

APPLICATION

BIOMASS: an R package for estimating above-ground biomass and its uncertainty in tropical forests

Maxime Réjou-Méchain^{*,1,2} , Ariane Tanguy¹, Camille Piponiot³, Jérôme Chave⁴ and Bruno Hérault⁵ 

¹French Institute of Pondicherry, UMIFRE 21/USR 3330 CNRS-MAEE, Pondicherry, India; ²UMR AMAP, IRD, F-34000 Montpellier, France; ³Université de la Guyane, UMR 'Ecologie des Forêts de Guyane' (AgroparisTech, Cirad, CNRS, Inra, Université des Antilles), Kourou Cedex F-97379, French Guiana; ⁴Laboratoire Evolution et Diversité Biologique UMR 5174, CNRS, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France; and ⁵Cirad, UMR 'Ecologie des Forêts de Guyane' (AgroparisTech, CNRS, Inra, Université de Guyane, Université des Antilles), Kourou Cedex F-97379, French Guiana

Summary

1. Estimating forest above-ground biomass (AGB), or carbon (AGC), in tropical forests has become a major concern for scientists and stakeholders. However, AGB assessment procedures are not fully standardized and even more importantly, the uncertainty associated with AGB estimates is seldom assessed.
2. Here, we present an R package designed to compute both AGB/AGC estimate and its associated uncertainty from forest plot datasets, using a Bayesian inference procedure. The package builds upon previous work on pantropical and regional biomass allometric equations and published datasets by default, but it can also integrate unpublished or complementary datasets in many steps.
3. BIOMASS performs a number of standard tasks on input forest tree inventories: (i) tree species identification, if available, is automatically corrected; (ii) wood density is estimated from tree species identity; (iii) if height data are available, a local height–diameter allometry may be built; else height is inferred from pantropical or regional models; (iv) finally, AGB/AGC are estimated by propagating the errors associated with all the calculation steps up to the final estimate. R code is given in the paper and in the appendix for the purpose of illustration.
4. The BIOMASS package should contribute to improved standards for AGB calculation for tropical forest stands, and will encourage users to report the uncertainties associated with stand-level AGB/AGC estimates in future studies.

Key-words: allometry, error propagation, forest carbon, height–diameter relationship

Introduction

The increase in carbon dioxide in the atmosphere and its effect on climate has brought forest carbon accounting to the forefront of the research and political agenda. Much attention is currently being given to tropical forests as they contain c. 55% of the carbon stored in the terrestrial vegetation, and tropical deforestation and degradation account for more than 10% of global anthropogenic greenhouse gas emissions (Le Quere *et al.* 2013). However, large uncertainties are still associated with these carbon stock estimates.

Forest carbon stocks, whether estimated through remote sensing or pure field-based approaches, are most often derived from forest inventories at least at one step. However, the way these forest inventories are converted into carbon stock estimates currently varies among studies. This results in variation among estimates that are not attributable solely to natural

variability, but also to methodological considerations, the focus of the present contribution.

Estimating above-ground biomass (AGB) accurately includes several steps, such as the assignment of a wood density value to a tree, the choice of a biomass allometric model or the choice of a height–diameter (H–D) model when tree height data are incomplete or absent. Errors associated with all these steps are rarely accounted for in the final AGB estimate (see however Chave *et al.* 2004; Molto, Rossi & Blanc 2013; Chen, Laurin & Valentini 2015; Mermoz *et al.* 2015). For instance, a major source of error comes from the uncertainty in inferring height when it has not been measured directly and accurately. Different generic models have been proposed that depend either on regional variation (Feldpausch *et al.* 2012) or on bioclimatic constraints (Chave *et al.* 2014), although it is advisable to develop local H–D allometries because local abiotic or biotic (e.g. competition) conditions may significantly impact H–D relationships.

