RESEARCH PAPER

WILEY | **Global Ecology and Biogeography** | A Journal of Macroecology

Developing allometric models to predict the individual aboveground biomass of shrubs worldwide

G. Conti[1] [iD] | L. D. Gorné[1] | S. R. Zeballos[1] | M. L. Lipoma[1] | G. Gatica[2] |
E. Kowaljow[1] | J. I. Whitworth-Hulse[1,3] | A. Cuchietti[1,4] | M. Poca[1,3] | S. Pestoni[1] |
P. M. Fernandes[5] [iD]

[1]Instituto Multidisciplinario de Biología Vegetal, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional de Córdoba, Facultad de Ciencias Exactas Físicas y Naturales, IMBiV (CONICET-UNC), Córdoba, Argentina

[2]Departamento de Biología, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan, San Juan, Argentina

[3]Instituto de Matemática Aplicada San Luis, Universidad Nacional de San Luis, CONICET, Grupo de Estudios Ambientales, IMASL (CONICET-Universidad Nacional de San Luis), San Luis, Argentina

[4]Secretaría de Ambiente y Desarrollo Sustentable de Argentina (SAyDS), Buenos Aires, Argentina

[5]Centro de Investigação e de Tecnologias Agroambientais e Biológicas (CITAB), Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal

Correspondence
L. D. Gorné, Instituto Multidisciplinario de Biología Vegetal (IMBIV-CONICET), Av. Vélez Sársfield 1611, Edificio de Investigaciones Biológicas y Tecnológicas, Ciudad Universitaria, CP 5000, Córdoba, Argentina.
Email: gorneld@gmail.com

Funding information
European Regional Development Fund, Grant/Award Number: POCI-01-0145-FEDER-006958; Fondo para la Investigación Científica y Tecnológica, Grant/Award Number: PICT 2016-3184; Portuguese Foundation for Science and Technology, Grant/Award Number: UID/AGR/04033/2013

Editor: Andrew Kerkhoff

Abstract

Aim: Existing global models to predict standing biomass are based on trees characterized by a single principal stem, well developed in height. However, their use in open woodlands and shrublands, characterized by multistemmed species with substantial crown development, generates a high level of uncertainty in biomass estimates. This limitation led us to (a) develop global models of shrub individual aboveground biomass based on simple allometric variables, (b) to compare the fit of these models with existing global biomass models, and (c) to assess whether models fit change when bioclimatic variables are considered.

Location: Global.

Time period: Present.

Major taxa studied: 118 species of shrubs.

Methods: We compile a database of 3,243 individuals across 49 sites distributed worldwide. Including stem basal diameter, height and crown diameter as predictor variables, we built potential models and compared their fit using generalized least squares. We used mixed effects models to determine if bioclimatic variables improved the accuracy of biomass models.

Results: Although the most important variable in terms of predictive capacity was stem basal diameter, crown diameter significantly improved the models' fit, followed by height. Four models were finally chosen, with the best model combining all these variables in the same equation [R^2 = 0.930, root mean square error (RMSE) = 0.476]. Selected models performed as well as established global biomass models. Including the individual bioform significantly improved the models' fit.

Main conclusions: Stem basal diameter, crown diameter and height measures could be combined to provide robust aboveground biomass (AGB) estimates of individual shrub species. Our study supplements well-established models developed for trees, allowing more accurate biomass estimation of multistemmed woody individuals. We further provide tools for a methodological standardization of individual biomass quantification in these species. We expect these results contribute to improve the quality of biomass estimates across ecosystems, but also to generate methodological consensus on field biomass assessments in shrubs.

1 | INTRODUCTION

Over recent decades, vegetation cover shifts due to land use changes have deeply affected the global carbon budget, representing *c.* 18% of the total carbon emissions globally (Le Quéré et al., 2018). In this context, carbon mitigation projects intended to protect and enhance forest cover, and thus carbon sequestration, have gained attention [e.g., Reducing Emissions from Deforestation and Forest Degradation (REDD)]. Most of these projects have taken place in tropical ecosystems and have been focused on tree species, particularly those with larger diameters, as they store large amounts of carbon (Lutz et al., 2018; Saatchi et al., 2011; Slik et al., 2013). Nonetheless, open forests and woodlands have received less attention in carbon accounting models and mitigation projects. However, some of these ecosystems (e.g., semi-arid ecosystems, mostly dominated by multistemmed trees and shrub species) have been recognized as drivers of global atmospheric CO_2 level variations (Poulter et al., 2014), and are expected to increase in coverage due to global climate change by the end of the 21st century (Huang, Yu, Guan, Wang, & Guo, 2016). One of the large sources of uncertainty in carbon stocks quantification is the lack of standard models to convert woody individual measurements into biomass estimations (Chave et al., 2005). Therefore, it is essential to gather ground-based information on carbon stored in vegetation across different ecosystems, especially in those where multistemmed trees and shrubs are a significant proportion of the total vegetation biomass, in order to implement and improve global climate change mitigation projects.

Ground-based estimates of aboveground biomass (AGB) are typically obtained by applying allometric models based on field measurements of biometric data at individual level. At present, several well-established global models use stem diameter at breast height (i.e., DBH, diameter at 1.3 m from the ground level) as an AGB predictor variable of tree individuals, which are often represented by a single stem, well developed in height (Brown, 1997; Chave et al.,). The choice of this biometric variable is based on the fact that DBH is relatively easy to obtain in the field (but see Paul et al., 2017) and has been shown to have good predictive capacity. Nevertheless, in both woodlands and open forests, where vegetation communities are dominated by multistemmed and/or small woody individuals, these generalized biomass models based on DBH do not fit accurately. This is because these woody species typically have a well-developed crown, branching off at stem heights below 1.3 m (Vesk, Warton, & Westoby, 2004). For the purpose of this study, "shrubs" are all woody non-climbing plants with multiple stems and/or small size that do not meet the tree definition criteria (i.e., a perennial woody plant with many secondary branches supported by a single main stem or trunk with clear apical dominance; Richardson & Rejmánek, 2011). We use the term "shrub" in its broadest definition in order

to use one unifying denomination. Paul et al. (2017) demonstrated the methodological limitation of using DBH, showing that AGB models based upon diameter tend to be less accurate for multistemmed woody individuals, mainly due to the lack of standardization in accounting for the diameter of multistemmed individuals.

Plant dimensions reflecting crown morphology (i.e., height and crown diameters) are more comprehensive AGB predictors of shrub species, as they define plant architecture better than diameter. Inclusion of these allometric variables would improve the fit of models used to quantify shrub AGB, representing alternative measures when stem diameter is difficult or unpractical to measure (Conti, Enrico, Casanoves, & Díaz, 2013; Hierro, Branch, Villarreal, & Clark, 2000; Hofstad, 2005; Murray & Jacobson, 1982). The relevance of allometric tools including height and crown size to predict AGB has just recently started to be widely recognized, especially for trees (Feldpausch et al., 2012; Goodman, Phillips, & Baker, 2014; Jucker et al., 2017; Ploton et al., 2016). Wood density has also been documented as another important predictor of stand-level biomass in tropical ecosystems (Chave et al., 2014), although its predictive potential has been questioned, particularly for AGB models developed for temperate species for which wood density is typically less variable (Swenson & Enquist, 2007).

