencies that were similar across the three techniques included 0.01 and ~0.21 cyc/m (or 100 and ~5 m). For the lidar data sets, vertical wavelengths of 17–25 m were also important.

3.4.6 Extrapolating GLAS biomass estimates to larger spatial scales

In this section, we incorporate the Fourier approach into the up-scaling strategy to biomass estimation developed by Baccini et al. (2012) and demonstrate its performance in the generation of a wall-to-wall map of biomass for the Xingu River basin, Brazil (Figure 3.9), by combining new field plot data, GLAS measurements, and Moderate Resolution Imaging Spectroradiometer (MODIS) imagery.

In addition to the 30 plots measured at Tapajós in 2010, we used data collected within a total of 62 GLAS footprints sampled over two different sites in the study area (São Félix and upper Xingu). We converted the plot data to biomass using methods similar to those described in section 3.3.2.1. Figure 3.10 shows the field-estimated biomass versus that predicted by models based on traditional (left) and Fourier (right) metrics derived for these sites. The selected models are given in Table
3.3 along with measures of predictive ability. As observed for Tapajós, the models based on Fourier transforms of profiles performed significantly better than traditional models. The RMSPE of the Fourier model was 32% for São Félix and 9.2% for the upper Xingu (Table 3.3), representing an improvement in predictive ability of 25% and 63%, respectively. The poorer performance in São Félix compared to Tapajós, where the average biomass is comparable, is explained by the fact that height measurements were only available for 15% of the trees in São Félix. We developed a diameter-to-height relationship to estimate height for the remaining 85% of the trees, but the model explained only 26% of the variability with a RMSE of 9 m (51%). If heights are estimated at Tapajós using a site-specific diameter-to-height relationship calibrated using 15% of the data (as in São Félix), the RMS scatter of GLAS biomass estimates about field measurements increases by 55%. This is more than the difference in RMSE% of 47% observed between Tapajós (using actual heights) and São Félix (using allometrically-derived heights), as illustrated in Figures 3.7 and 3.10.

We combined the three data sets (i.e., Tapajós, São Félix, and upper Xingu) to establish a regional regression model between the field-based estimates of biomass and Fourier transforms of co-located GLAS waveforms. We excluded from the analysis a total of 18 plots for which tree-level wood density information was not available. The final regression model explained 80% of the variation in biomass at the footprint scale with a RMSE of 45 Mg ha\(^{-1}\) (31%) (Figure 3.11). Exclusion of the São Félix plots from the analysis resulted in a RMSE of 39 Mg ha\(^{-1}\) (24%) and \(R^2\) of 0.91. However, the São Félix data set includes forest types that are not represented by the other sites and we decided to keep it in the analysis despite the poor performance of the diameter-to-height relationship. We note that the association between structure and biomass did not differ significantly among sites (same slopes), although differences were observed for the mean biomass (different intercepts).

The regional regression model was used to generate a biomass prediction for each GLAS footprint sampled in the study area between 2006 and 2008 (~600,000 shots). Because the quality of each GLAS lidar “shot” varies depending on factors such as atmospheric conditions and ground topography (Lefsky et al. 2005), we applied an intentionally conservative screening approach to flag and remove
unreliable GLAS observations (Baccini et al. 2012). We also used a global forest change data set based on time-series analysis of Landsat ETM+ images (Hansen et al. 2013) to remove any GLAS shot where changes took place between GLAS and MODIS acquisitions (2006-2008). As a result of the screening process, just 30% (~180,000) of the original GLAS samples were retained for the study area. However, this up-scaling from relatively few field observations (74 plots) to thousands of GLAS footprints provided what was effectively a pseudo-inventory of biomass for the study area, with unprecedented coverage.