In this paper, we present an open source R package (R Core Team 2016) dedicated to the estimation of AGB and its

*Correspondence author. E-mail: maxime.rejou@gmail.com

uncertainty from tropical forest inventories. The goal of the BIOMASS package is to streamline many of the steps that are now being conducted by operators as illustrated in Fig. 1).

Test datasets

1 KarnatakaForest, which contains tree inventory data from 96 forest plots (1 ha) established in the central Western Ghats of India (Ramesh *et al.* 2010).

2 NouraguesHD, which contains tree inventory data, including height measurements, from two 1-ha plots established in the Nouragues Ecological Research Station, in French Guiana (see Réjou-Méchain *et al.* 2015 for methods).

```
install.packages("BIOMASS", dependencies=TRUE)
library(BIOMASS)
data(NouraguesHD)
```

Accounting for diameter

Tree diameter is a compulsory entry in our package. Note that diameter measurements are prone to errors and that a careful check prior to the analyses may correct some obvious errors.

Accounting for wood density

Wood density, the oven dry weight divided by green volume of wood, displays a 10-fold variation among tree species (Chave

et al. 2009) and thus constitutes an important determinant of tree biomass. Because wood density is almost never measured for individual trees in forest inventories, a wood density value is often assigned to each tree from an independent database based on taxonomic identity (Flores & Coomes 2011). The `getWoodDensity` function from the BIOMASS package assigns a wood density value to each taxon using the global wood density (GWD) database as a reference (Chave *et al.* 2009; Zanne *et al.* 2009). Additional wood density values can be added using the `addWoodDensityData` argument. By default, `getWoodDensity` assigns to each taxon a species- or genus-level average if at least one wood density value in the same genus as the focal taxon is available in the reference database. For unidentified trees or if the genus is missing in the reference database, the stand-level mean wood density is assigned to the tree (based on trees for which a value was attributed). In the inputs of the function, each tree may be assigned to a stand through the `stand` argument, i.e. a vector of factors, such as plots, habitats or successional status. The `family` option also assigns to the trees a family-level wood density average, but it should be borne in mind that the taxon-average approach gives relatively poor estimates above the genus level (Flores & Coomes 2011). The `region` option allows the user to only consider wood density values from a specific region in the GWD database. However, Flores & Coomes (2011) showed that more accurate wood density averages are obtained with the global GWD database (without subsetting GWD by region).