Improving the fit of allometric models represents one of the most important steps in assessing AGB stocks (Chave et al., 2005; Skole, Samek, & Smalligan, 2011). Developing new generalized models for wide application depends on the availability of destructive sampling data, which are enormously time-consuming and expensive to acquire (Chave et al., 2014; Paul et al., 2016). Species-specific models locally developed for shrub species across ecosystems provide accurate AGB estimates (e.g., Conti et al., 2013; Hierro et al., 2000; Návar et al., 2004). However, their application outside the range of calibration, or in a different location, can generate significant biases (20–200%, e.g., Chave et al., 2014; Ishihara et al., 2015; Ketterings, Coe, Noordwijk, Ambagau, & Palm, 2001). The compilation of large databases from previous species- and site-specific studies based on destructively sampled individuals could be effectively used to reparametrize global allometric models, substantially improving global AGB estimations (e.g., Chave et al., 2014; Jucker et al., 2017; Paul et al., 2016; Vieilledent et al., 2012). Additionally, to fully test the applicability of global predictive biomass models based on allometric variables, bioclimatic variables need to be considered, given the varying relationship between individual allometric variables (scaling relationships) across different regions and bioforms (Blanchard et al., 2016). Therefore, it is imperative to assemble all the available information to improve global biomass models.

We analysed a globally distributed database of direct-harvest biomass of multistemmed and/or small woody species. The dataset

included 49 undisturbed vegetation and secondary forest sites, spanning a wide range of vegetation types, for a total of 3,243 woody individuals belonging to 118 different shrub species. The following questions were addressed: (a) which are the best global AGB models developed for shrubs based on commonly used allometric variables?; (b) how do our AGB models compare in performance with other globally fitted AGB models?; and (c) does the inclusion of different bioclimatic variables improve the prediction of AGB for shrubs? To the best of our knowledge, this work is the first compilation of biomass data for shrub species across different ecosystem types around the world, and one of the first that attempts to make progress in overcoming issues related to modelling AGB in multistemmed and/or small woody individuals globally.

2 | MATERIALS AND METHODS

2.1 | Database development

We compiled a database of published papers that developed AGB models using allometric variables as predictor variables. These models were developed based on destructively harvested woody individual plants considered within the definition of shrub detailed before.

To construct the database, we performed a search in Scopus and Google Scholar during the years 2016 and 2017 using the following English and Spanish search terms "(*mass OR biomasa OR weight OR peso) AND (shrub* OR arbusto* OR multistem* OR multitallo OR multi-tallo) AND (allometr* OR alométrico* OR equation* OR ecuacion* OR 'dimensional relationship*' OR 'relaciones dimensionales')". We identified 390 papers and examined whether they included the following measurements at individual plant level in order to be included in the database: total AGB; crown diameters, their average, or crown area in order to estimate mean crown diameter (CD); individual height (H); and, if available, a measure of total stem diameter (basal diameter, measured below 30 cm stem height (BD); stem diameter at 30 cm height (D30) and/or DBH). We excluded data from individuals explicitly sampled after regrowth. In the case of multistemmed individuals, the measured stem diameter had to represent the total basal area at a specified stem height. In general, the most common stem diameters sampled in shrubs were basal diameter (below 30 cm stem height: D5, D10, D20 or not exactly defined), diameter at 30 cm, diameter of the longest stem and diameter at breast height. Some studies reported the sum of all stem diameters, while others did not report how multiple stems were dealt with in diameter estimation. This is why we decided to obtain a standardized basal stem diameter estimation (BDest) based on the different stem diameters compiled (see Supporting Information Figure S1) to maximize the number of included harvested individuals, but at the expense of increasing the model's biases. More accurate estimations of AGB are possible by precisely defining the basal stem height at the diameter at which it is sampled. This would certainly generate more precise predictions than those that were obtained here.

After the selection and filtering processes we finally retained 35 articles that were used to construct the database (see details in Supporting Information Table S1). When data were solely displayed in figures, we used a data extraction software (Tummers, 2006). When data were not publicly available, the dataset associated with the published reference was requested from the authors. More detailed information about the database compilation, analysis and screening are in Supporting Information Text S1.

Wood density (the oven-dry wood mass divided by its green volume, and denoted as ρ, g/cm^3) of the sampled species was also included. If ρ for a given species was reported in the original study, we included that value. Otherwise, we assumed the mean value for the species reported in the global wood density database (Chave et al., 2009), as well as in other references (see Supporting Information Table S2 for details).

To test if model parameters change under particular bioclimatic conditions, we included the species' bioform, as well as the corresponding biome and the global aridity index (GAI) category for the study site. We subcategorized shrub species' bioforms as mangroves, subshrubs, shrubs (i.e., small size woody individuals typically multistemmed) and "shrubs sometimes small trees" (SST) (i.e., medium sized woody plants, with variable architecture from multistemmed to single-stemmed) (Zizka, Govender, & Higgins, 2014). This categorization followed the authors' description of the species and the available information on local floras or digital repositories (see specific references in Supporting Information Text S1). Taxonomic information for each species was carefully checked for consistency using The Plant List (http://www.theplantlist.org/). Each individual shrub in the database was assigned to one of six biome types based on its geographic location: "Tropical & subtropical forests & shrublands"; "Temperate coniferous forests"; "Temperate mixed forests"; "Savannas, woodlands & Mediterranean forests"; "Grasslands & shrublands" or "Deserts & xeric shrublands" [classification adapted from Olson et al. (2001)]. The location of each study site was georeferenced and used to obtain environmental data. We chose the GAI as an index summarizing mean annual precipitation (MAP) and mean annual potential evapotranspiration (MAE) (GAI = MAP/MAE). We defined different regions according to the following GAI categories as follows: GAI < 0.03, Hyper Arid; GAI between 0.03 and 0.2, Arid; GAI between 0.2 and 0.5, Semi-Arid; GAI between 0.5 and 0.65, Dry sub-humid; and GAI > 0.65, Humid (Middleton & Thomas, 1992).

