

1 Allometric equations for integrating remote sensing imagery into forest 2 monitoring programs

3 Tommaso Jucker¹, John Caspersen², Jérôme Chave³, Cécile Antin⁴, Nicolas Barbier⁵, Frans
4 Bongers⁶, Michele Dalponte⁷, Karin Y. van Ewijk⁸, David I. Forrester⁹, Matthias Haeni¹⁰,
5 Steven I. Higgins¹¹, Robert J. Holdaway¹², Yoshiko Iida¹³, Craig Lorimer¹⁴, Peter L.
6 Marshall¹⁵, Stéphane Momo¹⁶, Glenn R. Moncrieff¹⁷, Pierre Ploton¹⁸, Lourens Poorter¹⁹,
7 Kassim Abd Rahman²⁰, Michael Schlund²¹, Bonaventure Sonké²², Frank J. Sterck²³, Anna T.
8 Trugman²⁴, Vladimir A. Usoltsev²⁵, Mark C. Vanderwel²⁶, Peter Waldner²⁷, Beatrice M. M.
9 Wedeux²⁸, Christian Wirth²⁹, Hannsjörg Wöll³⁰, Murray Woods³¹, Wenhua Xiang³², Niklaus
10 E. Zimmermann³³ and David A. Coomes³⁴

11 ¹Forest Ecology and Conservation group, Department of Plant Sciences, University of Cambridge, UK
12 (tj272@cam.ac.uk)

13 ²Faculty of Forestry, University of Toronto, 33 Willcocks Street, Toronto, ON, M5S 3B3, Canada and Swiss
14 Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland
15 (john.caspersen@utoronto.ca)

16 ³Laboratoire Evolution et Diversité Biologique, UMR5174, CNRS/ Université Paul Sabatier Bâtiment 4R1, 118
17 route de Narbonne F-31062 Toulouse cedex 9, France (jerome.chave@univ-tlse3.fr)

18 ⁴Institut de Recherche pour le Développement, UMR AMAP, Montpellier, France and Institut Français de
19 Pondichéry, UMIFRE CNRS-MAE 21, Puducherry, India (cecile.madelaine-antin@cirad.fr)

20 ⁵Institut de Recherche pour le Développement, UMR AMAP, Montpellier, France (nicolas.barbier@ird.fr)

21 ⁶Forest Ecology and Forest Management Group, Wageningen University, PO Box 47, 6700 AA Wageningen,
22 the Netherlands (Frans.Bongers@wur.nl)

23 ⁷Department of Sustainable Agro-ecosystems and Bioresources, Research and Innovation Centre, Fondazione E.
24 Mach, Via E. Mach 1, 38010 San Michele all'Adige, Italy (michele.dalponte@fmach.it)

25 ⁸Department of Geography and Planning, Queen's University, Kingston, ON, Canada (ekyv@queensu.ca)

26 ⁹Chair of Silviculture, Faculty of Environment and Natural Resources, Freiburg University, Tennenbacherstr. 4,
27 79108 Freiburg, Germany (david.forrester@waldbau.uni-freiburg.de)

28 ¹⁰Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland
29 (matthias.haeni@wsl.ch)

30 ¹¹Department of Botany, University of Otago, PO Box 56, Dunedin 9016, New Zealand
31 (steven.higgins@otago.ac.nz)
32 ¹²Landcare Research, PO Box 69040, Lincoln 7640, New Zealand (holdawayr@landcareresearch.co.nz)
33 ¹³Kyushu Research Center, Forestry and Forest Products Research Institute, Kurokami, Chuo-ku, Kumamoto
34 860-0862, Japan (yyoshiko503@gmail.com)
35 ¹⁴Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, Madison, WI 53706, USA
36 (clorimer@wisc.edu)
37 ¹⁵Faculty of Forestry, University of British Columbia, 2424 Main Mall, BC, V6T 1Z4, Vancouver, Canada
38 (peter.marshall@ubc.ca)
39 ¹⁶Institut de Recherche pour le Développement, UMR AMAP, Montpellier, France and Laboratoire de
40 Botanique systématique et d'Ecologie, Département des Sciences Biologiques, Ecole Normale Supérieure,
41 Université de Yaoundé I, Yaoundé, Cameroon (takoudjournomo@gmail.com)
42 ¹⁷Fynbos Node, South African Environmental Observation Network (SAEON), Centre for Biodiversity
43 Conservation, Kirstenbosch Gardens, Private Bag X7, Rhodes Drive, Claremont, 7735, Cape Town, South
44 Africa (glenn.moncrieff@gmail.com)
45 ¹⁸Institut de Recherche pour le Développement, UMR AMAP, Montpellier, France (p.ploton@gmail.com)
46 ¹⁹Forest Ecology and Forest Management Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen,
47 the Netherlands (lourens.poorter@wur.nl)
48 ²⁰Forest Research Institute of Malaysia, Kepong 52109, Selangor, Malaysia (rahmank@frim.gov.my)
49 ²¹Department of Earth Observation, Friedrich-Schiller University, Loebdergraben 32, 07743 Jena, Germany
50 (michael.schlund@uni-jena.de)
51 ²²Laboratoire de Botanique systématique et d'Ecologie, Département des Sciences Biologiques, Ecole Normale
52 Supérieure, Université de Yaoundé I, Yaoundé, Cameroon (bsonke_1999@yahoo.com)
53 ²³Forest Ecology and Forest Management Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen,
54 the Netherlands (frank.sterck@wur.nl)
55 ²⁴Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, 08544, USA
56 (att@princeton.edu)
57 ²⁵Botanical Garden of the Russian Academy of Sciences (Ural branch), 620144 Yekaterinburg, Russia and Ural
58 State Forest Engineering University, 620100 Yekaterinburg, Russia (usoltsev50@mail.ru)
59 ²⁶Department of Biology, University of Regina, 3737 Wascana Pkwy, Regina, SK, S4S 0A2, Canada
60 (mark.vanderwel@uregina.ca)
61 ²⁷Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland
62 (peter.waldner@wsl.ch)
63 ²⁸Forest Ecology and Conservation group, Department of Plant Sciences, University of Cambridge, UK
64 (beatrice.wedeux@gmail.com)
65 ²⁹Systematic Botany and Functional Biodiversity, Institute of Biology, University of Leipzig, Leipzig, Germany
66 and German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
67 (cwirth@uni-leipzig.de)

68 ³⁰Conservation and Natural Resources Management, Sommersbergsestr. 291, A-8990, Bad Aussee, Austria
69 (hannsjorg.woll@gmail.com)

70 ³¹Ontario Ministry of Natural Resources, North Bay, ON, P1A 4L7, Canada (murray.woods@ontario.ca)

71 ³²Faculty of Life Science and Technology, Central South University of Forestry and Technology, Changsha,
72 Hunan, 410004, China (xiangwh2005@163.com)

73 ³³Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland
74 (niklaus.zimmermann@wsl.ch)

75 ³⁴Forest Ecology and Conservation group, Department of Plant Sciences, University of Cambridge, UK
76 (dac18@cam.ac.uk)

77 **Correspondence:** Tommaso Jucker; Forest Ecology and Conservation group, Department of
78 Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK; Email:
79 tj272@cam.ac.uk; Phone: +44 (0)1223 333911; Fax: +44 (0)1223 333953.

80 **Statement of authorship:** T.J. and D.A.C. designed the study, with the assistance of J.P.C.
81 and J.C.; T.J. and J.P.C. compiled the database, with all authors providing data; T.J. analysed
82 the data and wrote the first draft of the manuscript, with all authors contributing substantially
83 to revisions.