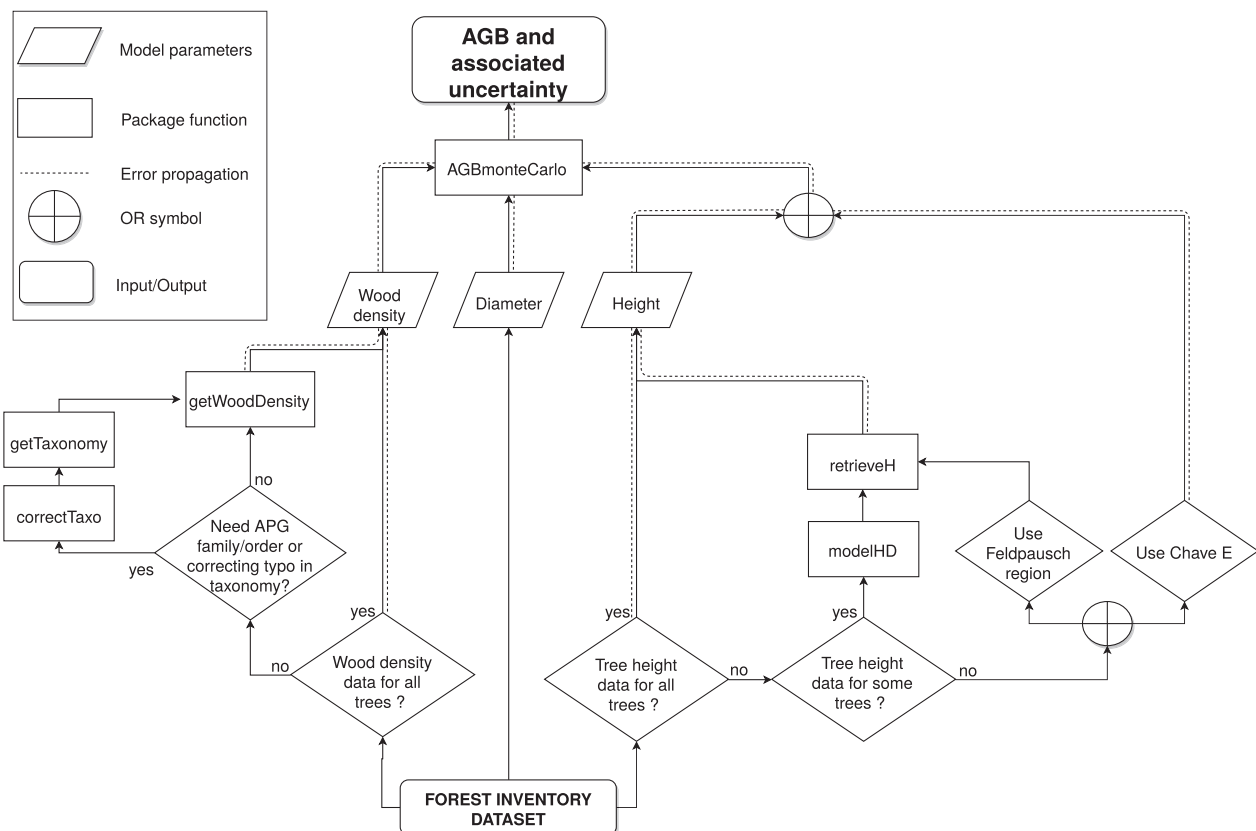


Fig. 1. Workflow of the BIOMASS package.

In our experience, taxonomic names in forest inventories are prone to typos. We implemented the function `correctTaxo` to correct typos in scientific names using the Taxonomic Name Resolution Service (TNRS; Boyle *et al.* 2013) via the Taxosaurus interface (<http://taxosaurus.org/>). This function has been adapted from the `tnrs` function of the `taxize` R package (Chamberlain & Szöcs 2013). Note that the taxonomy of the GWD database has been also submitted to TNRS to match the outputs from `correctTaxo`. The `getTaxonomy` function provides the APG III order and family names with genus names as entry.

```
Taxo <- correctTaxo(genus=NouraguesHD$genus,
species=NouraguesHD$species)
WDdata <- getWoodDensity(genus=Taxo
$genusCorrected, species=Taxo$speciesCorrected,
stand=NouraguesHD$plotId)
```

Accounting for tree height

Tree height plays a key role in AGB calculation. Using tree height measurements for all trees is thus by far the most accurate method to infer AGB (Chave *et al.* 2014). However, accurate tree height measurements are rare, or rarely exhaustive, in closed-canopy tropical forests. Hence, a recommended strategy is to construct a stand-specific H–D allometry using a subset of well-measured trees (Réjou-Méchain *et al.* 2015). Trees selected for height measurement should be ideally sampled equally across size classes. Several H–D models have been proposed in the literature [see for instance the `lmfor` R package (Mehtatalo 2017)]. In the BIOMASS package the `modelHd` function contains five models to fit height H–D relationships in the tropics: three log–log polynomial models with order varying from 1 to 3, a three-parameter Weibull model (Feldpausch *et al.* 2012) and a two-parameter Michaelis–Menten model (Molto *et al.* 2014). By default, `modelHd` fits all the models and outputs the associated statistics (RSE and bias) for model assessment.

```
Hmodel <- modelHd(D=NouraguesHD$D, H=NouraguesHD
$H, method="log2", useWeight=TRUE)
Hlocal <- retrieveH(D=NouraguesHD$D, model=
Hmodel)
```