A first overview of the database showed that it included 49 study sites (Figure 1a), from published and unpublished sources, for a total of 3,243 harvested shrub individuals, belonging to 118 species and 35 families (see Supporting Information Table S2), spanning a wide range of woody shrub sizes (BD: 0.5–41.74 cm; H: 0.2–20.76 m; CD: 0.2–13.02 m; AGB: 0.01–926.30 kg, ρ: 0.38–1.07 g/cm^3). The database sites were distributed along different combinations of temperature and precipitation, but with a stronger representation under 1,000 mm MAP and semi-arid sites (GAI between 0.2 and 0.5) (Figure 1b,d). "Savannas, woodlands & Mediterranean forests" were the most represented biome type

(a)

(b)
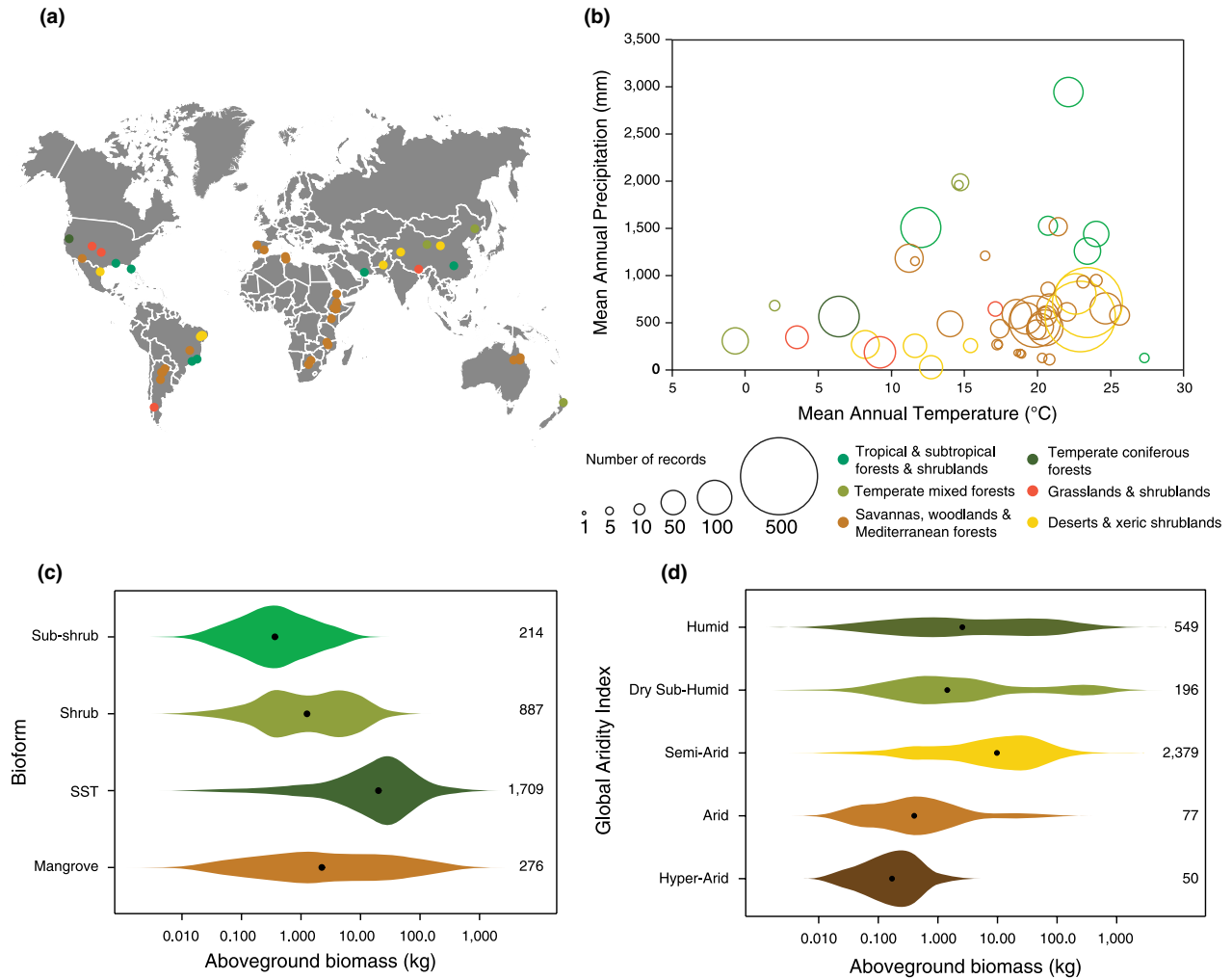
FIGURE 1 Overview of the allometric database. Panel (a) shows the geographic location of the included sites, and (b) the distribution of these sites considering their mean annual precipitation and temperature among forest types. Circle sizes reflect the number of individuals measured at each location. Panels (c) and (d) show violin plots of the distribution in terms of individual aboveground dry biomass across bioforms and the global aridity index categories, respectively. In the case of bioform, 157 individuals were excluded because their bioform was not reported in the source study. SST = shrubs sometimes small trees

(Figure 1b), while most individuals fell within the SST bioform category (Figure 1c).

2.2 | Statistical analysis

2.2.1 | Developing a global shrub species aboveground model

Allometric model construction was based on regressing AGB as a dependent variable against one or several independent variables. The independent variables considered were stem basal diameter (BD or BDest, cm), tree height (H, m), mean crown diameter (CD, m) and wood density (ρ). First, we fitted a log-log model relating AGB to each individual allometric variable separately to explore which was the best AGB predictor. The tested log-log models had the following mathematical form (in the simplest version):

$$AGB = \exp\left(\alpha + \beta_i \ln\left(X_i\right) + \varepsilon\right) \qquad (1)$$

where X refers to the putative independent variables, α and β are model parameters, and ε is an error term.

For practical use, estimated biomass predictions computed using a log-log model must be back-transformed to the original, plant-biomass scale. Because this transformation is nonlinear, and there is variability in the observed data around the fitted relationship, a simple exponential-based transformation (a "naïve" transformation) would generate bias (Baskerville, 1972; Clifford, Cressie, England, Roxburgh, & Paul, 2013). Consequently, correction factors are typically calculated to remove this bias when back-transforming. Clifford et al. (2013) reviewed this issue and provided routines to implement those corrections. We tested the naïve version together with the recommended correction factors [El-Shaarawi-Viveros estimator

(EV) and the minimum bias estimator (MB) from El-Shaarawi and Viveros (1997) and Shen and Zhu (2008), respectively] and the traditional correction factor (restricted maximum likelihood, REML) proposed by Baskerville (1972) across our dataset.

We built several models including all allometric predictor variables in all possible combinations and compared the fit of the models obtained using generalized least squares (GLS). To carry out these analyses the total database was randomly divided into a training dataset (n = 2,919, 90% of the data) and a fitting dataset (n = 324, 10% of the data). The fitting dataset was exclusively used to evaluate our models' performance and to compare with other published models. The distributions across both training and fitting datasets of the bioclimatic variables considered are shown in Supporting Information Figure S2. For the training dataset, we first cleaned all missing values for the included variables to generate a complete subset of data without blanks (n = 1,444), in order to obtain comparable statistical descriptors between the developed models, and to run model selection. The preferred statistical model was selected according to the Akaike information criterion (AIC), a likelihood criterion that penalizes the number of parameters (Burnham & Anderson, 2002); the coefficient of determination (R^2); and the root mean square error (RMSE), calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\text{AGB}_{\text{obs}} - \text{AGB}_{\text{est}} \right)^2} \qquad (2)$$

where AGB_{obs} is the observed AGB and AGB_{est} is the estimated AGB. We used the GLS procedure using the maximum likelihood (ML) method to obtain the model's AIC and then, using the REML method, we computed R^2 and RMSE.