84 **Running head:** Global diameter and biomass allometries

85 **Key words:** aboveground biomass; airborne laser scanning; carbon mapping; crown
86 architecture; height–diameter allometry; stem diameter distributions

87 **Type of paper:** Primary research article

88 **Abstract**

89 Remote sensing is revolutionizing the way we study forests, and recent technological
90 advances mean we are now able – for the first time – to identify and measure the crown
91 dimensions of individual trees from airborne imagery. Yet in order to make full use of these
92 data for quantifying forest carbon stocks and dynamics, a new generation of allometric tools
93 which have tree height and crown size at their centre are needed. Here, we compile a global
94 database of 108753 trees for which stem diameter, height and crown diameter have all been
95 measured, including 2395 trees harvested to measure aboveground biomass. Using this
96 database, we develop general allometric models for estimating both the diameter and
97 aboveground biomass of trees from attributes which can be remotely sensed – specifically
98 height and crown diameter. We show that tree height and crown diameter jointly quantify the
99 aboveground biomass of individual trees, and find that a single equation predicts stem
100 diameter from these two variables across the world’s forests. These new allometric models
101 provide an intuitive way of integrating remote sensing imagery into large-scale forest
102 monitoring programs, and will be of key importance for parameterizing the next generation of
103 dynamic vegetation models.

104 **Introduction**

105 Forests are a key component of the terrestrial carbon cycle (Beer *et al.*, 2010; Pan *et al.*,
106 2011), making forest conservation of critical importance for mitigating climate change
107 (Agrawal *et al.*, 2011). Yet effectively managing forests as carbon sinks is predicated on the
108 assumption that carbon stocks can be quantified with accuracy across extensive and often
109 remote areas. Traditionally, forest carbon stocks have been assessed by measuring the
110 diameter (and sometimes height) of trees in permanent field plots, and then using allometric
111 equations to estimate biomass (Malhi *et al.*, 2006; Pan *et al.*, 2011; Anderson-Teixeira *et al.*,
112 2015). Recently, however, we have begun to see a move towards remote sensing as the
113 primary tool for monitoring forest carbon (Saatchi *et al.*, 2011; Baccini *et al.*, 2012; Avitabile
114 *et al.*, 2016). Airborne laser scanning (ALS) is particularly promising in this regard (Asner &
115 Mascaro, 2014; Asner *et al.*, 2014), allowing the 3D structure of entire forest landscapes to be
116 reconstructed in detail using high-frequency laser scanners mounted on airplanes or
117 unmanned aerial vehicles. Importantly, advances in both sensor technology and computation
118 mean we are now able – for the first time – to reliably identify and measure the crown
119 dimensions of individual trees using ALS (Yao *et al.*, 2012; Duncanson *et al.*, 2014;
120 Shendryk *et al.*, 2016), marking a fundamental shift in the way we census forests. To
121 facilitate this transition, we aim to develop allometric equations for estimating a tree’s
122 diameter and aboveground biomass based on attributes which can be remotely sensed –
123 namely tree height and crown diameter – enabling airborne imagery to be fully integrated into
124 existing carbon monitoring programs (Fig. 1).

125 While ALS opens the door to rapidly and accurately measuring the height and crown
126 dimensions of millions of trees (Duncanson *et al.*, 2015), it also poses the challenge of how

127 best to use these data to estimate aboveground biomass. Current allometries rely on stem
128 diameter as a key input for estimating biomass (e.g., Chave *et al.* 2014). But because
129 diameters cannot be measured directly through ALS, new approaches that have tree height
130 and crown dimensions at their centre are needed. We see two possible solutions for
131 integrating tree-level ALS data into biomass monitoring programs: the first is to use tree
132 height and crown dimensions to predict diameters, allowing biomass to be estimated using
133 existing allometric equations (Dalponte & Coomes, 2016). The second is to develop
134 equations that estimate biomass directly from tree height and crown size, thereby bypassing
135 diameter altogether.

136 *Approach 1: estimating diameter*

137 Theory based on the mechanical and hydraulic constraints to plant growth predicts that tree
138 height (H , in m) should scale with diameter (D , in cm) following a power-law relationship
139 with an invariant scaling exponent of $2/3$ ($H \propto D^{2/3}$; West *et al.*, 1999). This would suggest
140 that measuring tree height should be sufficient for estimating diameter. However, growing
141 evidence indicates that this is unlikely to be the case (Muller-Landau *et al.*, 2006): not only
142 do H – D allometries vary considerably among and within species, as well as in relation to
143 climate and stand structure (Banin *et al.*, 2012; Lines *et al.*, 2012; Hulshof *et al.*, 2015; Jucker
144 *et al.*, 2015), but power-law relationships also fail to adequately capture the asymptotic nature
145 of height growth (Muller-Landau *et al.*, 2006; Banin *et al.*, 2012; Feldpausch *et al.*, 2012;
146 Iida *et al.*, 2012; Chave *et al.*, 2014). Trees typically invest heavily in height growth when
147 young to escape shaded understories – rapidly approaching their maximum height – but then
148 continue to grow in diameter throughout their lives (King, 2005). This makes estimating the
149 diameter of large trees challenging, as trees of similar height can have very different

150 diameters – which is problematic given that large-diameter trees hold most of the biomass
151 (Slik *et al.*, 2013; Bastin *et al.*, 2015). In this context, information on crown size may prove
152 key to accurately estimating a tree’s diameter. While height growth tends to slow rapidly in
153 large trees, lateral crown expansion does not, requiring a continued investment in stem
154 growth on the tree’s part to ensure structural stability and hydraulic function (Sterck &
155 Bongers, 2001; King & Clark, 2011; Iida *et al.*, 2012). As a result, crown width and stem
156 diameter tend to be strongly coupled, even in large trees (Hemery *et al.*, 2005).

157 *Approach 2: estimating aboveground biomass*

158 Estimating the diameter of individual trees from remotely sensed data is an appealing
159 prospect: not only would it provide a way to quantify biomass stocks, but would also allow
160 other forest attributes of interest to be reconstructed with ease (e.g., stem diameter
161 distributions). However, it also presents a challenge from the point of view of biomass
162 estimation, as biomass allometries typically have diameter as a squared term in the equation
163 (Zianis *et al.*, 2005; Chave *et al.*, 2014; Chojnacky *et al.*, 2014), meaning that even small
164 errors in diameter predictions can strongly influence the accuracy of biomass estimates. A
165 better approach may therefore be to estimate a tree’s aboveground biomass directly from
166 crown architectural properties which can be measured from airborne imagery, without the
167 need to first predict diameter. Specifically, both tree height (Hunter *et al.*, 2013; Chave *et al.*,
168 2014) and crown dimensions (Henry *et al.*, 2010; Goodman *et al.*, 2014; Ploton *et al.*, 2016)
169 are known to relate strongly to aboveground biomass, although it remains to be tested
170 whether they can be used to accurately estimate biomass without needing to also account for
171 stem diameter.

172 Here we compile a global dataset consisting of 108753 trees for which stem diameter, height
173 and crown diameter have all been measured, including 2395 trees which have been harvested
174 to measure aboveground biomass. The dataset is representative of the world's major tree-
175 dominated biomes and spans a huge gradient in tree size (Fig. 2). We use these data to
176 develop allometric equations that enable the precise and unbiased estimation of a tree's
177 diameter and aboveground biomass based on its height and horizontal crown dimensions, and
178 use the following questions to guide our processes: (i) Can a tree's diameter be estimated
179 accurately based on its height alone, or do we also need to account for its crown dimensions?
180 (ii) Can a single universal equation be used to model diameter, or do different scaling
181 relationships among forest types, biogeographic regions and tree functional types need to be
182 accommodated for? (iii) Can a tree's aboveground biomass be estimated directly from its
183 height and crown diameter, thereby eliminating the need to first predict its diameter?