If the study site misses tree height measurements, two alternatives may be used in the `retrieveH` function. First, the continent- or region-specific H–D models proposed by Feldpausch *et al.* (2012) may be used through the `region` argument. Second, a generic H–D model based on a single bioclimatic predictor *E* (eqn 6a in Chave *et al.* 2014) may be used through the `coord` argument.

```
Hfeld <- retrieveH(D=NouraguesHD$D, region=
"GuianaShield")
HchaveE <- retrieveH(D=NouraguesHD$D, coord=cbind
(NouraguesHD$long, NouraguesHD$lat))
```

AGB calculation

Once tree diameter, wood density and tree height have been retrieved for each tree, the generalized allometric model eqn 4 of Chave *et al.* (2014) can be used with the `computeAGB` function where AGB values are reported in Mg instead of in kg (Mg is the conventional unit for AGB values).

```
AGBtree <- computeAGB(D=NouraguesHD$D, WD=WDdata
$meanWD, H=Hlocal$H)
```

In the course of our package development, we identified an inaccurate statement in Chave *et al.* (2014). Contrary to a statement made in the paper, eqn 7 of that paper was not fitted directly with the data but was obtained by combining eqns 4 and 6. The directly fitted equation, including Baskerville (1972) correction, has the following expression:

$$\text{AGB} = \exp(-2.024 - 0.896 \cdot E + 0.920 \cdot \ln(\text{WD}) + 2.795 \cdot \ln(D) - 0.0461 \cdot [\ln(D)^2])$$

where the bioclimatic compound parameter *E* is the same as the one given in Chave *et al.* (2014). This is the equation implemented in the current version of `computeAGB` function when the site coordinates are given.

```
AGBtree <- computeAGB(D=NouraguesHD$D, WD=WDdata
$meanWD, coord=cbind(NouraguesHD$long,
NouraguesHD$lat))
```

Propagating errors

In the BIOMASS package, the `AGBmonteCarlo` function allows the user to propagate different sources of error up to the final AGB estimate, as explained below.

DIAMETER MEASUREMENT ERROR

In the `Dpropag` argument of the `AGBmonteCarlo` function, the user can set diameter measurement errors by providing either a residual standard error (RSE) or a vector of errors (SD values) associated with diameters. Random and independent errors are assigned to diameters at each iteration in the `AGBmonteCarlo` function using a truncated normal distribution with a range of diameter values of (0.1, 500) cm. By default, the error propagation assumes that no errors were made on diameter measurements. An example of diameter measurement error is given in the `Dpropag` argument, taken from the Barro Colorado Island permanent plot where two types of errors were identified: large and small errors on 5 and 95% of the trees respectively (Chave *et al.* 2004).

WOOD DENSITY ERROR

To account for uncertainty in wood density values, the user can provide either a RSE or a vector of errors (SD values) associated with wood density values through the `errWD` argument of the `AGBmonteCarlo` function. At each iteration, errors

are assigned randomly and independently to wood density values using a truncated normal distribution with a range of (0.08, 1.39) g cm⁻³, the minimum and maximum wood density values observed in the GWD database respectively. As a possibility, the `getWoodDensity` function outputs prior values on the uncertainty on wood density values using the mean SD at the species, genus and family levels calculated on taxa having at least 10 wood density values in the GWD database.

TREE HEIGHT ERROR

Here again, the user can provide either a RSE or a vector of errors (SD values) associated with tree height values through the `errH` argument of the `AGBmonteCarlo` function. Random errors are assigned independently to tree height values using a truncated normal distribution with a range of (1.3, $\text{maxH} + 15$) m, where maxH is the maximum tree height of the dataset. The RSE for the generic H–D models, or the RSE resulting from the local H–D models, are stored in the outputs of the `retrieveH` and `modelHD` function respectively. If real tree height measurements are available for all trees, the user should decide whether a height-related random error should be added in the error propagation.