In the final stage, the “randomForest” nonparametric machine-learning algorithm (Brieman 2001) was used to model the relationship between the GLAS footprint-level predictions of biomass and MODIS NBAR (Nadir Bidirectional reflectance distribution function Adjusted Reflectance) data acquired over the study area. The NBAR product is screened for clouds, normalized for viewing and solar conditions, and composited on an 8-day interval (Ju et al. 2010). Pre-processing of the NBAR data included generation of a radiometrically consistent, cloud-free mosaic of standardized spectral reflectance information for the ca. 2007-2008 timeframe. The mosaic was assembled at the pixel level with only the highest-quality observations (NBAR QA values ≤ 3) being retained. A two-year timeframe was needed to obtain cloud-free observations. In addition to the spectral mean, the standard deviation and the EVI2 and NDII vegetation indices were computed. Following a sensitivity analysis, the MODIS-based randomForest models were developed using only pixels having 5 or more GLAS footprints inside each pixel. Average biomass values computed from ≥ 5 footprint-level estimates were required to adequately capture the spatial variability in biomass often observed within ca. 500 m MODIS pixels (Baccini et al. 2012). The final MODIS-based model explained 73% of the variance in biomass with a RMSE of 25 Mg ha⁻¹ (Figure 3.11). Uncertainty analysis based on a cross-validation approach where 80% of the pixels were used for calibration of the randomForest model and the remaining 20% for the assessment of the predictions resulted in a RMSE of 26 Mg ha⁻¹ and R² of 0.72. The biomass map produced with this model is shown in Figure 3.12. At the scale of the MODIS pixel, biomass is estimated to range from 0 to 251 Mg ha⁻¹. The mean biomass of the mapped area is
estimated at $81 \pm 50$ Mg ha$^{-1}$. The mean in the smaller region including protected areas and indigenous lands (the black polygon in Figure 3.12) is considerably higher, estimated at $112 \pm 31$ Mg ha$^{-1}$.

3.5 ACKNOWLEDGEMENTS

The research described in this paper was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration, under the Terrestrial Ecology program element. F.G. Gonçalves was partially funded by the CAPES Foundation, Brazilian Ministry of Education, through the CAPES/Fulbright Doctoral Program (process BEX-2684/06-3). The authors would like to thank Brazil’s Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/MCTI) and Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio/MMA) for research authorizations, and the Santarém office of the Large Scale Biosphere-Atmosphere Experiment in Amazonia (LBA) for providing logistical support. They would also like to thank E. Oliveira (UFAC), the IFT crew (R. Brito, A. Viana, M. Vitorino, and A. Jesus), and the local assistants J. Oliveira, R. Santos, I. Silva, and E. Pedroso for the invaluable help with the field acquisitions. Airborne lidar data were acquired with support from the United States Agency for International Development and the U.S. Department of State with the technical assistance of the Brazilian Corporation for Agricultural Research (EMBRAPA) and the U.S. Forest Service Office of International Programs.

References


estimates and forest structure metrics at footprint level. Remote Sensing of Environment, 115(11), 2786–2797.


Figure 3.1. Geographical location of the Tapajós National Forest, PA, Brazil, outlined in yellow. The white lines are GLAS tracks from 2003–2009, the red rectangles are ALS measured in 2012 and 2013, and the blue targets are the locations of the plots sampled in this study (2010 and 2013).
Figure 3.2. Side view of the modeled canopy structure of (a) a primary forest (AGB = 421.0 Mg ha$^{-1}$), and (b) an early successional forest (AGB = 12.1 Mg ha$^{-1}$) measured at Tapajós, represented in their correct relative sizes. In (a), we display the ellipsoidal model in (1) for the tallest tree, which is 42.8 m (note that some trees are displayed as half ellipsoids for visualization purposes). In (b), we show the 1-m height bins that were used to calculate vegetation density profiles as in (2).
Figure 3.3. GLAS waveform for the primary forest plot displayed in Figure 3.2a. The ground/canopy return is shown in black (with the ground return in bold), and the noise component is shown in gray, as determined by the threshold values calculated as $\mu_{\text{noise}} + 4\sigma_{\text{noise}}$ (note the difference in the noise distribution between the initial and final portions of the waveform, with $\mu_{\text{noise}}$ represented by a vertical gray line for each portion). The elevation of the ground peak is indicated by the dashed red line, and the cumulative return by the solid red line. Some structural metrics are also shown, including the mean ($Z$) and maximum ($Z_{\text{max}}$) heights of the profile, quartiles of the cumulative return ($Z_{25}$, $Z_{50}$, and $Z_{75}$), and the front slope ($m_{\phi} = \tan \phi$), as described in section 3.3.4.
Figure 3.4. Relative vegetation density profiles from field (shaded polygon), GLAS (black line), and ALS (gray line) measurements for nine field plots at Tapajós. All profiles were normalized to unity by dividing by the maximum amplitude. The plots are arranged alphabetically in order of increasing biomass (a proxy for age), with panels a–d showing secondary forests, and panels e–i showing old-growth forests.
Figure 3.5. Relationship between species diversity (as measured by the Shannon index) and (a) mean stand-level wood density (calculated on a per stem basis), (b) basal area, (c) mean height, and (d) aboveground biomass for the 30 plots measured in 2010. The Shannon index is calculated as 