184 **Materials and methods**

185 ALLOMETRIC DATABASE

186 We compiled a global database of trees for which stem diameter (D , in cm), height (H , in m)
187 and crown diameter (CD , in m) were all measured. Trees were selected for inclusion in the
188 database based on the following criteria: (i) only trees with $D \geq 1$ cm and $H \geq 1.3$ m were
189 considered; (ii) trees from managed plantations and agroforestry systems were excluded; (iii)
190 trees known or presumed to be severely damaged were removed (e.g., broken stems or major
191 branches; see Fig. S1); (iv) only trees whose geographic location was recorded were retained;
192 and (v) from a taxonomic perspective trees had to, at a minimum, be identifiable as either
193 angiosperms or gymnosperms (note that tree ferns and palms were excluded from the
194 analysis). Our search yielded a total of 108753 trees which met the above requirements. For
195 2395 of these, total oven-dry aboveground biomass (AGB , in kg) was additionally measured
196 by harvesting and weighing trees. The database spans a large range of tree sizes (D : 1.0–
197 293.0 cm; H : 1.3–72.5 m; CD : 0.1–41.0 m; AGB : 0.1–76063.5 kg), captures a wide spectrum
198 of tree forms and functional types (1492 tree species from 127 families), and covers the major
199 forest types and climatic conditions found in the world's forests (see Fig. 2 for an overview
200 of the database). A full list of data sources and associated measurement protocols is provided
201 in Appendix S1 of Supporting Information. The database is publicly available through
202 figshare (<https://dx.doi.org/10.6084/m9.figshare.3413539.v1>), with data from the Alberta
203 Permanent Sample Plots (<https://www.agric.gov.ab.ca/app21/forestrypage>) and the
204 International Cooperative Programme on Air Pollution Effects on Forests ([http://icp-](http://icp-forests.net/page/data-requests)
205 [forests.net/page/data-requests](http://icp-forests.net/page/data-requests)) archived separately and available upon request through the
206 above links.

207 *Forest biome classification*

208 Scaling relationships between D , H and CD are strongly influenced by climate (Lines *et al.*,
209 2012; Hulshof *et al.*, 2015), as well as varying among species (Poorter *et al.*, 2006) and
210 geographic regions (Banin *et al.*, 2012). To capture this degree of variation – which we
211 expect to be of key importance to accurately estimating both D and AGB – each tree in the
212 database was assigned to one of five biome types based on its geographic location: boreal
213 forests, temperate coniferous forests, temperate mixed forests, woodlands and savannas
214 (which combines temperate and tropical savannas, as well as Mediterranean woodlands) or
215 tropical and subtropical forests (biome classification follows Olson *et al.*, 2001). In the same
216 way, trees were also assigned to one of six biogeographic regions: Australasia, Afrotropics,
217 Nearctic, Indo-Malaya, Neotropics or Palearctic. Transitions among forest biomes reflect
218 strong climatic gradients (Whittaker 1975; Stephenson 1998; Fig. 2b), whereas biogeographic
219 realms define regions which share a common evolutionary history (Udvardy, 1975). Olson *et*
220 *al.*'s (2001) map of the world's terrestrial ecoregions, which defines the geographic
221 distribution of the world's major biome and biogeographic regions, is available for download
222 from <http://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world>.

223 APPROACH 1: ESTIMATING DIAMETER

224 *Model development*

225 To determine how to most accurately estimate a tree's diameter based on its crown
226 architectural properties, we compared a set of regression models in which D was expressed as
227 a function of either H , CD or the compound variable $H \times CD$ (which tests whether both
228 height and crown size are needed to predict D). We chose to model the combined effect of H
229 and CD using a compound variable (as opposed to including the two predictors separately in

230 the model) to avoid issues with collinearity resulting from the non-independence of H and
231 CD (Dormann *et al.*, 2013). Furthermore, preliminary analyses revealed that $H \times CD$ was as
232 good (if not better) a predictor of D than a model with H and CD as separate explanatory
233 variables (Table S2).

234 Typically, allometric equations are derived by fitting a linear regression directly to raw data
235 (which in most cases have been log-transformed). Yet this approach will tend to
236 underestimate the slope of a bivariate line when the independent variable is measured with
237 error (also known as regression dilution bias; Fuller, 1987; Warton *et al.*, 2006). In the case
238 of forest inventory data this systematic bias is made worse by the inherently unbalanced size
239 distribution of trees, as small stems – which vastly outnumber large ones – come to dominate
240 the signal of the regression (Duncanson *et al.*, 2015). As a solution to this problem,
241 Duncanson *et al.* (2015) proposed fitting allometric models to binned data as opposed to raw
242 values. Because this method reduces tree-level variation in allometric attributes to a mean
243 value, it has the drawback of inevitably underestimating the true uncertainty of the model.
244 However, a preliminary analysis of the data revealed it to be the only approach able to
245 adequately capture underlying allometric scaling relationships (see Appendix S2 for a
246 detailed discussion). As a compromise, we therefore chose to adopt Duncanson *et al.*'s (2015)
247 binning method to estimate allometric relationships, but also develop a framework for
248 robustly quantifying and propagating model uncertainty when working with binned data (see
249 “*Model uncertainty and error propagation*” section below).

250 We calculated the mean H , CD and $H \times CD$ for each of 50 stem diameter logarithmic bins of
251 constant width (logarithmic binning was chosen to better capture the right-skewed

252 distribution of D). Linear log-log models were then fit to the binned data using least-squares
253 regression (as implemented in the R statistical software; R Core Development Team, 2013):

$$\ln(D) = \alpha + \beta \ln(H) + \varepsilon \quad (1)$$

$$\ln(D) = \alpha + \beta \ln(CD) + \varepsilon \quad (2)$$

$$\ln(D) = \alpha + \beta \ln(H \times CD) + \varepsilon \quad (3)$$

254 where α and β are parameters to be estimated from the data and ε is an error term [which is
255 assumed to be normally distributed, with a mean of zero and a standard deviation σ ,
256 $N(0, \sigma^2)$].

257 Models 1–3 can be thought of as global allometric equations, as they assume that scaling
258 relationships between D , H and CD are invariant across forest types, biogeographic regions
259 and tree functional groups (e.g., angiosperms and gymnosperms). To determine the extent to
260 which regional or group-specific allometries improve the accuracy of D estimates compared
261 to those of a global model, we used mixed-effects models to develop two further equations.
262 First, the relationship between D and the independent variable (e.g., $H \times CD$) was allowed to
263 vary among forest types nested within biogeographic regions (i.e., random intercept and slope
264 model, where forest type and biogeographic region were treated as nested random effects). In
265 the second model, the relationship between D and the independent variable was further
266 allowed to vary among angiosperm and gymnosperm trees (i.e., separate α and β estimates
267 were calculated for each functional group/forest type/biogeographic region combination).
268 Note that in order to fit these models, the data binning processes was repeated and separate
269 mean values of H , CD and $H \times CD$ were calculated for each combination of functional group,
270 forest type and biogeographic region.

271 *Generating predictions*

272 Allometric models, such as those described above, can be used to estimate D for any tree
273 whose H and CD are known. Using Model 3 as an example, predicted diameter values (D_{pred})
274 are obtained as follows: $D_{pred} = \exp[\alpha + \beta \ln(H \times CD) + \varepsilon]$. Assuming ε is normally
275 distributed [i.e., $N(0, \sigma^2)$], the mean of $\exp(\varepsilon)$ can be approximated by $\exp(\sigma^2/2)$, where
276 σ^2 is the mean square error of the regression (Baskerville, 1972). An unbiased estimate of D
277 can therefore be calculated using the following equation:

$$D_{pred} = \exp[\alpha + \beta \ln(H \times CD)] \times \exp\left[\frac{\sigma^2}{2}\right] \quad (4)$$

278 *Model validation*

279 To evaluate and compare the predictive accuracy of the different D models, we: (i) divided
280 the database into a training set (90% of the data) and a validation set (remaining 10% of the
281 data, used exclusively to evaluate model performance). Trees assigned to the validation
282 dataset were selected following a size-stratified random sampling approach which aimed to
283 capture the full range of D in the database; (ii) D models were fit to the training dataset using
284 the binning approach described above; (iii) fitted equations were used to predict D for all
285 trees in the validation dataset [as outlined in equation (4)]; and (iv) the predictive error of
286 each model was quantified by comparing predicted and observed D values (D_{pred} and D_{obs} ,
287 respectively) of trees in the validation dataset (see below for a description of the model-
288 performance metrics used). Steps (i–iv) were repeated 100 times to avoid the randomization
289 procedure in step (i) having an undue effect on the model evaluation process.