After selecting the best statistical models, we reparametrized the models using the complete analysis dataset to include all data points available for each particular model, given that different allometric variables had different sample sizes. Then, for each final model we reported the number of individuals used to obtain their particular parameters.

2.2.2 | Testing established allometric models

We used the fitting dataset to compare the performance of the final selected models together with already existing and well-established global models in predicting the individual AGB. The already published global models selected were:

1. The one proposed by Chave et al. (2014) using DBH, H and ρ as predictive variables for pantropical trees,

$$\text{AGB}_{\text{est}} = 0.0673 \left(\rho \times \text{DBH}^2 \times \text{H} \right)^{0.976} \qquad (3)$$

2. A global tree model proposed by Jucker et al. (2017) for angiosperms, using H and CD as predictive variables,

$$\text{AGB}_{\text{est}} = 0.016 \left((\text{H} \times \text{CD})^{2.013} \right) e^{\frac{0.204^2}{2}}, \qquad (4)$$

3a. The shrubs and small trees model ("SHRUB" model)

$$\text{AGB}_{\text{est}} = e^{(2.428 \ln D10) - 3.007} \times 1.128 \qquad (5)$$

3b. The multistemmed tree model ("MULTI" model),

$$\text{AGB}_{\text{est}} = e^{(2.474 \ln D10) - 2.757} \times 1.0787 \qquad (6)$$

with both 3a and 3b developed by Paul et al. (2016) based only upon stem diameter at 10 cm height (D10) at the Australian continental scale.

As Chave et al.'s model required DBH as an input variable, the final training subset was reduced to 122 individuals containing all the variables required by all models, using the original DBH value recorded (without conversion). We compared the RMSE of each model (from the observed versus predicted AGB relationship) to have a general approximation of the behaviour of particular models. Furthermore, we calculated the average relative systematic error of model predictions (or bias, in %) using the biomass data without log-transformation (Chave et al., 2005; Jucker et al., 2017), as shown below:

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\text{AGB}_{\text{est}} - \text{AGB}_{\text{obs}}}{\text{AGB}_{\text{obs}}} \right) 100 \qquad (7)$$

All model comparisons were performed using their back-transformed versions to estimate AGB from the fitting dataset. In some cases, we then log-transformed data to facilitate graphical comparisons.

2.2.3 | Testing the effect of bioclimatic variables on models' performance

The models developed so far considered the complete dataset without accounting for climatic and ecological differences, assuming that scaling relationships between BD, H and CD are invariant across zones with varying hydrological balance, biomes, or distinctive bioforms. To determine if these variables improved the accuracy of the four proposed biomass models, we used mixed effects models to analyse if their parameters vary as a function of these variables. To that end, the relationship between AGB and the independent variables (e.g., BD, H, CD) for each case was allowed to vary among bioforms, biomes and GAI categories (random intercepts and slope model, only random intercept, and only random slope). We could not analyse if these variables also interact between one another (nested analysis), as not all bioforms were represented in all biomes and climatic conditions, resulting in a very unbalanced model. The final models were compared with each original model, to evaluate if the inclusion of bioclimatic variables improved the fit of the models.

All statistical analyses were performed using R statistical software version 3.4.0 (R Core Team, 2017). Climatic data extraction was carried out using the extract function from the "raster" package in R (Hijmans & van Etten, 2017). Model comparisons and mixed effect models were performed using the "nlme", "lme4", "bbmle" and "MuMIn" packages (Barton, 2016; Bates, Mächler, Bolker, & Walker, 2014; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017).

3 | RESULTS

3.1 | Developing a global shrub species aboveground model

Preliminary analyses using only single predictor variables (i.e., BD, H and CD) to explore which of them better explained AGB variability showed BD as the best single predictor (Figure 2a). CD also showed a good fit (Figure 2b), with greater predictive power than H, which showed a comparatively poorer fit when used as a single predictive variable of AGB, with a larger dispersion than BD and CD (Figure 2c).
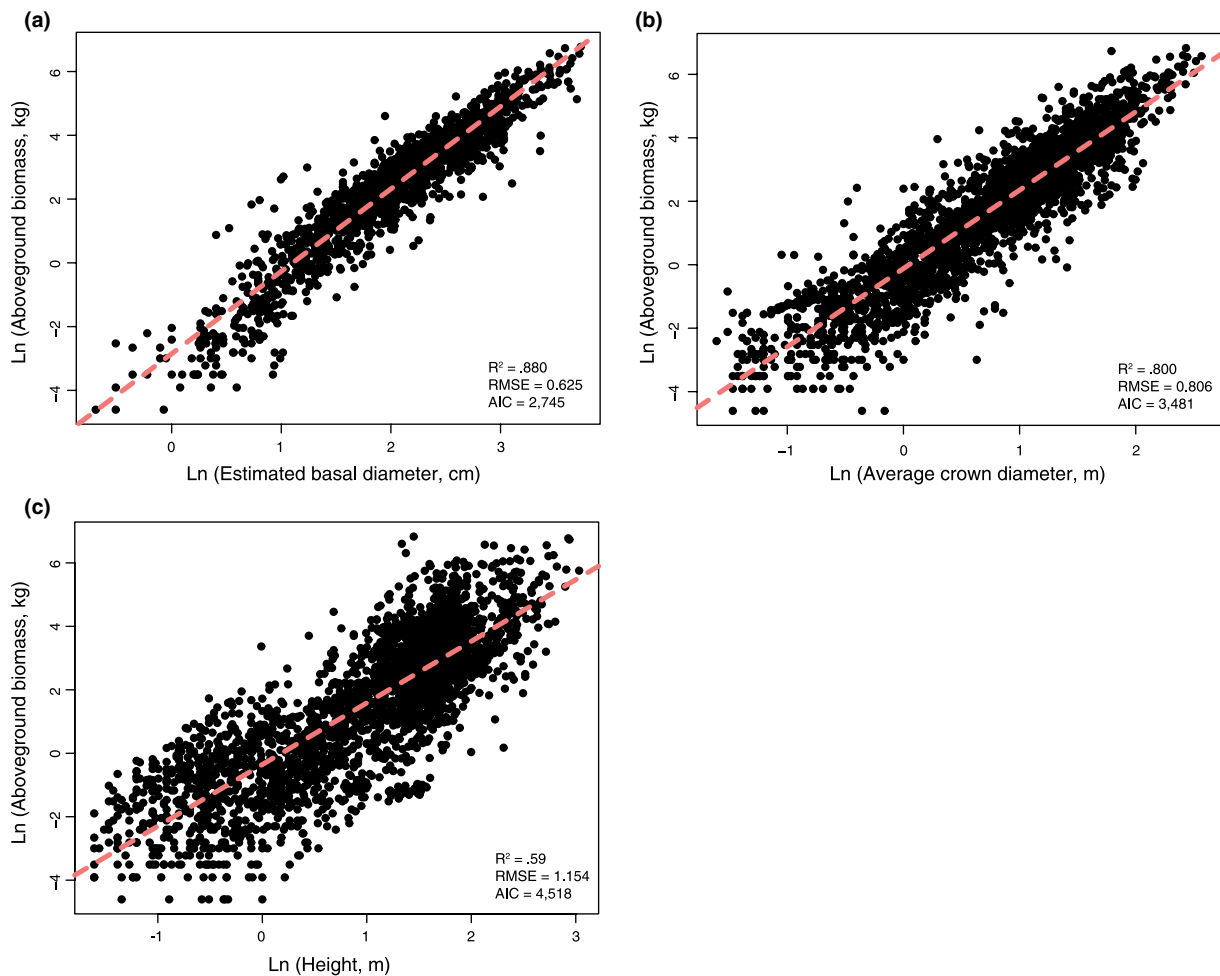
The analysis to find the best predictive model of AGB showed that, in general, models including BD had the highest predictive capacity. Even when BD was the only predictive variable in the model, goodness of fit was superior to any other model including all the remaining variables. However, the inclusion of other allometric variables (particularly CD, but also H) significantly improved the fit of