ALLOMETRIC MODEL ERROR

To account for uncertainty in Chave *et al.* (2014) model parameters, eqns 4 and 7 have been inferred using a Markov chain Monte Carlo algorithm (code in Appendix S1, Supporting Information) in a Bayesian framework with uninformative priors. Chains of posterior vector, i.e. the model coefficients and RSE (RSEp), were stored to be used in the final error propagation step in the `AGBmonteCarlo` function (see below). To infer the models, we used the tree destructive dataset of Chave *et al.* (2014) available at http://chave.ups-tlse.fr/pantropical_allometry.htm.

FINAL PROPAGATION

Errors are propagated through a Monte Carlo scheme in the `AGBmonteCarlo` function, where the following steps are repeated n times:

- 1 When appropriate, diameter, wood density and height values are simulated adding a random error for each tree as described above.
- 2 Allometric model parameters are set up randomly picking up a single vector of posterior parameters applied simultaneously to all the trees of a stand (i.e. corresponding to a systematic error).
- 3 Tree AGB is computed using (1) and (2).
- 4 A single error term is added for all trees independently from the distribution $N(0, \text{RSEp})$.

The procedure yields a distribution of n tree AGB values (n , the number of iterations, is an argument of the `AGBmonteCarlo` function), which are then summed by stand to provide n stand AGB values from which summary statistics are reported (mean, median and 95% credibility interval).

```
AGBmc <- AGBmonteCarlo(D=NouraguesHD$D, WD=WDdata$meanWD, H=Hlocal$H,
  errWD=WDdata$sdWD, errH=Hlocal$RSE, Dpropag=
  "chave2004")
```

As an option in the `AGBmonteCarlo` function, the results can be provided in carbon instead of biomass (Carbon argument) using a mean biomass to carbon ratio of 0.471 and a standard deviation of 0.206 (calculated from carbon content of tropical angiosperms stems compiled by Thomas & Martin 2012a, b).

Figure 2 illustrates the outputs from the `AGBmonteCarlo` function for the two Nouragues plots using three different approaches. It shows that for those two plots, the mean AGB and its uncertainty were highly consistent between methods, indicating that both Chave' eqn 7 and Feldpausch regional H–D model perform well in that study area.

Focusing on a single plot from Nouragues, Fig. 3 shows how the different sources of errors contribute to the global error and how these relative contributions depend on the number of sampled trees. It confirms that the allometric error is the most important (Chave *et al.* 2004; Molto, Rossi & Blanc 2013), followed by the H–D model error. It also shows that the global error decreases rapidly with sampling size, because random errors are averaged. Note that this averaging effect depends on the tree size distribution of the stand and is thus site-specific.

Availability

The current stable version of the package requires R 3.3.0 and can be downloaded from CRAN (<https://cran.r-project.org/web/packages/BIOMASS/>) or installed directly in R with

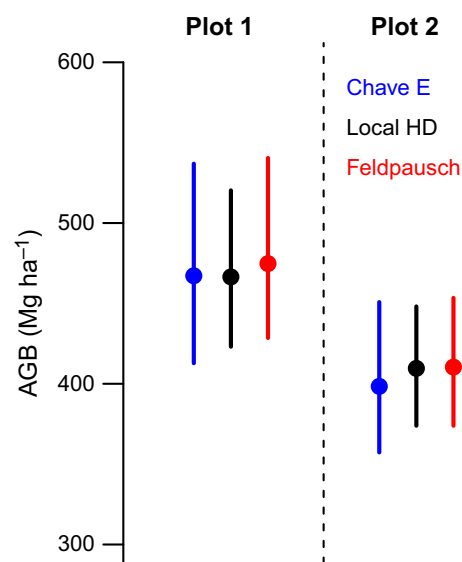


Fig. 2. Comparison of above-ground biomass estimation and uncertainty from three different methods in two 1-ha plots from French Guiana (Nouragues). Chave E: Chave *et al.* 2014 eqn 7 model; Local H–D: use of a mix of directly measured tree height (H) and of estimated ones by a local height–diameter (H–D) model; Feldpausch: use of the regional Guiana Shield Feldpausch H–D model. The R code used for generating this figure is available in Appendix S2.