290 For each D model we calculated two measures of average error: the root mean square error
291 (RMSE, in cm) and the relative systematic error (or bias, in %).

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (D_{obs} - D_{pred})^2}$$

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N \left(\frac{D_{obs} - D_{pred}}{D_{obs}} \right) \times 100$$

292 Additionally, a third model performance statistic was used to compare the predictive
 293 accuracy of the D models across functional groups (angiosperms and gymnosperms), forest
 294 types and biogeographic regions. Following the approach of Chave *et al.* (2014), we
 295 calculated the tree-level coefficient of variation (CV) in D for trees of functional group i ,
 296 growing in forest type j and in biogeographic region k as follows:

$$\text{CV}_{ijk} = \frac{\text{RMSE}_{ijk}}{\frac{1}{N} \sum_{i=1}^N D_{obs_{ijk}}}$$

297 where RMSE_{ijk} is the RMSE of trees belonging to functional group i , growing in forest type j
 298 and in biogeographic region k , whereas the denominator corresponds to the mean observed D
 299 for this same group of trees. Standardizing the RMSE by the mean D is a necessary step in
 300 order to compare model errors across functional groups, forest types or biogeographic
 301 regions, as errors in D are strongly dependent on tree size (Colgan *et al.*, 2013).

302 *Model uncertainty and error propagation*

303 As discussed previously, while data binning is well suited to estimating average allometric
 304 scaling relationships, it inevitably underestimates the true variability in these relationships
 305 among individual trees. Specifically, the data binning approach will tend to underestimate σ –
 306 the residual standard deviation – which makes quantifying and propagating uncertainty a

307 challenge. In a linear modelling framework $\sigma = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}$, where n is the number of
 308 observations, y_i is the i th observation of the response variable, and \hat{y}_i is the corresponding
 309 predicted value obtained from the model. The reason why data binning generally
 310 underestimates σ is that the difference between observed and predicted values (i.e., the
 311 residuals, $y_i - \hat{y}_i$) is calculated not for individual trees, but for mean values obtained by
 312 averaging across multiple trees. However, by using an independent dataset (the 10% of trees
 313 set aside for model validation), we can compare predicted and observed estimates of D
 314 generated for individual trees to get a much more realistic estimate of the true value of σ for a
 315 given model (which we refer to as σ_v):

$$\sigma_v = \sqrt{\frac{\sum (\ln(D_{obs}) - \ln(D_{pred}))^2}{n - 2}}$$

316 Using this simple approach we were able to generate realistic estimates of the predictive
 317 uncertainty of models fit using the data binning method (see Fig. S3). To enable users to
 318 robustly propagate uncertainty when using the equations developed here, we report σ_v values
 319 for all fitted models. Furthermore, in Appendix S5 we provide R code for replicating the
 320 entire analysis.

321 *Scaling-up from diameter to aboveground biomass*

322 Approach 1 aims to predict D from crown attributes, with the idea that D estimates can then
 323 be fed into existing biomass equations. To quantify the extent to which replacing field-
 324 measured D values with predicted ones influences the accuracy of *AGB* estimates, we used
 325 Chave *et al.*'s (2014) general biomass equation as a baseline. In Chave *et al.* (2014) *AGB* is
 326 expressed as the following function of D , H and wood density [ρ , in g cm^{-3}]; which we

327 obtained from the global wood density database of Chave *et al.* (2009) and Zanne *et al.*,
328 (2009)]: $AGB = 0.0673 \times (D^2 \times H \times \rho)^{0.976} \times \exp\left[0.357^2/2\right]$. Using this equation, we
329 estimated *AGB* for trees in the database with a known biomass (i.e., trees that had been
330 destructively harvested and weighed) using both field-measured and predicted *D* values as
331 inputs to the biomass model. Only trees with $D \geq 5$ cm were used for this purpose ($n = 1859$
332 trees with field-measured *AGB*), as trees smaller than this threshold contribute negligibly to
333 forest carbon stocks and were not used to calibrate Chave *et al.*'s (2014) equation. By
334 comparing observed *AGB* values with those predicted using Chave *et al.*'s (2014) equation,
335 we were then able to determine whether the underlying *D* models described previously can be
336 used to generate accurate biomass estimates. Additionally, this also allowed us to compare
337 the predictive accuracy of Approaches 1 and 2 – the latter of which aims to estimate *AGB*
338 directly from *H* and *CD* (see following section).

339 APPROACH 2: ESTIMATING ABOVEGROUND BIOMASS

340 Instead of estimating *D* first, a better approach to predicting the biomass of individual trees
341 from crown architectural attributes might be to relate *AGB* directly to *H* and *CD*. To test this,
342 we used data for trees with measured *AGB* to explore a number of alternative models relating
343 *AGB* to *H* and/or *CD*. Preliminary analyses revealed the compound variable $H \times CD$ to be a
344 far superior predictor of *AGB* than either *H* or *CD* alone. We therefore focus on the following
345 log-log regression model of *AGB*:

$$\ln(AGB) = \alpha + \beta \ln(H \times CD) + \varepsilon \quad (5)$$

346 Model development and validation followed the same steps described for Approach 1. As for
347 previous equations, the model was fit to binned mean values of $H \times CD$ (as opposed to raw

348 data). To allow a comparison with Approach 1, only trees with $D \geq 5$ cm were used to
349 develop the model. We further tested whether modelling angiosperms ($n = 1069$) and
350 gymnosperms ($n = 790$) separately would improve model accuracy, as these two functional
351 groups differ strongly in crown architecture (Poorter *et al.*, 2012; Hulshof *et al.*, 2015) as
352 well as wood density (Chave *et al.*, 2009). Given the relatively small number of trees with
353 measured *AGB* values, we did not explore the extent to which the relationship between *AGB*
354 and $H \times CD$ varies among forest types or biogeographic regions. The predictive accuracy of
355 equation (5) was compared against that of *AGB* models which include D as a predictor (i.e.,
356 Approach 1) on the basis of RMSE and bias.

357 Results

358 APPROACH 1: ESTIMATING DIAMETER

359 Of the candidate models we tested for estimating D , ones relying on H or CD alone as
360 predictors of D proved unsuitable. Despite exhibiting relatively low RMSE (13.7 cm), a
361 height-only model tended to systematically overestimate D (bias = 24.7%). This occurred
362 because D – H relationships were non-linear on a log-log scale, as H tended to asymptote in
363 large trees. As a result, a power-law tended to overestimate D for small and medium-sized
364 trees, while severely underestimating that of large ones (Fig. S4). Conversely, a model with
365 only CD as a predictor of D had higher RMSE (16.6 cm), but showed lower overall
366 systematic bias (-4.5%). However, the average bias masks a tendency of the crown diameter-
367 only model to overestimate D for large trees, while underpredicting the size of smaller stems
368 (Fig. S4). In contrast to the previous two models, $H \times CD$ proved a much better predictor of
369 D (Fig. 3). The best-fit global D model was:

$$D_{pred} = 0.557 \times (H \times CD)^{0.809} \times \exp\left[0.056^2/2\right] \quad (6)$$

370 Equation (6) had both lower RMSE (9.7 cm) and average systematic bias (-1.2%) compared
371 to models based on H or CD alone. Importantly, the model showed no evidence of over- or
372 underpredicting D across a wide range of tree sizes (Fig. 3b). Using the independent
373 validation dataset, we estimated σ_v [i.e., the standard deviation of $\ln(D_{obs}) - \ln(D_{pred})$] of the
374 model to be 0.45.