the models (Figure 3). When BD was not included as a predictive variable, the model based on CD and H had the best fit, but with significantly lower predictive capacity (Table 1). Finally, the inclusion of ρ slightly improved the fit of the proposed models (see statistical descriptors in Supporting Information Table S3).

We selected four final best-performing models for predicting shrub AGB (Table 1) according to the fit of the models and the number of variables included. The best AGB model included BD, CD and H as independent variables (Model 1):

$$AGB_{est} = e^{(-2.281 + 1.525\,Ln\,(BD) + 0.831\,Ln\,(CD) + 0.523\,Ln\,(H))} \qquad (8)$$

The results showed that Model 1 tended to overestimate AGB_{est} at lower observed AGB values, showing an average relative error of +33% for shrubs weighing less than 10 kg (Figure 4a). At higher AGB values the mean error was reduced (Figure 4b). Comparison of models fits (Supporting Information Figure S3) showed that Model 1 and Model 2 had similar average systematic bias (−0.96 and −0.16%,

(a)

R² = .880
RMSE = 0.625
AIC = 2,745

(b)

R² = .800
RMSE = 0.806
AIC = 3,481

(c)

R² = .59
RMSE = 1.154
AIC = 4,518

FIGURE 2 Regression analysis between dry aboveground biomass of woody shrub individuals (AGB, kg) and allometric variables: (a) stem basal diameter (BD, cm), (b) height (H, m) and (c) mean crown diameter (CD, m). Statistical descriptors were obtained based on the same number of individuals for all models ($n = 1,444$). After obtaining comparable statistical descriptors, each model was rerun using the maximum number of individuals available for each variable (BD, $n = 1,933$; CD, $n = 2,620$; H, $n = 2,877$). Each dot represents an individual shrub. The fitted model is represented by the dashed line. RMSE = root mean square standard error; AIC = Akaike information criterion

CONTI ET AL.

Global Ecology
and Biogeography

A Journal of
Macroecology

WILEY | 7
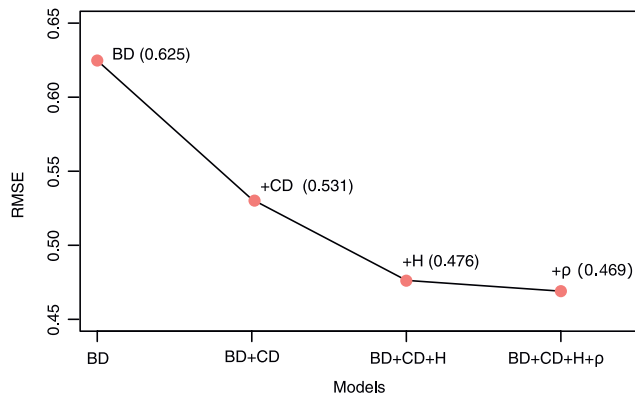
FIGURE 3 Effect of including different allometric variables on the root mean square error (RMSE) of global shrub biomass models. BD = stem basal diameter; BD + CD = stem basal diameter + mean crown diameter; BD + CD + H = stem basal diameter + mean crown diameter + height; BD + CD + H + ρ = stem basal diameter + mean crown diameter + height + wood density. The order of inclusion of the allometric variables followed their comparative goodness of fit. Values in parentheses are the RMSE for each particular model

respectively) but Model 1 had a better general fit in terms of AIC, R^2 and RMSE, although it included one variable more than Model 2. Model 3 had also good fit, considering that includes a single predictive variable, with an average systematic bias of −8.14%. Finally, Model 4 including CD and H, but not BD, had the lowest predictive capacity within the proposed models (average systematic bias of +12.9%).

Surprisingly, when comparing bias between predicted and observed values obtained for each model after applying the REML, MB and EV correction factors, together with the naive estimate, the latter gave the lowest bias in all cases but for Model 4. Only for this case, the application of the REML correction factor reduced the model RMSE (see details in Supporting Information Figure S4). We retained the simplest version of each of the selected models, and computed the REML correction factor only for the reported version of Model 4.

3.2 | Testing established allometric models

Model 1 (Equation 8) and Chave et al.'s model (Equation 3) were the best fitted models (Figure 5 and Supporting Information Figure S5a,b). However, Chave et al.'s model presented the lowest RMSE and average systematic bias (+1.5%) , while Model 1 comparatively showed a higher RMSE and average systematic bias (+10.4%). Across this comparison, both Paul et al.'s models (Equations 5 and 6), based on BD only, had higher RMSE than Model 1 and Model 2. Equation 6 (MULTI) fitted the data better than Equation 5 (SHRUB), which showed a higher underestimation of observed AGB (average systematic bias of +7.3% for MULTI versus +21.0% for SHRUB). The comparison of Paul et al.'s models with Model 3, also based on BD only, showed that Model 3 had an intermediate fit between the MULTI and SHRUB models in terms of RMSE (0.591) and average systematic bias (+13.7%) (Supporting Information Figure S5c,d). Finally, the model proposed by Jucker et al. (2017) (Equation 4), including only H and CD, showed the poorest performance across models. This model presented the highest RMSE and average systematic bias (in absolute terms; −67.3%), with a general underestimation of observed values (Figure 5). When Jucker et al.'s model was compared to Model 4 based on the same variables (CD and H), Model 4 performed better in terms of RMSE (0.653) and average systematic bias (+53.3%) (Supporting Information Figure S5e,f).