$$-\sum_{i=1}^{S} p_i \ln p_i,$$

where $p_i$ is the number of individuals belonging to the $i$th species divided by the total number of individuals. The curved lines show the fit of a logarithmic model to the data.
Figure 3.6. Relationship between height-based metrics derived from GLAS (open symbols) and ALS (filled symbols) versus the corresponding field metrics: (a) mean canopy height, (b) standard deviation, (c–e) quartile heights, and (f) maximum canopy height. The dotted lines are 1:1 lines, the solid lines are the estimated regressions (GLAS: black, ALS: gray), and “r” is the pooled correlation coefficient. The box-and-whisker plot in panel (g) shows the distribution of the differences between corresponding lidar and field metrics, calculated separately for primary (PF) and secondary (SF) forests.
Figure 3.7. Field-estimated aboveground biomass versus that predicted by traditional and Fourier metrics derived from field, ALS, and GLAS profiles for Tapajós, as indicated in the panels. The specification of the models is given in Table 3.2 along with validation results. The diagonal lines are $x = y$. 
Figure 3.8. Graphical representation of the 500 Fourier-based models with the smallest RMSPE. The maximum number of predictors allowed was 6 and all models included an intercept (see text for details).
Figure 3.9. Location of the Xingu River basin, Brazil. The area mapped (the red rectangle) encompasses the entire basin (blue polygon), with a buffer of 50 Km (total area of 1,107,730 km²). The polygon in green is an area chosen for a REDD project in the municipality of São Félix do Xingu, PA, which includes protected areas and indigenous lands. The black points are field plots used to calibrate GLAS, and the polygon in beige is the Tapajós National Forest.
Figure 3.10. Field-estimated aboveground biomass versus that predicted by traditional (left) and Fourier (right) metrics derived from GLAS profiles acquired over two sites in the Xingu region (São Félix and Upper Xingu). The specification of the models is given in Table 3.3 along with validation results. The diagonal lines are $x = y$. 
Figure 3.11. Left: Performance of the regional GLAS-biomass model based on Fourier metrics; Right: Average biomass estimates from GLAS (Fourier model) versus “randomForest” MODIS estimates.
Figure 3.12. Aboveground live biomass density map for the Xingu River basin, Brazil, produced from field plot data, GLAS lidar measurements, and MODIS imagery. The inset figure on the right shows details of the area chosen for a REDD project in the municipality of São Félix do Xingu, PA, with biomass aggregated in classes of 50 Mg ha$^{-1}$ to better show the spatial patterns. The histogram on the bottom shows the frequency distribution of biomass for the full map in units of Mg ha$^{-1}$. 
Table 3.1. Root mean square deviation (RMSD), bias (i.e., mean difference), and results of simple regressions between corresponding field and lidar metrics: mean canopy height (MCH), standard deviation of height (SD), quartile heights (H25, H50, and H75), and maximum canopy height (H100). The number of plots was 39 for ALS and 30 for GLAS.