375 While the global D model presented in equation (6) was able to produce unbiased estimates
376 of D across a wide range of species, climate zones and tree sizes (Fig. 3), scaling
377 relationships between D and $H \times CD$ did vary among both forest types and functional groups

378 (Fig. 4). Incorporating these differences in the modelling processes further improved the
 379 precision of D estimates (Fig. 5 and Table S2). In particular, accounting for the different
 380 scaling relationships of angiosperms and gymnosperms reduced the RMSE of the model to
 381 8.1 cm, the average CV to 35.8% (from 43.3% in the global D model), and σ_v to 0.35 (Table
 382 S2). These gains in precision were especially evident when attempting to predict D for
 383 angiosperm trees in boreal and temperate coniferous forests, which tend to be dominated by
 384 gymnosperms (Fig. 5b). A full list of group-, forest type- and region-specific D equations is
 385 provided in Appendix S4.

386 APPROACH 2: ESTIMATING ABOVEGROUND BIOMASS

387 AGB was strongly related to $H \times CD$, with a linear log-log relationship holding across more
 388 than six orders of magnitude variation in tree mass (Fig. 6). Scaling relationships between
 389 AGB and $H \times CD$ varied consistently among functional groups, with gymnosperms
 390 exhibiting higher scaling constants ($\alpha = 0.109$ vs 0.016) but smaller scaling exponents ($\beta =$
 391 1.790 vs 2.013) compared to angiosperm trees (Fig. 6). The best-fit AGB model which
 392 accounted for different scaling relationships among angiosperms and gymnosperms was:

$$AGB_{pred} = (0.016 + \alpha_G) \times (H \times CD)^{(2.013 + \beta_G)} \times \exp\left[0.204^2/2\right] \quad (7)$$

393 where α_G and β_G are functional-group dependent parameters which represent the difference in
 394 the scaling constant α and scaling exponent β between angiosperm and gymnosperm trees.
 395 For gymnosperms $\alpha_G = 0.093$ and $\beta_G = -0.223$, whereas for angiosperms both parameters are
 396 set to zero. The estimated σ_v of the model was 0.69.

398 *AGB* estimates obtained using Chave *et al.*'s (2014) biomass equation and field-measured *D*
399 values as inputs showed a close agreement with observed *AGB* values (RMSE = 0.86 Mg;
400 Fig. 7a), but had a tendency to overestimate *AGB* (bias = 27.7%). As expected, replacing
401 field-measured *D* values with ones predicted using the global *D* model [i.e., equation (6),
402 corresponding to Approach 1] increased the RMSE of the model predictions to 1.78 Mg (Fig.
403 7b). However, the average systematic bias in the *AGB* predictions was little affected (bias =
404 30.1%, the overestimation arising from the use of the biomass function, not the global *D*
405 model). This suggests that diameter estimates obtained using the global *D* model can be
406 scaled up to biomass without introducing a systematic bias. In contrast to Approach 1, using
407 equation (7) to estimate *AGB* directly from $H \times CD$ (i.e., Approach 2) resulted in
408 substantially lower average bias in *AGB* estimates, regardless of tree mass (bias = -4.3%; Fig.
409 7c). Furthermore, Approach 2 had the advantage of reducing the RMSE of the model
410 predictions to 1.70 Mg.

411 Discussion

412 We developed general allometric models for estimating both the stem diameter and
413 aboveground biomass of trees based on crown architectural properties which can be remotely
414 sensed: tree height and crown diameter. Here we discuss how these allometric models can be
415 used to integrate remote sensing imagery – particularly ALS data – into forest monitoring
416 programs, allowing carbon stocks to be mapped with accuracy across forest landscapes and
417 shedding light on the processes which govern the structure and dynamics of forest
418 ecosystems.

419 STEM DIAMETER ALLOMETRIES FOR REMOTE SENSING IMAGERY

420 We found that estimating stem diameter required accounting for both height and crown size –
421 the latter of which proved essential for differentiating between trees of similar height but
422 having substantially different trunk sizes (King, 2005; King & Clark, 2011). Using a simple
423 metric which combines these two allometric dimensions – $H \times CD$ – we were able to derive a
424 global equation for estimating stem diameter which proved robust across a large range of tree
425 sizes, forest types and tree species (Fig. 3). Our results highlight how allocation to height
426 growth and lateral crown expansion are strongly coordinated in trees (Sterck & Bongers,
427 2001; King, 2005; Iida *et al.*, 2012), and illustrate how these developmental constraints can
428 be exploited for the purposes of estimating stem diameter.

429 While we did find that a single allometric function can be used to estimate diameter without
430 introducing systematic bias, incorporating different scaling relationships among forest types,
431 biogeographic regions and functional groups into the models helped improve the predictive
432 accuracy of the allometric equations (Figs 4 and 5; Table S2). Particularly important in this
433 respect was accounting for differences between angiosperms and gymnosperms (Fig. 5b).

434 This is not surprising given the contrasting crown architecture of these two groups:
435 gymnosperms generally exhibit strong apical dominance and invest heavily in height growth,
436 whereas angiosperm trees have a greater ability to plastically adapt the shape and size of their
437 crown to suit their competitive environment (Poorter *et al.*, 2012; Hulshof *et al.*, 2015). These
438 differences in crown architecture – coupled with clearly distinct leaf biochemical profiles –
439 also mean that angiosperm and gymnosperm trees can be easily distinguished using a variety
440 of remote sensing products (e.g., aerial photographs, hyperspectral sensors and ALS;
441 Dalponte *et al.* 2012). Consequently, we suggest that users select group-specific diameter
442 equations (which we provide in Appendix S4) wherever possible, as these can be employed
443 with little or no need for additional field data. As our ability to remotely map tree species
444 improves (e.g., through the development of spectral libraries derived from hyperspectral
445 sensors; Asner, 2013), it is conceivable that species-specific diameter equations could also be
446 utilized in the future. Similarly, other aspects known to influence crown architecture (e.g.,
447 tree packing density; Jucker *et al.*, 2015) could also be incorporated to further refine the
448 models we develop here.

449 The diameter allometries we develop here open the door to a more general and robust
450 framework for monitoring forest carbon stocks using ALS. Currently, the standard approach
451 for estimating carbon stocks from ALS data involves calculating summary statistics from
452 ALS point clouds for a given pixel of land (e.g., top canopy height) and relating these to
453 carbon estimates obtained from permanent field plots in a regression framework (Asner &
454 Mascaro, 2014; Asner *et al.*, 2014). Despite recent attempts to generalize this “area based”
455 approach (e.g., Asner & Mascaro 2014), most models for estimating carbon stocks from ALS
456 summary statistics are highly site-specific and can only be applied with confidence to the
457 particular patch of forest they were calibrated for. Working at tree-level provides an intuitive

458 solution to the issue of developing a general approach for mapping forest carbon stocks, and
459 would allow a direct comparison to field-based aboveground carbon estimates. This “tree-
460 centric” approach is not without its limitations, the biggest of which is the implicit
461 assumption that individual trees can be reliably identified and measured from ALS point
462 clouds (something which can be challenging in dense, multi-layered canopies). However,
463 recent years have seen substantial progress in this respect, as both ALS instruments and the
464 algorithms used to delineate trees from ALS data have improved considerably (Popescu *et*
465 *al.*, 2003; Yao *et al.*, 2012; Duncanson *et al.*, 2014; Paris *et al.*, 2016; Shendryk *et al.*, 2016).
466 For example, Paris *et al.* (2016) recently developed a segmentation method which was able to
467 correctly delineate the crowns of 97% and 77% of canopy dominant and understorey trees,
468 respectively, as well as accurately measuring the crown dimensions of all segmented trees.
469 Equally promising is Shendryk *et al.*'s (2016) algorithm which segments trees from the
470 bottom up (mimicking the approach used to process terrestrial laser scanning data; Calders *et*
471 *al.* 2014). As ALS technology continues to improve, “tree-centric” carbon monitoring
472 programs are becoming not only feasible, but oftentimes preferable to traditional “area
473 based” approaches (Duncanson *et al.*, 2015; Dalponte & Coomes, 2016).