3.3 | Testing the effect of bioclimatic variables on models' performance

The models' parameters changed significantly in response to the inclusion of bioclimatic variables (Table 2). However, for Model 1, and Model 4 only in the case of bioform, model performance was considerably increased in terms of AIC and RMSE. For example, the inclusion of the variable bioform in Model 1 decreased AIC and RMSE by 17 and 7.3%, respectively, whereas the inclusion of biome or GAI reduced AIC and RMSE by no more than 5%. The pattern was maintained across the four proposed models with bioform as the variable that most increased model performance according to the AIC criteria. However, for the case of Models 2 and 3, AIC reduction was

TABLE 1 Comparison of shrub aboveground biomass regression models

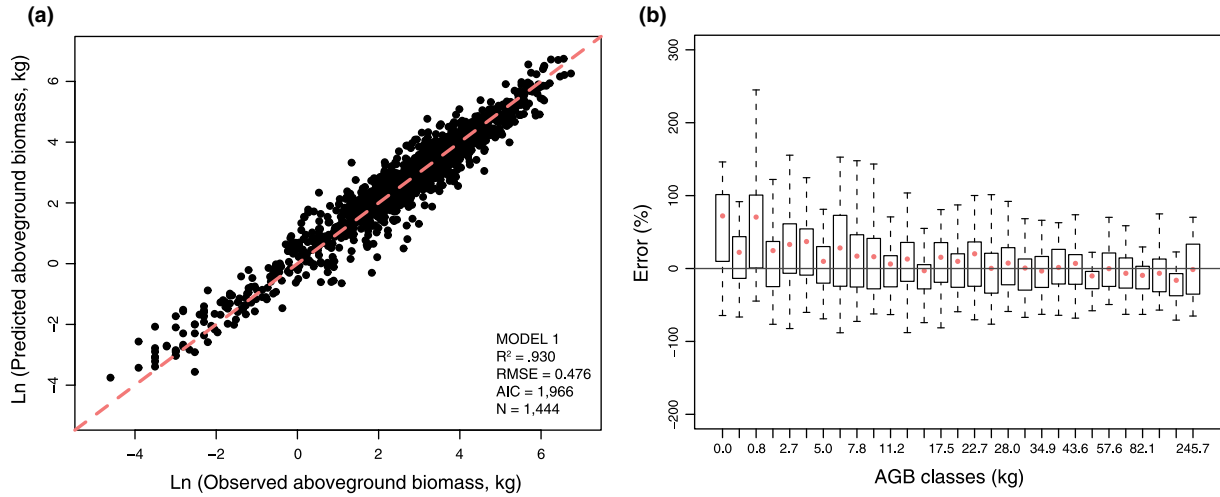| Models	Parameters	R^2	RMSE	AIC
Model 1	AGB_{est} = exp(−2.281 + 1.525 Ln (BD) + 0.831 Ln (CD) + 0.523 Ln (H))	0.930	0.476	1,966
Model 2	AGB_{est} = exp(−2.057 + 1.741 Ln (BD) + 0.945 Ln (CD))	0.913	0.531	2,278
Model 3	AGB_{est} = exp(−2.869 + 2.584 Ln (BD))	0.880	0.625	2,745
Model 4	AGB_{est} = exp(−0.370 + 1.903 Ln (CD) + 0.652 Ln (H))*1.403	0.859	0.677	2,979

Note. AGB_{est} = estimated aboveground dry biomass (kg); BD = stem basal diameter (cm); H = height (cm); CD = mean crown diameter (m); R^2 = coefficient of determination; RMSE = root mean square error; AIC = Akaike information criterion. All regression analyses were statistically significant ($p < 0.0001$). Statistical descriptors were obtained considering the same number of individuals (n = 1,444) for comparison. After obtaining statistical descriptors, model coefficients were recalculated using the maximum number of individuals available for each case: Model 1, n = 1633; Model 2, n = 1658; Model 3, n = 1933 and Model 4, n = 2,578.

FIGURE 4 Goodness of fit for Model 1 based on multiple allometric variables [stem basal diameter (BD), height (*H*), mean crown diameter (CD)] for predicting the aboveground biomass of shrub individuals. (a) Predicted and observed aboveground biomass (AGB) values; the dashed line corresponds to a 1:1 relationship. Each dot represents a woody individual. (b) Mean relative error [Error = (AGB_{est} – AGB_{obs}/ AGB_{obs}) × 100] for different AGB classes, with the bars delimiting the interquartile range (boxes) and 95% limits (dotted lines) of the errors. Note that when the predicted value is greater than the observed value, overestimation occurs and the error is positive; conversely, negative error values represent underestimation. Pink points represent the mean error for each AGB class. Total number of AGB values was divided into classes of the same length, where each number shown on the AGB axis is representative of the initial AGB value for each class. RMSE = root mean square error; AIC = Akaike information criterion
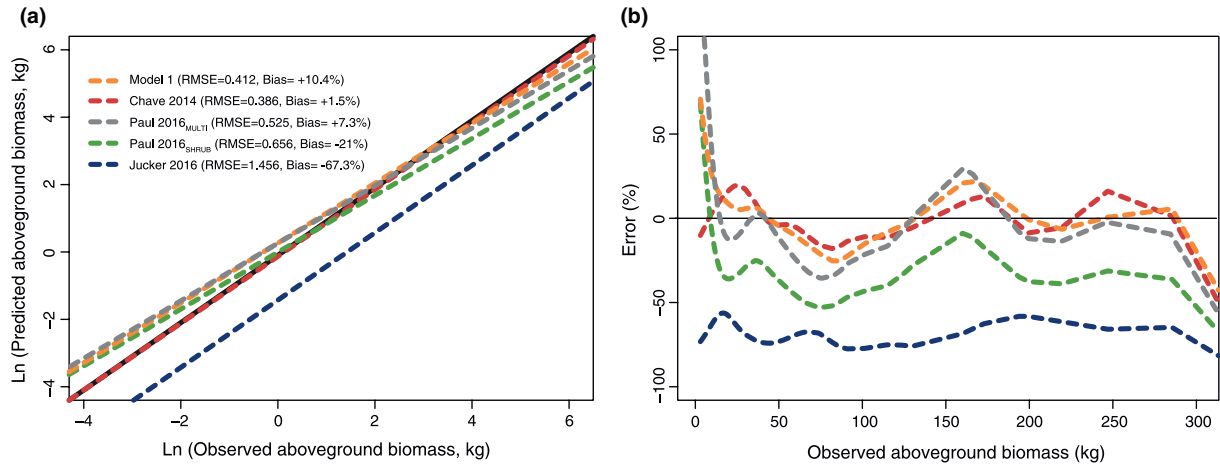
FIGURE 5 Comparison of models' performances using the fitting dataset. (a) Predicted and observed aboveground biomass (AGB) values for the selected models. (b) Mean relative errors [Error = (AGB_{est} – AGB_{obs}/AGB_{obs}) × 100] for the compared models across the observed aboveground biomass values. The compared models correspond to Model 1 proposed here; the global model proposed by Chave et al. (2014), including diameter at breast height (DBH), height (*H*) and wood density (ρ); the model proposed by Jucker et al. (2017) for angiosperms including mean crown diameter (CD) and *H*; and the shrub models proposed by Paul et al. (2016) using stem diameter at 10 cm height (D10) as a main variable (MULTI and SHRUB models). RMSE = root mean square error; Bias = average relative systematic error of model predictions

also higher when both biome (AIC decreased by 11.2 and 15.5% for Models 2 and 3, respectively) and GAI (AIC decreased by 9.4 and 14%, for Models 2 and 3, respectively) were included. Supporting Information Tables S5 to S8 show the results for all remaining models.