<table>
<thead>
<tr>
<th>Metric</th>
<th>1:1 line</th>
<th>Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSD (m)</td>
<td>Bias (m)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SF</td>
</tr>
<tr>
<td>ALS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCH</td>
<td>3.3 (21.6%)</td>
<td>-2.9 (-25.0%)</td>
</tr>
<tr>
<td>SD</td>
<td>2.2 (39.8%)</td>
<td>0.8 (22.7%)</td>
</tr>
<tr>
<td>H25</td>
<td>3.5 (33.3%)</td>
<td>-2.9 (-36.6%)</td>
</tr>
<tr>
<td>H50</td>
<td>3.3 (23.1%)</td>
<td>-1.9 (-18.9%)</td>
</tr>
<tr>
<td>H75</td>
<td>4.0 (21.5%)</td>
<td>-1.8 (-13.9%)</td>
</tr>
<tr>
<td>H100</td>
<td>6.0 (19.6%)</td>
<td>-0.5 (-0.4%)</td>
</tr>
<tr>
<td>GLAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCH</td>
<td>2.8 (17.0%)</td>
<td>-1.8 (-15.4%)</td>
</tr>
<tr>
<td>SD</td>
<td>2.6 (42.4%)</td>
<td>-0.3 (-5.5%)</td>
</tr>
<tr>
<td>H25</td>
<td>2.4 (21.5%)</td>
<td>-0.6 (-7.3%)</td>
</tr>
<tr>
<td>H50</td>
<td>3.3 (21.4%)</td>
<td>-0.7 (-6.0%)</td>
</tr>
<tr>
<td>H75</td>
<td>4.6 (22.7%)</td>
<td>-1.7 (-11.1%)</td>
</tr>
<tr>
<td>H100</td>
<td>5.9 (17.1%)</td>
<td>-2.4 (-11.6%)</td>
</tr>
</tbody>
</table>
Table 3.2. Root mean square prediction error (RMSPE), cross-validation coefficient of determination ($R^2_{CV}$), and confidence for the selected regression models at Tapajós. The predictors included in the traditional models are described in section 3.3.4. $A$ and $\phi$ are, respectively, the amplitude and phase of the Fourier transform of the profile at the spatial frequency given by the subscript number following the symbol over 100 (e.g., $\phi_{01}$ is the phase at 0.01 cyc/m and $A_{18}$ is the amplitude at 0.18 cyc/m).

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictors</th>
<th>RMSPE (%)</th>
<th>Mean (95% CI)</th>
<th>SD</th>
<th>Range</th>
<th>$R^2_{CV}$</th>
<th>Confidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best traditional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>MCH</td>
<td>25.2</td>
<td>24.8–25.6</td>
<td>6.9</td>
<td>8.7–51.5</td>
<td>0.92</td>
<td>&gt; 99.9</td>
</tr>
<tr>
<td>ALS</td>
<td>MCH, $m_{gr}$, $R_{gtr}$</td>
<td>33.0</td>
<td>32.2–33.8</td>
<td>13.6</td>
<td>4.9–64.6</td>
<td>0.91</td>
<td>&gt; 99.9</td>
</tr>
<tr>
<td>GLAS</td>
<td>H50, SD</td>
<td>29.0</td>
<td>28.4–29.5</td>
<td>9.1</td>
<td>4.1–52.7</td>
<td>0.88</td>
<td>&gt; 99.9</td>
</tr>
<tr>
<td><strong>Best Fourier</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field</td>
<td>$\phi_{01}$, $A_{18}$, $A_{21}$, $A_{25}$, $A_{30}$</td>
<td>21.6</td>
<td>21.2–22.0</td>
<td>6.2</td>
<td>5.6–41.5</td>
<td>0.94</td>
<td>&gt; 99.9</td>
</tr>
<tr>
<td>ALS</td>
<td>$\phi_{01}$, $A_{04}$, $A_{16}$, $\phi_{16}$, $A_{21}$, $\phi_{25}$</td>
<td>24.0</td>
<td>23.5–24.5</td>
<td>8.2</td>
<td>4.2–60.1</td>
<td>0.94</td>
<td>99.8</td>
</tr>
<tr>
<td>GLAS</td>
<td>$A_{01}$, $\phi_{19}$, $A_{20}$, $A_{21}$, $\phi_{22}$</td>
<td>21.6</td>
<td>21.2–22.1</td>
<td>6.7</td>
<td>2.9–49.1</td>
<td>0.93</td>
<td>&gt; 99.9</td>
</tr>
</tbody>
</table>
Table 3.3. Root mean square prediction error (RMSPE), cross-validation coefficient of determination \( (R_{CV}^2) \), and confidence for the selected regression models at São Félix (SFX) and the Upper Xingu (UX). The predictors included in the traditional models are described in section 3.3.4. \( A \) and \( \phi \) are, respectively, the amplitude and phase of the Fourier transform of the profile at the spatial frequency given by the subscript number following the symbol over 100 (e.g., \( \phi_{01} \) is the phase at 0.01 cyc/m and \( A_{18} \) is the amplitude at 0.18 cyc/m).