474 In addition to mapping carbon stocks, characterising the relationships between stem diameter
475 and crown dimensions also has important implications for advancing our understanding of
476 forest dynamics. The most obvious application of the diameter allometries developed here is
477 for characterizing tree size distributions from airborne imagery, something which has proved
478 challenging using traditional “area-based” approaches (Maltamo & Gobakken, 2014). Tree
479 size distributions are an emergent property of forest ecosystems – arising from demographic
480 processes and competition for space among individual trees (Enquist *et al.*, 2009; Kohyama
481 *et al.*, 2015) – and are of key interest for understanding forest dynamics, structure and

482 responses to disturbance (Coomes *et al.*, 2003; Enquist *et al.*, 2009). Intriguingly, recent work
483 suggests that scaling relationships between diameter and crown size govern how trees utilize
484 canopy space and compete for light, thereby having a direct influence on tree size
485 distributions (Taubert *et al.*, 2015; Fariior *et al.*, 2016). ALS data, coupled with allometric
486 equations for converting crown dimensions to diameter distributions, would allow us to
487 empirically test this theory across large spatial scales and diverse forest types. In a similar
488 vein, diameter allometries provide a simple solution for integrating ALS data into individual-
489 based models of forest dynamics (e.g., Shugart *et al.* 2015), allowing these models to be more
490 easily parameterized and validated.

491 ESTIMATING ABOVEGROUND BIOMASS FROM CROWN DIMENSIONS

492 Using the subset of trees that were destructively harvested and weighed, we showed that *AGB*
493 was strongly related to tree height and crown size (Fig. 6). These results give weight to recent
494 reports which have highlighted how accounting for crown size can substantially improve
495 *AGB* estimation, especially in the case of large trees where a considerable proportion of the
496 biomass is stored in large branches (Henry *et al.*, 2010; Goodman *et al.*, 2014; Ploton *et al.*,
497 2016). The strong link between crown dimensions and *AGB* has important implications for
498 “tree-centric” carbon mapping approaches, as it suggests that *AGB* can be estimated directly
499 from remotely-sensed measurements of tree height and crown width without needing to first
500 predict diameter (Fig. 7c). This is particularly appealing as it reduces the number of steps in
501 the *AGB* estimation process (each of which carries a certain degree of error), and also
502 eliminates the need to select an equation from the literature for scaling from diameter to *AGB*.
503 Our analysis revealed clear differences in the *AGB* scaling relationships of angiosperms and
504 gymnosperms (Fig. 6), presumably reflecting differences in both crown architecture and

505 wood density among these two groups (Chave *et al.*, 2009; Poorter *et al.*, 2012; Hulshof *et*
506 *al.*, 2015). It may well be that *AGB* scaling relationships also vary systematically among
507 forest types or biogeographic regions, and that accounting for these differences could further
508 improve the predictive accuracy of the biomass allometries presented here. Unfortunately, the
509 relatively modest sample size of trees with measured *AGB* at our disposal meant we were
510 unable to robustly test these assumptions. Despite recent efforts to compile comprehensive
511 allometric databases (e.g., Chave *et al.* 2014; Falster *et al.* 2015), the number of trees with
512 measured *AGB* remains relatively small, geographically biased and heavily skewed towards
513 smaller stems. This is even more so when attempting to find trees that have been felled and
514 weighed and whose crown dimensions have also been recorded. Future studies developing
515 *AGB* equations should take care to also record the crown dimensions of harvested trees (e.g.,
516 Henry *et al.*, 2010; Goodman *et al.*, 2014; Ploton *et al.*, 2016). In this regard, perhaps the
517 most promising solution for bolstering existing allometric databases is terrestrial laser
518 scanning, which captures tree architecture in exquisite detail and provides a non-destructive
519 method for accurately estimating *AGB* (Calders *et al.*, 2015). Most importantly, this would
520 provide access to biomass data for large trees (e.g., ≥ 10 Mg), which tend to be
521 disproportionately rare in allometric databases – including the one we have assembled here
522 (only 2.4% of measured trees had a mass ≥ 10 Mg; see Fig. 2c).

523 SEEING THE FOREST AND THE TREES

524 Accurate assessments of forest carbon stocks are essential for initiatives to mitigate climate
525 change – such as the UN’s programme for Reducing Emissions from Deforestation and
526 Forest Degradation (REDD+) – to be implemented successfully (Agrawal *et al.*, 2011). Yet
527 monitoring carbon stocks across large and sometimes remote areas of forest poses a real

528 challenge, particularly in countries where national-scale forest inventory programs are not in
529 place. In this context, remote sensing technologies such as ALS promise to revolutionize the
530 way we census forests (Asner *et al.*, 2014). It is our hope that the allometric equations
531 developed here can help us move towards a more general and robust approach for monitoring
532 forests from the air.

533 **Acknowledgements**

534 We thank the co-authors, field data collectors and funding agencies of the original data
535 sources used in this study. We are also grateful to all curators of open-access databases from
536 which we drew data for this study. In particular, we wish to thank Daniel Falster and Remko
537 Duursma for compiling the biomass and allometry database (BAAD) for woody plants;
538 Michael Keller and Maiza Nara for providing us access to data from the Sustainable
539 Landscapes Brazil project; Kristina Anderson-Teixeira and her co-authors for archiving
540 allometric data from the CTFS-ForestGEO forest dynamics plot at the Smithsonian
541 Conservation Biology Institute (Virginia, USA); and to KaDonna Randolph of the USDA
542 forest service for her assistance in accessing the Forest Health Monitoring (FHM) database.
543 We thank Bruno Hérault and an anonymous reviewer for their thoughtful and constructive
544 comments on an earlier draft of our manuscript. T.J. was funded by NERC (grant number:
545 NE/K016377/1). This work has benefited from ANR grants to J.C. (CEBA, ref. ANR-10-
546 LABX-25-01 and TULIP, ref. ANR-10-LABX-0041). The Sustainable Landscapes Brazil
547 project was supported by the Brazilian Agricultural Research Corporation (EMBRAPA), the
548 US Forest Service, and USAID, and the US Department of State. Data collection for the
549 UNECE ICP Forests PCC Collaborative Database was co-financed by national or regional
550 organisations and by the European Commission under regulations (EEC) No 2158/86, Forest
551 Focus (EC) No 2152/200, FutMon (EC) LIFE07 ENV/D/218.

552 **References**

- 553 Agrawal A, Nepstad D, Chhatre A (2011) Reducing emissions from deforestation and forest
554 degradation. *Annual Review of Environment and Resources*, **36**, 373 – 396.
- 555 Anderson-Teixeira KJ, Davies SJ, Bennett AC et al. (2015) CTFS-ForestGEO: a worldwide
556 network monitoring forests in an era of global change. *Global Change Biology*, **21**, 528–
557 549.
- 558 Asner GP (2013) Biological diversity mapping comes of age. *Remote Sensing*, **5**, 374–376.
- 559 Asner GP, Mascaro J (2014) Mapping tropical forest carbon: calibrating plot estimates to a
560 simple LiDAR metric. *Remote Sensing of Environment*, **140**, 614–624.
- 561 Asner GP, Knapp DE, Martin RE et al. (2014) Targeted carbon conservation at national
562 scales with high-resolution monitoring. *Proceedings of the National Academy of
563 Sciences*, **111**, E5016–E5022.
- 564 Avitabile V, Herold M, Heuvelink GBM et al. (2016) An integrated pan-tropical biomass
565 map using multiple reference datasets. *Global Change Biology*, **22**, 1406–1420.
- 566 Baccini A, Goetz SJ, Walker WS et al. (2012) Estimated carbon dioxide emissions from
567 tropical deforestation improved by carbon-density maps. *Nature Climate Change*, **2**,
568 182–185.
- 569 Banin L, Feldpausch TR, Phillips OL et al. (2012) What controls tropical forest architecture?
570 Testing environmental, structural and floristic drivers. *Global Ecology and
571 Biogeography*, **21**, 1179–1190.
- 572 Baskerville GL (1972) Use of logarithmic regression in the estimation of plant biomass.
573 *Canadian Journal of Forest Research*, **2**, 49–53.
- 574 Bastin J-F, Barbier N, Réjou-Méchain M et al. (2015) Seeing Central African forests through
575 their largest trees. *Scientific Reports*, **5**, 13156.
- 576 Beer C, Reichstein M, Tomelleri E et al. (2010) Terrestrial gross carbon dioxide uptake:
577 global distribution and covariation with climate. *Science*, **329**, 834–838.
- 578 Calders K, Newnham G, Burt A et al. (2015) Nondestructive estimates of above-ground
579 biomass using terrestrial laser scanning. *Methods in Ecology and Evolution*, **6**, 198–208.
- 580 Chave J, Coomes DA, Jansen S, Lewis SL, Swenson NG, Zanne AE (2009) Towards a
581 worldwide wood economics spectrum. *Ecology Letters*, **12**, 351–366.
- 582 Chave J, Réjou-Méchain M, Búrquez A et al. (2014) Improved allometric models to estimate
583 the aboveground biomass of tropical trees. *Global Change Biology*, **20**, 3177–3190.
- 584 Chojnacky DC, Heath LS, Jenkins JC (2014) Updated generalized biomass equations for
585 North American tree species. *Forestry*, **87**, 129–151.
- 586 Colgan MS, Asner GP, Swemmer T (2013) Harvesting tree biomass at the stand level to