4 | DISCUSSION

By analysing a worldwide dataset of 3,243 woody individuals, we developed four different global models to estimate shrub AGB

biomass, considering a combination of the most frequently sampled allometric variables (BD, CD and *H*). Our work supports the effectiveness of generic biomass allometric models developed from large datasets for shrub species, consistent with comparable models developed for trees and multistemmed woody individuals across forest ecosystems (Chave et al., ; Jucker et al., 2017; Paul et al., 2016). The research presented here has the added value of presenting alternative models for those cases where stem diameter is difficult, or even impossible, to obtain in the field, a very common situation in the case of multistemmed woody individuals. Furthermore, we found that the

TABLE 2 Absolute change (%) in the predictive capacity of models after including different bioclimatic variables

Models	AIC	RMSE	AIC change (%)	RMSE change (%)
Model 1	1,966	0.476		
Model 1 + bioform	1,631	0.441	**17.04**	**7.35**
Model 1 + biome	1,875	0.452	4.63	5.04
Model 1 + GAI	1,881	0.457	4.32	3.99
Model 2	2,278	0.531		
Model 2 + bioform	1,912	0.493	**16.07**	**7.16**
Model 2 + biome	2,022	0.477	**11.24**	**10.17**
Model 2 + GAI	2,064	0.485	**9.39**	**8.66**
Model 3	2,745	0.625		
Model 3 + bioform	2,235	0.558	**18.58**	**10.72**
Model 3 + biome	2,319	0.530	**15.52**	**15.20**
Model 3 + GAI	2,360	0.538	**14.03**	**13.92**
Model 4	2,979	0.677		
Model 4 + bioform	2,492	0.614	**16.35**	**9.31**
Model 4 + biome	2,882	0.647	3.26	4.43
Model 4 + GAI	2,967	0.668	0.40	1.33

Note. AIC = Akaike information criterion; RMSE = root mean square error; GAI = global aridity index. Values in bold represent changes >5%. Statistical descriptors were obtained considering the same number of individuals (n = 1,444) for comparison. Mixed models including random intercept and slope had the best fit in all cases, except for Model 4 for which mixed model including random slope but fixed intercept had the best fit for the case of bioform.

fitness of the models was improved by including bioclimatic variables, showing that scaling relationships differed mainly for different bioforms. However, the original models still fit the data satisfactorily and their use is simpler and more practical. Here, we discuss several aspects that arise from the results that need to be considered to use these models for accurate estimation of individual AGB across shrub species and ecosystems.

4.1 | Developing a global shrub species aboveground model

Substantial work has been dedicated to develop species- and site-specific models for shrub species worldwide, encompassing a huge range of species and environments including subtropical forests of Mexico, the USA, Canada (e.g., Alaback, 1986; Ludwig, Reynolds, & Whitson, 1975; Murray & Jacobson, 1982; Návar et al., 2004), China (Zeng, Liu, Feng, & Ma, 2010) and South America (Conti et al., 2013; Haase & Haase, 1995; Hierro et al., 2000; Sampaio, 2005); European Mediterranean shrublands (Paton, Nuñez, Bao, & Muñoz, 2002); and the Patagonian steppe (Oñatibia, Aguiar, Cipriotti, & Troiano, 2010), among others. The variables tested were seldom other than stem diameter-related variables, crown-related variables and height, all with good fits. These local studies represent enormous efforts to improve biomass models to include these woody species in local and global carbon budgets. However, these models are mostly species-specific and rely on a low number of individuals (typically < 100), which limits application to the local scale or to a few species. Approaches aimed to develop a global biomass model based on a relatively large sample

size (>1,000) have the added value of greatly reducing uncertainty in parameter estimates, as already shown by models developed for trees around the world (i.e., Chave et al., 2004; Jucker et al., 2017; Roxburgh, Paul, Clifford, England, & Raison, 2015). For the case of shrubs, to our knowledge, only Paul et al. (2016) have compiled a huge dataset across several functional types at continental level in Australia, including multistemmed shrubs and small trees, and using stem diameter, H and ρ as the main predictor variables. The analysis presented in Paul et al. (2016) showed that AGB prediction based only on BD, or even including H and ρ, tended to be less accurate for multistemmed shrubs in comparison with trees. The authors explained this low model fit by the problems associated with diameter sampling across species belonging to the former (Paul et al., 2017). Our work went further by compiling a global dataset of shrub individuals, including not only BD, H and ρ, but also CD, as predictor variables of a global shrub biomass model with potential wide application across different phytogeographic regions and climatic conditions.

Stem diameter seems to be the best single predictive variable of biomass across woody individuals, as found by other studies (Brown, 1997; Chave et al., 2005; Paul et al., 2016). However, accurate field measurement of shrub BD can be demanding, particularly in highly branched or spiny species where stem diameter measurement is difficult, slow and also dangerous due to exposure to potential hazards, often resulting in inaccurate measures as the operator needs to physically crawl or bend to get close access to the stem (Paul et al., 2017). In such cases, an alternative model without BD (Model 4) is needed to accurately estimate AGB, even if there is a goodness of fit trade-off.

In agreement with Goodman et al. (2014) and Jucker et al. (2017) our results indicate that AGB models incorporating a crown-related variable have significantly improved predictive power, even more so than including height and wood density (Figure 3). The crown represents a relatively higher amount of biomass in shrubs than in trees, in which the main stem accounts for the major proportion of biomass, and this is probably why this variable markedly increased the predictive power of shrub AGB models (a RMSE reduction of 15%), followed by *H* (an additional RMSE reduction of 10%) and lastly by ρ (other additional RMSE reduction of 1.5%). Although the inclusion of several variables could be impractical for field sampling, our study showed that it is difficult to capture the volume occupied by a multistemmed individual with only one variable (or dimension).

Our models did not improve significantly after including ρ, in contrast with what was found by other authors for tree models (Chave et al., 2005, 2014). A potential explanation of this pattern could be related to the level of sampling (measurement error) as was also suggested by Paul et al. (2016), given that species' ρ values were obtained from databases rather than at the individual level. There is accumulated evidence that ρ varies as a function of height, ramification degree and age (Chave et al., 2009). Hence, database-derived ρ values could increase variability, rather than reduce it. It could be interesting to test if reducing this methodological bias could significantly improve the predictive power of biomass models.