<table>
<thead>
<tr>
<th>Site</th>
<th>Variables</th>
<th>RMSPE (%)</th>
<th>Mean (95% CI)</th>
<th>SD</th>
<th>Range</th>
<th>( R_{CV}^2 )</th>
<th>Confidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Best traditional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFX</td>
<td>H50, ( R_g )</td>
<td>42.4 (41.8–43.0)</td>
<td>9.4</td>
<td>17.5–106.4</td>
<td>0.39</td>
<td>&gt; 99.9</td>
<td></td>
</tr>
<tr>
<td>UX</td>
<td>MCH, H25, H75, H100</td>
<td>24.9 (24.4–25.5)</td>
<td>9.1</td>
<td>5.5–47.7</td>
<td>0.65</td>
<td>&gt; 99.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Best Fourier</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFX</td>
<td>( A_{06}, A_{07}, A_{23}, A_{28}, A_{29}, \phi_{11} )</td>
<td>31.7 (31.2–32.2)</td>
<td>7.4</td>
<td>7.8–52.9</td>
<td>0.62</td>
<td>&gt; 99.9</td>
<td></td>
</tr>
<tr>
<td>UX</td>
<td>( A_{07}, A_{15}, A_{19}, A_{24}, \phi_{24}, \phi_{25} )</td>
<td>9.2 (9.0–9.5)</td>
<td>4.5</td>
<td>1.3–24.6</td>
<td>0.88</td>
<td>&gt; 99.9</td>
<td></td>
</tr>
</tbody>
</table>
3.6 APPENDIX

Bayesian model averaging

Given the data \((y, X)\), we estimated the expected value of AGB for each plot, \(E[AGB|y,X]\), using the weighted average

\[
E[AGB|y,X] = \sum_{k=1}^{K} E[AGB|y,X,M_k] P(M_k|y,X) \quad (A1)
\]

where \(K\) is the number of regression models \(M\) considered, \(E[AGB|y,X,M_k]\) is the AGB prediction under model \(M_k\), and \(P(M_k|y,X)\) is the weight, or posterior probability that \(M_k\) is correct, given by the Bayes’ theorem

\[
P(M_k|y,X) = \frac{P(y|M_k,X) P(M_k)}{\sum_{j=1}^{K} P(y|M_j,X) P(M_j)} \quad (A2)
\]