- 587 assess the accuracy of field and airborne biomass estimation in savannas. *Ecological*
588 *Applications*, **23**, 1170–1184.
- 589 Coomes DA, Duncan RP, Allen RB, Truscott J (2003) Disturbances prevent stem size-
590 density distributions in natural forests from following scaling relationships. *Ecology*
591 *Letters*, **6**, 980–989.
- 592 Dalponte M, Coomes DA (2016) Tree-centric mapping of forest carbon density from airborne
593 laser scanning and hyperspectral data. *Methods in Ecology and Evolution*.
- 594 Dalponte M, Bruzzone L, Gianelle D (2012) Tree species classification in the Southern Alps
595 based on the fusion of very high geometrical resolution multispectral/hyperspectral
596 images and LiDAR data. *Remote Sensing of Environment*, **123**, 258–270.
- 597 Dormann CF, Elith J, Bacher S et al. (2013) Collinearity: A review of methods to deal with it
598 and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.
- 599 Duncanson LI, Cook BD, Hurtt GC, Dubayah RO (2014) An efficient, multi-layered crown
600 delineation algorithm for mapping individual tree structure across multiple ecosystems.
601 *Remote Sensing of Environment*, **154**, 378–386.
- 602 Duncanson L, Rourke O, Dubayah R (2015) Small sample sizes yield biased allometric
603 equations in temperate forests. *Scientific Reports*, **5**, 17153.
- 604 Enquist BJ, West GB, Brown JH (2009) Extensions and evaluations of a general quantitative
605 theory of forest structure and dynamics. *Proceedings of the National Academy of*
606 *Sciences*, **106**, 7046–7051.
- 607 Falster DS, Duursma RA, Ishihara MI et al. (2015) BAAD: a biomass and allometry database
608 for woody plants. *Ecology*, **96**, 1445.
- 609 Farrior CE, Bohlman SA, Hubbell S, Pacala SW (2016) Dominance of the suppressed:
610 power-law size structure in tropical forests. *Science*, **351**, 155–157.
- 611 Feldpausch TR, Lloyd J, Lewis SL et al. (2012) Tree height integrated into pantropical forest
612 biomass estimates. *Biogeosciences*, **9**, 3381–3403.
- 613 Fuller WA (1987) *Measurement error models* (ed Fuller WA). John Wiley, New York, 440
614 pp.
- 615 Goodman RC, Phillips OL, Baker TR (2014) The importance of crown dimensions to
616 improve tropical tree biomass estimates. *Ecological Applications*, **24**, 680–689.
- 617 Hemery GE, Savill PS, Pryor SN (2005) Applications of the crown diameter–stem diameter
618 relationship for different species of broadleaved trees. *Forest Ecology and Management*,
619 **215**, 285–294.
- 620 Henry M, Besnard A, Asante WAA et al. (2010) Wood density, phytomass variations within
621 and among trees, and allometric equations in a tropical rainforest of Africa. *Forest*
622 *Ecology and Management*, **260**, 1375–1388.
- 623 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution

- 624 interpolated climate surfaces for global land areas. *International Journal of Climatology*,
625 **25**, 1965–1978.
- 626 Hulshof CM, Swenson NG, Weiser MD (2015) Tree height-diameter allometry across the
627 United States. *Ecology and Evolution*, **5**, 1193–1204.
- 628 Hunter MO, Keller M, Victoria D, Morton DC (2013) Tree height and tropical forest biomass
629 estimation. *Biogeosciences*, **10**, 8385–8399.
- 630 Iida Y, Poorter L, Sterck FJ, Kassim AR, Kubo T, Potts MD, Kohyama TS (2012) Wood
631 density explains architectural differentiation across 145 co-occurring tropical tree
632 species. *Functional Ecology*, **26**, 274–282.
- 633 Jucker T, Bouriaud O, Coomes DA (2015) Crown plasticity enables trees to optimize canopy
634 packing in mixed-species forests. *Functional Ecology*, **29**, 1078–1086.
- 635 King DA (2005) Linking tree form, allocation and growth with an allometrically explicit
636 model. *Ecological Modelling*, **185**, 77–91.
- 637 King DA, Clark DA (2011) Allometry of emergent tree species from saplings to above-
638 canopy adults in a Costa Rican rain forest. *Journal of Tropical Ecology*, **27**, 573–579.
- 639 Kohyama TS, Potts MD, Kohyama TI, Kassim AR, Ashton PS (2015) Demographic
640 properties shape tree size distribution in a Malaysian rain forest. *The American*
641 *Naturalist*, **185**, 367–379.
- 642 Lines ER, Zavala MA, Purves DW, Coomes DA (2012) Predictable changes in aboveground
643 allometry of trees along gradients of temperature, aridity and competition. *Global*
644 *Ecology and Biogeography*, **21**, 1017–1028.
- 645 Malhi Y, Wood D, Baker TR et al. (2006) The regional variation of aboveground live
646 biomass in old-growth Amazonian forests. *Global Change Biology*, **12**, 1107–1138.
- 647 Maltamo M, Gobakken T (2014) Predicting tree diameter distributions. In: *Forestry*
648 *applications of airborne laser scanning* (eds Maltamo M, Næsset E, Vauhkonen J), pp.
649 177–191. Springer, Berlin.
- 650 Muller-Landau HC, Condit RS, Chave J et al. (2006) Testing metabolic ecology theory for
651 allometric scaling of tree size, growth and mortality in tropical forests. *Ecology Letters*,
652 **9**, 575–588.
- 653 Olson DM, Dinerstein E, Wikramanayake ED et al. (2001) Terrestrial ecoregions of the
654 world: a new map of life on earth. *BioScience*, **51**, 933–938.
- 655 Pan Y, Birdsey RA, Fang J et al. (2011) A large and persistent carbon sink in the world's
656 forests. *Science*, **333**, 988–93.
- 657 Paris C, Valduga D, Bruzzone L (2016) A hierarchical approach to three-dimensional
658 segmentation of LiDAR data at single-tree level in a multilayered forest. *IEEE*
659 *Transactions on Geoscience and Remote Sensing*, 1–14.
- 660 Ploton P, Barbier N, Momo ST et al. (2016) Closing a gap in tropical forest biomass