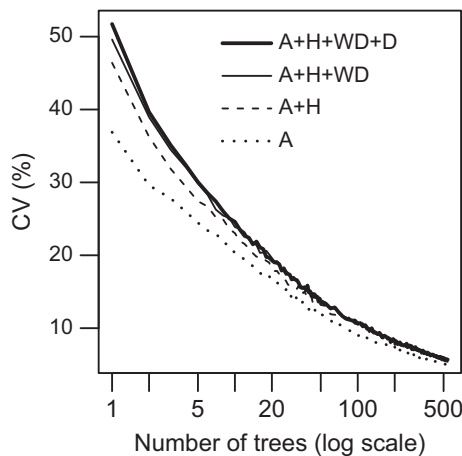


Fig. 3. Above-ground biomass uncertainty (coefficient of variation, CV) as a function of the number of sampled trees and using different sources of errors for a 1-ha plot from French Guiana (Nouragues). A, allometric error; H, height–diameter error (from the Guiana Shield Feldpausch model); WD, wood density error; D, diameter error (Chave *et al.* 2004). The R code used for generating this figure is available in Appendix S2.

`install.packages("BIOMASS")`. A vignette (i.e. a tutorial) is available for this package (<https://cran.r-project.org/web/packages/BIOMASS/vignettes/VignetteBiomass.html>). BIOMASS depends on five existing R packages: (i) *raster* (Hijmans 2016); (ii) *msm* (Jackson 2011); (iii) *minpack.lm* (Elzhov *et al.* 2016); (iv) *httr* (Wickham 2016); and (v) *jsonlite* (Ooms 2014).

Authors' contributions

M.R.M. and A.T. conceived and designed the package; B.H. and C.P. developed the Metropolis algorithm for the allometric models; all authors contributed to developing R functions; M.R.M. wrote the paper with inputs from the other authors.

Acknowledgements

We thank two anonymous reviewers for their valuable insights as well as some users for their feedbacks on the first version of the package. We acknowledge financial support from Investissement d'Avenir grants managed by Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX-25-01; TULIP: ANR-10-LABX-0041). Salaries of M.R.M. and A.T. were supported by the French Ministry of Foreign affairs. The authors have no conflicts of interest to declare.

Data accessibility

The wood carbon content data from which the mean and standard deviation is used in the package are available from <https://doi.org/10.5061/dryad.69sg2> (Thomas & Martin 2012b). The GWD dataset is available from <http://datadryad.org/handle/10255/dryad.235> (Zanne *et al.* 2009) and the Karnataka Forest dataset is available from <https://doi.org/10.6084/m9.figshare.c.3303531.v1>, both of which are included in the BIOMASS package as well.

References

- Baskerville, G.L. (1972) Use of logarithmic regression in the estimation of plant biomass. *Canadian Journal of Forest Research*, **2**, 49–53.
- Boyle, B., Hopkins, N., Lu, Z. *et al.* (2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics*, **14**, 1.