In (A2), \(P(M_k)\) is the probability that \(M_k\) is the true model prior to observing the data, and \(P(y|M_k,X)\) is the marginal likelihood of \(M_k\), obtained by integrating its likelihood, \(P(y|\theta_k,M_k)\), with the prior density on the parameters, \(P(\theta_k|M_k)\)

\[
P(y|M_k,X) = \int P(y|\theta_k,M_k) P(\theta_k|M_k) d\theta_k \quad (A3)
\]

where \(\theta_k = (\beta_0, \beta_j, \sigma)\) as in (8). We approximated \(P(y|M_k,X)\) by employing a prior structure as in Feldkircher et al. (2009)\(^{10}\), which leads to a simple closed-form

---

\(^8\) For the analysis based on traditional metrics, all possible models were considered. For the Fourier analysis, due to the large number of candidate metrics (i.e. 60), we averaged over a reduced set of models supported by the data. (A1) was approximated using a Markov chain Monte Carlo (MCMC) sampler as described by Raftery et al. (1997).

\(^9\) The probability of the data given the model and the parameter values.
expression that is proportional to $R^2$ and includes a penalty for the number of predictors included in the model, $p_k$

$$P(y|M_k,X,g) \propto (1 + g)^{-\frac{p_k}{2}} \left( 1 - \frac{g}{1 + g} R_k^2 \right)^\left(\frac{n-1}{2}\right)$$  \hspace{1cm} (A4)

where $n$ is the number of observations in the data set, and $g$ is a scalar parameter that determines how much importance is attributed to the priors (see footnote 4), controlling in part how the weights of (A2) are distributed across the model space $M_1, M_2, \ldots, M_K$. We specified $g$ automatically following methods of Liang et al. (2008) (i.e. using the hyper-$g$ prior), which reduce the risks associated with the selection of an arbitrary fixed $g^{11}$ by relying on a Bayesian approach to update its estimate for each model $M$, taking into account the noise level $\sigma$ of the data (Feldkircher et al. 2009).

Prior model probabilities $P(M_k)$ were specified using the beta-binomial framework of Ley & Steel (2009), which discriminates among models according to their sizes, $p_k$. For each model, $P(M_k)$ was specified as

$$P(M_k) = \delta^{p_k} (1 - \delta)^{P-p_k}$$  \hspace{1cm} (A5)

where $P$ is the total number of candidate predictors, and $\delta$ is the probability of any given predictor being included in $M_k$, given the expected model size $m$. Following Ley & Steel (2009), $\delta$ was made random, beta-distributed$^{12}$ with shape parameters $a$ and $b$, i.e. $\delta \sim \text{Be}(a, b)$, where $a$ was set to 1 and $b$ to $(P - m)/m$, leading to a compound distribution proportional to

---

$^{10}$ The natural conjugate framework employing Zellner’s $g$ prior, which expresses the lack of prior knowledge on the distribution of the parameters of (8), $P(\theta_k|M_k)$, by specifying $P(\beta_0) \propto 1, P(\sigma) \propto 1/\sigma$, and $\beta_i \sim \mathcal{N}(0, g\sigma^2(X'X)^{-1})$.

$^{11}$ i.e. the risk of artificially inflating the weights $P(M_k|y,X)$ of “wrong” models or, on the other hand, underestimating the weights of models that reflect the “true” underlying relationship.

$^{12}$ Note that by making $\delta$ random with a beta distribution, the beta-binomial approach increases the prior uncertainty over models by dispersing the binomial distribution in (A5). This reduces the impact that the choice of $m$ has on the weights of (A2) relative to fixing $\delta = m/P$ or simply setting an uninformative flat prior over all models $P(M_k) = 1/K$ (or, equivalently, $\delta = 0.5$).
$P(M_k) \propto \Gamma(1 + p_k) \Gamma\left(\frac{P - m}{m} + P - p_k\right)$  \hspace{1cm} (A6)