- 661 estimation: accounting for crown mass variation in pantropical allometries.
662 *Biogeosciences*, **13**, 1571–1585.
- 663 Poorter L, Bongers L, Bongers F (2006) Architecture of 54 moist-forest tree species: traits,
664 trade-offs, and functional groups. *Ecology*, **87**, 1289–1301.
- 665 Poorter L, Lianes E, Moreno-de las Heras M, Zavala MA (2012) Architecture of Iberian
666 canopy tree species in relation to wood density, shade tolerance and climate. *Plant*
667 *Ecology*, **213**, 707–722.
- 668 Popescu SC, Wynne RH, Nelson RF (2003) Measuring individual tree crown diameter with
669 lidar and assessing its influence on estimating forest volume and biomass. *Canadian*
670 *Journal of Remote Sensing*, **29**, 564–577.
- 671 R Core Development Team (2013) *R: A language and environment for statistical computing*.
672 R Foundation for Statistical Computing, Vienna, Austria.
- 673 Saatchi SS, Harris NL, Brown S et al. (2011) Benchmark map of forest carbon stocks in
674 tropical regions across three continents. *Proceedings of the National Academy of*
675 *Sciences*, **108**, 9899–9904.
- 676 Shendryk I, Broich M, Tulbure MG, Alexandrov S V (2016) Bottom-up delineation of
677 individual trees from full-waveform airborne laser scans in a structurally complex
678 eucalypt forest. *Remote Sensing of Environment*, **173**, 69–83.
- 679 Shugart HH, Asner GP, Fischer R, Huth A, Knapp N, Le Toan T, Shuman JK (2015)
680 Computer and remote-sensing infrastructure to enhance large-scale testing of individual-
681 based forest models. *Frontiers in Ecology and the Environment*, **13**, 503–511.
- 682 Slik JWF, Paoli G, Mcguire K et al. (2013) Large trees drive forest aboveground biomass
683 variation in moist lowland forests across the tropics. *Global Ecology and Biogeography*,
684 **22**, 1261–1271.
- 685 Stephenson NL (1998) Actual evapotranspiration and deficit: biologically meaningful
686 correlates of vegetation distribution across spatial scales. *Journal of Biogeography*, **25**,
687 855–870.
- 688 Sterck FJ, Bongers F (2001) Crown development in tropical rain forest trees: patterns with
689 tree height and light availability. *Journal of Ecology*, **89**, 1–13.
- 690 Taubert F, Jahn MW, Dobner H-J, Wiegand T, Huth A (2015) The structure of tropical
691 forests and sphere packings. *Proceedings of the National Academy of Sciences*,
692 201513417.
- 693 Udvardy MD (1975) *A classification of the biogeographical provinces of the world*. IUCN
694 Occasional Paper No. 18, Morges, Switzerland.
- 695 Warton DI, Wright IJ, Falster DS, Westoby M (2006) Bivariate line-fitting methods for
696 allometry. *Biological Reviews*, **81**, 259–91.
- 697 West GB, Brown JH, Enquist BJ (1999) A general model for the structure and allometry of

- 698 plant vascular systems. *Nature*, **400**, 122–126.
- 699 Whittaker RJ (1975) *Communities and ecosystems*, 2nd edn. Macmillan, New York.
- 700 Yao W, Krzystek P, Heurich M (2012) Tree species classification and estimation of stem
701 volume and DBH based on single tree extraction by exploiting airborne full-waveform
702 LiDAR data. *Remote Sensing of Environment*, **123**, 368–380.
- 703 Zanne AE, Lopez-Gonzalez G, Coomes DA et al. (2009) Global wood density database.
704 *Dryad Digital Repository*, <http://dx.doi.org/10.5061/dryad.234>.
- 705 Zianis D, Muukkonen P, Mäkipää R, Mencuccini M (2005) Biomass and stem volume
706 equations for tree species in Europe. *Silva Fennica Monographs*, **4**, 1–63.
- 707

708 **Figure legends**

709 **Fig. 1:** Schematic diagram illustrating how airborne laser scanning (ALS) imagery can be
710 integrated into forest inventory programs. State-of-the-art algorithms that detect and measure
711 individual tree crowns from ALS point clouds are combined with existing field data to
712 estimate the diameter and aboveground biomass of remotely sensed trees.

713 **Fig. 2:** Overview of the allometric database. Panel (a) shows the geographic coverage of the
714 database in relation to the world's biomes (map adapted from Olson *et al.*, 2001). Circle size
715 reflects the number of trees measured at each location (on a logarithmic scale). Panel (b)
716 highlights differences in mean annual precipitation and temperature among forest types.
717 Climate data were obtained from the WorldClim database (Hijmans *et al.*, 2005), which
718 consists of gridded annual mean values covering the period between 1950-2000 (data
719 available from: <http://www.worldclim.org/current>). In (c) violin plots show the size
720 distribution – in terms of diameter and aboveground biomass – of trees in the database. The
721 number of records available for each forest type is displayed on the right.

722 **Fig. 3:** Goodness-of-fit for the global diameter model [i.e., equation (6) in the main text],
723 tested on an independent random sample of the data corresponding to 10% of measured trees
724 ($n = 10875$). Panel (a) compares predicted and observed diameter values, with the dashed line
725 corresponding to a 1:1 relationship. The density of overlapping points is represented by a
726 colour gradient which ranges from blue (low point density) to red (high point density). Panel
727 (b) reports the mean relative error (i.e., $\frac{D_{pred} - D_{obs}}{D_{obs}} \times 100$) for different diameter size classes,
728 with the bars delimiting the interquartile range (thick lines) and 95% limits (thin lines) of the
729 errors.

730 **Fig. 4:** Relationship between stem diameter and the product of tree height and crown
731 diameter ($H \times CD$). Panel (a) shows the distribution – on a logarithmic scale – of the raw
732 data (in grey) and of the mean $H \times CD$ values in each diameter size class (black circles).
733 Panel (b) illustrates fitted relationships between diameter and $H \times CD$ for each forest type
734 separately, while (c) reports the slopes of these relationships ($\pm 95\%$ confidence intervals) for
735 angiosperms and gymnosperms separately.

736 **Fig. 5:** Comparison of model performance between the global diameter model [i.e., equation
737 (6) in the main text] and (a) a model that allows scaling relationships to vary among forest
738 types and biogeographic regions, and (b) one where angiosperms and gymnosperms are also
739 modelled separately. The coefficient of variation (CV) of the absolute errors ($\pm 95\%$ range
740 across 100 simulations) is reported for angiosperms (open symbols) and gymnosperms
741 (closed symbols) according to forest type and biogeographic region. Boxplots along each axis
742 capture the distribution of the model errors, while the dashed line indicates a 1:1 relationship.

743 **Fig. 6:** Relationship between aboveground biomass and the product of tree height and crown
744 diameter. Gymnosperm (filled circles; $n = 1049$) and angiosperm trees (empty circles; $n =$
745 1346) are shown separately. For illustrative purposes, 536 trees with a stem diameter of less
746 than 5 cm are also shown.

747 **Fig. 7:** Aboveground biomass (AGB) estimation accuracy. Panels (a–c) show predicted
748 *versus* observed AGB values for trees greater than 5 cm in diameter ($n = 1859$). In panel (a),
749 AGB was estimated using Chave *et al.*'s (2014) equation (where AGB is expressed as a
750 function of diameter, height and wood density). Panel (b) illustrates the predictive accuracy
751 of Chave *et al.*'s (2014) equation when field-measured diameters are replaced with ones
752 predicted using the global diameter model (i.e., Approach 1). Panel (c) corresponds to a

753 model in which AGB is expressed directly as a function of tree height and crown diameter
754 (i.e., Approach 2). For panels (**a–c**), the dashed line corresponds to a 1:1 relationship, while
755 the solid line is a regression spline fit to the data points to highlight how predictive accuracy
756 varies with tree size. The RMSE and bias of each set of predictions is reported in the lower
757 right-hand corner. Panel (**d**) shows the probability density distribution of the absolute errors
758 (i.e., $AGB_{pred} - AGB_{obs}$) for each AGB function.

759 **Supporting information**

760 Additional supporting information may be found in the online version of this article:

761 **Appendix S1:** Database generation

762 **Appendix S2:** Data binning

763 **Appendix S3:** Diameter model comparison

764 **Appendix S4:** Region-, forest type- and group-specific diameter equations

765 **Appendix S5:** R code for implementing data binning approach