- Chamberlain, S.A. & Szöcs, E. (2013) Taxize: taxonomic search and retrieval in R. *F1000Research*, **2**, 191. doi: 10.12688/f1000research.2-191.v1.
- Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S. & Perez, R. (2004) Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 409–420.
- Chave, J., Coomes, D., Jansen, S., Lewis, S.L., Swenson, N.G. & Zanne, A.E. (2009) Towards a worldwide wood economics spectrum. *Ecology Letters*, **12**, 351–366.
- Chave, J., Réjou-Méchain, M., Búrquez, A. *et al.* (2014) Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, **20**, 3177–3190.
- Chen, Q., Laurin, G.V. & Valentini, R. (2015) Uncertainty of remotely sensed aboveground biomass over an African tropical forest: propagating errors from trees to plots to pixels. *Remote Sensing of Environment*, **160**, 134–143.
- Elzhov, T.V., Mullen, K.M., Spiess, A.-N. & Bolker, B. (2016) *Minpack.lm: R interface to the Levenberg–Marquardt nonlinear least-squares algorithm found in minpack, plus support for bounds*. Available at: <https://CRAN.R-project.org/package=minpack.lm> (accessed 22 September 2016).
- Feldpausch, T.R., Lloyd, J., Lewis, S.L. *et al.* (2012) Tree height integrated into pantropical forest biomass estimates. *Biogeosciences*, **9**, 3381–3403.
- Flores, O. & Coomes, D.A. (2011) Estimating the wood density of species for carbon stock assessments. *Methods in Ecology and Evolution*, **2**, 214–220.
- Hijmans, R.J. (2016) *Raster: Geographic data analysis and modeling*. Available at: <https://CRAN.R-project.org/package=raster> (accessed 22 September 2016).
- Jackson, C.H. (2011) Multi-state models for panel data: the *msm* package for R. *Journal of Statistical Software*, **38**, 1–29.
- Le Quere, C., Andres, R.J., Boden, T. *et al.* (2013) The global carbon budget 1959–2011. *Earth System Science Data Discussions*, **5**, 165–185.
- Mehtatalo, L. (2017) *lmfor: Functions for forest biometrics*. Available at: <https://CRAN.R-project.org/package=lmfor> (accessed 22 September 2016).
- Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V. & Gourlet-Fleury, S. (2015) Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sensing of Environment*, **15**, 307–317.
- Molto, Q., Rossi, V. & Blanc, L. (2013) Error propagation in biomass estimation in tropical forests. *Methods in Ecology and Evolution*, **4**, 175–183.
- Molto, Q., Hérault, B., Boreux, J.-J., Daullet, M., Rousteau, A. & Rossi, V. (2014) Predicting tree heights for biomass estimates in tropical forests—a test from French Guiana. *Biogeosciences*, **11**, 3121–3130.
- Ooms, J. (2014) *The jsonlite package: A practical and consistent mapping between json data and r objects*. arXiv:1403.2805 [stat.CO]. Available at: <http://arxiv.org/abs/1403.2805> (accessed 22 September 2016).
- R Core Team. (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/> (accessed 22 September 2016).
- Ramesh, B.R., Swaminath, M.H., Patil, S.V., Pélissier, R., Venugopal, P.D., Aravajy, S., Elouard, C. & Ramalingam, S. (2010) Forest stand structure and composition in 96 sites along environmental gradients in the central Western Ghats of India. *Ecology*, **91**, 3118.
- Réjou-Méchain, M., Tymen, B., Blanc, L., Fauset, S., Feldpausch, T.R., Monteagudo, A., Phillips, O.L., Richard, H. & Chave, J. (2015) Using repeated small-footprint LiDAR maps to infer spatial variation and dynamics of a high-biomass neotropical forest. *Remote Sensing of Environment*, **169**, 93–101.
- Thomas, S. & Martin, A. (2012a) Carbon content of tree tissues: a synthesis. *Forests*, **3**, 332–352.
- Thomas, S. & Martin, A. (2012b) Dryad wood carbon content database. *Dryad Digital Repository*, <https://dx.doi.org/10.5061/dryad.69sg2>
- Wickham, H. (2016) *Httr: Tools for working with urls and http*. <https://CRAN.R-project.org/package=httr>.
- Zanne, A.E., Lopez-Gonzalez, G., Coomes, D.A. *et al.* (2009) Global wood density database. *Dryad Digital Repository*, <http://datadryad.org/handle/10255/dryad.235>.

Received 4 January 2017; accepted 26 January 2017

Handling Editor: Sarah Goslee

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. Metropolis algorithm used to generate the posterior distribution of Chave *et al.* (2014)'s model parameters.

Appendix S2. R codes used to produce Figs 2 and 3.