where $\Gamma(\eta)$ denotes the gamma function $(\eta - 1)!$. 
Figure 3.A1. Relationships between crown radius and (a) stem diameter (on the square root scale), (b) crown depth, and (c) total height, with the estimated regression lines for secondary (gray) and old-growth forests (black). Panel (d) shows observed crown radius versus that predicted by a multiple regression model including all three variables listed above (see Table 3.A1).
Figure 3.A2. Relative vegetation density profiles derived from field measurements of crown radius ($C_R$), crown depth ($C_D$), and total height ($H_T$) for nine 12.5 m x 50 m subplots sampled in 2010 (shaded polygons). The dotted lines represent profiles estimated using modeled $C_R$ (first eq. in Table 3.A1) instead of actual $C_R$ measurements. The solid lines show profiles estimated from stem diameter alone (i.e., using modeled $C_R$, $C_D$, and $H_T$ as in the second part of Table 3.A1).
Figure 3.A3. Dendrogram for an agglomerative clustering of the 2010 dataset based on the Hellinger distance (Kindt & Coe 2005). Sites that are grouped into the same cluster are more similar in species composition than sites that are grouped into different clusters. As expected, the clustering algorithm found clear differences in species composition between secondary (SF) and old-growth (OGF) forests. It was also able to detect differences between primary (PF) and selectively-logged (SLF) forests within the OGF group, and between two PF sites located less than 10 km apart: “Tapajós km 67” (plots 27, 28, 29, and 30) and “Tapajós km 72” (plots 8, 9, 10, and 11).
Figure 3.A4. Field profiles derived for six plots at Tapajós that were measured in 2010 and remeasured in 2013. The 2010 plots were temporary and the trees were not tagged or mapped to facilitate location and remeasurement in 2013. As a result, the observed differences reflect the effect of not only temporal but also spatial variation, with the estimated distance between plot centers provided for each case.
Table 3.A1. Results from the regression of crown radius ($C_R$, m) on stem diameter ($\sqrt{D}$, cm), crown depth ($C_D$, m), and total height ($H_T$, m). These equations were used in this study to predict $C_R$ for trees where this measurement was not taken in the field (75% of the plot area for the plots measured in 2010). For reference, we also present equations for predicting height and crown dimensions from stem diameter alone. Nonlinear models were also explored, but did not improve the fit.

<table>
<thead>
<tr>
<th>Equation</th>
<th>Coefficient (standard error)</th>
<th>$n$</th>
<th>$R^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_R = \beta_0 + \beta_1 \sqrt{D} + \beta_2 C_D + \beta_3 H_T$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$PF$</td>
<td>$-0.280 (0.151)$</td>
<td>415</td>
<td>0.59</td>
<td>0.9 (28%)</td>
</tr>
<tr>
<td>$SF$</td>
<td>$-1.090 (0.111)$</td>
<td>502</td>
<td>0.74</td>
<td>0.6 (29%)</td>
</tr>
<tr>
<td>$H_T = \beta_0 + \beta_1 \ln(D)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$PF$</td>
<td>$-17.514 (1.546)$</td>
<td>415</td>
<td>0.58</td>
<td>5.2 (28%)</td>
</tr>
<tr>
<td>$SF$</td>
<td>$-6.003 (0.705)$</td>
<td>502</td>
<td>0.57</td>
<td>3.3 (28%)</td>
</tr>
<tr>
<td>$C_D = \beta_0 + \beta_1 \sqrt{D}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$PF$</td>
<td>$-0.367 (0.159)$</td>
<td>415</td>
<td>0.55</td>
<td>0.9 (30%)</td>
</tr>
<tr>
<td>$SF$</td>
<td>$-1.331 (0.136)$</td>
<td>502</td>
<td>0.59</td>
<td>0.8 (37%)</td>
</tr>
<tr>
<td>$C_P = \beta_0 + \beta_1 \sqrt{D}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$PF/SF$</td>
<td>$-1.347 (0.261)$</td>
<td>917</td>
<td>0.50</td>
<td>2.5 (40%)</td>
</tr>
</tbody>
</table>
Chapter 4. Conclusions

Methods for obtaining accurate, spatially explicit estimates of biomass density in tropical forests are required to reduce uncertainties in the global carbon cycle, and to support international climate agreements and emerging carbon markets. In this research, we used airborne laser scanning, space-based lidar observations, and detailed in situ measurements made at the Tapajós National Forest, Brazil, to advance methods of biomass estimation in tropical regions from remotely sensed structure.

The objective of the first study (Chapter 2) was to use field plot data collected at Tapajós to gain a better understanding of the uncertainty associated with plot-level biomass estimates obtained specifically for calibration of remote sensing measurements in tropical forests. We found that the overall uncertainty in the field biomass was typically 25% for both secondary and primary forests, but ranged from 16 to 53%. Co-location errors were typically 13–26% and dominated the overall uncertainty in both mid-successional and primary stands. In comparison to the co-location error, measurement and allometric errors were relatively small. The individual contributions were generally below 15%, and slightly lower in secondary forests than in primary forests. In primary forests, the error of estimating biomass for the 5–10 cm diameter class contributed less than 1% to the total variance and could safely be neglected. However, this error was about seven times larger in mid-successional forests, being comparable to other sources in magnitude. In secondary forests, the projection of biomass values backward in time three years induced errors of the order of 7–19%. This term dominated the uncertainties in early successional stands, accounting for about half of the total variance on average, and represented the second largest component in mid-successional stands.

These results show that the uncertainty in the field biomass can represent a dominant term in the overall error budget for biomass maps derived from remote sensing, and therefore should be taken into account in calibration and validation efforts. Because co-location and temporal errors have the potential to account for a large fraction of the total variance, they emerge as obvious targets for reducing uncertainty in studies relating tropical forest biomass to remotely sensed data.
Although measurement and allometric errors were relatively unimportant when considered alone, combined they accounted for roughly 30% of the total variance on average (as much as 64% in individual plots) and therefore should not be ignored. Steps can be taken to reduce uncertainties in height and wood density, as well as in allometric equations. However, reducing co-location and temporal errors may be a more cost-effective solution for reducing the overall uncertainty when resources are limited.

The objectives of the second study (Chapter 3) were to (1) test and refine methods for the extraction of structural information from lidar data; (2) determine the accuracy of lidar estimates of structure in relation to detailed field measurements of vertical structure; (3) develop and validate structure-based models for optimal prediction of aboveground biomass; and (4) integrate the resulting models in a multi-stage scaling approach to biomass estimation and produce a wall-to-wall map of biomass for a test site in the Amazon.

We found that lidar and field-based vegetation density profiles obtained for stands spanning a wide range in structure and biomass at Tapajós showed good qualitative and quantitative agreement. Simple regressions between corresponding field and lidar metrics showed that GLAS and ALS were similarly sensitive to structural differences, explaining 74–87% of the variation in field measurements. The RMS differences between field and remote sensing metrics ranged between 11 and 20%, and were typically smaller than the uncertainty in field-based measurements of individual tree height. A new approach to biomass estimation based on Fourier transforms of lidar profiles resulted in models that explained 93–94% of the variation in biomass, with a cross-validation RMS error of 35–38 Mg ha⁻¹ (22–24%). This represented an improvement over the prediction performance of traditional models of 25% for GLAS and 27% for ALS, with the improvement occurring primarily in high-biomass forests. The best Fourier models included information at 0.01 cyc/m (100 m) and a combination of Fourier transforms including vertical wavelengths between ~4 and 6 m. There was some indication that vertical wavelengths of 17–25 m are also important.
Test of the Fourier approach using data from two different sites in the Amazon suggested that the improvement in biomass estimation performance is consistent across sites. Future studies should further evaluate how preferred Fourier frequencies vary among forests differing in structure to shed light on the potential of Fourier transforms and to develop a better basis for estimating biomass and carbon stocks from these measurements at local to regional scales. On the other hand, studies should also focus on understanding biophysical mechanisms by which canopies are organized and developing mechanistic approaches to biomass estimation, in addition to exploring statistical correlations between lidar metrics and biomass. Dynamic insight as to why given Fourier frequencies correlate with biomass may help to generate improved biomass estimation models.