4.2 | Practical field considerations in order to accurately estimate shrub biomass

Much work has been devoted to discussing how to adequately estimate individual AGB in trees (Brown, 1997; Chave et al., 2014; Feldpausch et al., 2012; Ketterings et al., 2001; Ploton et al., 2016). On the contrary, protocols for shrubs are not well developed (but see Chojnacky & Milton, 2008). In the quest to establish general recommendations for assessing shrub individual biomass, our results revealed that any measure of the total basal area of an individual is by far the best proxy of its total AGB. This coincides with the findings of Haase and Haase (1995). However, measuring stem diameter in shrubs can be problematic when individuals ramify from the base. In these cases, its field measurement can add uncertainty and bring about huge methodological errors. Through the analysis of the database used here, we detected that the decision on how to sample shrub stem diameter was entirely up to the researcher. There was no methodological standardization on what to do in the case of multistemmed individuals, which often have more than 20 ramifications from the base. To develop our models we chose to consider as BD any stem diameter sampled above the root collar and below 30 cm of stem height. This decision allowed us to include more studies and simplified the standardization procedure. However, to reduce the amount of noise introduced into collated datasets, we recognize the need to standardize the height selected to measure the stem diameter. We therefore recommend

sampling the stem diameter at 10 cm stem height (D10), as proposed by Paul et al. (2016). If ramifications exist below this level, then we recommend also recording their diameters (D_i) to obtain a single value of basal stem diameter (BD) representing all basal area at this height, calculated as follows:

$$BD = \sqrt{\sum D_i^2} \tag{9}$$

Paul et al. (2016) also presented simple equations to convert measures of D0, D30, D50 and D130 to D10 that we include in Supporting Information Table S4 to have a complete standardized protocol to recommend. Improved accuracy of the models being developed could be achieved if more rigor is applied to the height of stem diameter measurement required to derive BD.

Even though stem diameter measures could be standardized across protocols and studies, sometimes it can be impossible to sample this variable in the field. Measuring crown diameter has then the advantage of being relatively easy to sample in the field for small- and medium-sized shrub individuals (Northup, Zitzer, Archer, McMurtry, & Boutton, 2005) without methodological constraints. To assess average crown diameter in the field, we recommend measuring the maximum crown diameter and its perpendicular diameter with a metric tape, and then calculating their mean.

We found a large dispersion in shrub height values across our dataset, which could be related to methodological aspects that are not usually addressed. To have a comparable measure of height, it is important to know if it refers to the standing maximum height of the shrub crown, the total length of the highest branch as maximum height, or the apparent average maximum height of the crown. It is also important to know if height was measured using a metric tape, a telescopic stick, a laser rangefinder, a clinometer, or if it was visually estimated. In all, these methodological issues may introduce significant bias when comparing different datasets, as shown here. However, the inclusion of height as a predictive variable in a biomass model, improved its performance, as was also found for trees (Feldpausch et al., 2012). We recommend considering height as the shortest distance between the upper boundary of the main photosynthetic tissues of a plant and the ground level, expressed in metres and measured on the standing plant before harvest. For estimating the height of short shrubs, a metric tape can be used, but a telescopic stick with metre marks or a laser rangefinder could be useful for taller shrubs.

Several methodologies have been developed for estimating crown area, height and stem diameter from remote sensing (e.g., Barbier, Couteron, Proisy, Malhi, & Gastellu-Etchegorry, 2010; Jucker et al., 2017), yet their calibration still relies on the accuracy of ground-based biomass (Baccini et al., 2012; Saatchi et al., 2011; Le Toan et al., 2011). It is important to understand the uncertainty linked to the use of different methodologies in order to reduce potential bias in the estimations.

Estimation of shrub biomass can be improved by considering plant dimensions other than diameter. However, the goal of developing and applying AGB allometric models is to enable site-specific estimation of biomass based on plot measurements of plant individual

allometric variables (BD, CD and H). Given that site-level sampling errors are often the largest source of error in these estimates (Paul et al., 2017), they are likely to be minimized by increasing sample size. Any decision on increasing the number of predictive variables should not be at the expense of reducing sample size. On the other hand, a generalized allometric model may be applied across multiple sites, but validation is important when an existing generic multispecies model is applied to a species not represented within the existing model. According to Paul et al. (2018), there are no low-cost (i.e., small N) options for such validation. In those cases, sampling intensities of N > 50 individuals should be used to validate the existing model (Roxburgh et al., 2015), in such a way that even if the validation fails, the new dataset is sufficiently large to develop a new species-specific model. If the validation is successful, the new data may be combined with the existing dataset to provide a revised generic allometric model (Paul et al., 2018). All these decisions will definitely depend on the objective and the budget of the particular research project.

We developed models to estimate individual AGB of shrubs worldwide; however, we should mention that biomass estimates generally require both above- and belowground biomass. Paul et al. (2019) developed models to estimate individual belowground biomass of Australian shrubs and multistemmed trees, based on diameter at 10 cm stem height. This has to be considered when planning root biomass quantification based on aboveground allometric predictive variables. Other studies have also estimated tree belowground biomass based on total AGB (Cairns, Brown, Helmer, & Baumgardner, 1997), and based on the joint variation of root to shoot ratios and DBH (Ledo et al., 2018). A standardized protocol for quantifying belowground biomass across distinct woody life-forms and different ecosystems is also an imperative future step.

4.3 | Testing established allometric models

The four proposed models were compared with previously published and established biomass models for trees (Chave et al., 2014; Jucker et al., 2017) and small and/or multistemmed trees and shrubs (Paul et al., 2016). Our models had similar fit to that of the most used tree biomass model (Chave et al., 2014), even with a relatively lower sample size. We also showed that Chave et al.'s model had a good fit for shrubs, even when the fitting subset was in part (8.7%) out of the range of applicability of Chave et al.'s model, including some individuals with DBH < 5 cm. However, it is important to note that, although it is possible to use Chave et al.'s model to predict the biomass of individual shrubs with good fit, this is only achievable if DBH can be effectively sampled on those individuals, which is unlikely for most multistemmed woody individuals. As an example, to carry out model comparison we had to reduce our fitting dataset by 62.3% in order to include only shrub individuals for which DBH was sampled.

Model 3 based only on BD also performed well in comparison to those proposed by Paul et al. (2016) based on the same variable. This is particularly true for Paul et al's MULTI model based on a dataset with similar applicability range, but comprising

a significantly greater number of individuals (N = 5,397), which explains its better performance in comparison with Model 3. However, Paul et al.'s SHRUB model yielded higher underestimates for AGB > 10 kg in comparison with Model 3 and Paul et al.'s MULTI model. Differences in the fit of these models can be explained mainly by the fact that Paul et al.'s SHRUB model considered individuals smaller than those included here (but our data are still within its range of applicability), which could produce higher bias for individuals less represented in their dataset (>10 kg). When comparing Model 4 with the model proposed by Jucker et al. (2017), also including H and CD as predictive variables, we found that our model performed better due to a consistent underestimation of Jucker et al.'s predicted values. Differences in model fit in this case could be mainly due to the different procedures used to fit biomass models (GLS in our models in comparison with data binning in Jucker et al.'s model).

In general, our models represent a well-fitted set of shrub biomass models, complementing well-established and commonly used AGB tree models and spanning across diverse ecosystems unrepresented previously. They provide further evidence of the effectiveness of generic biomass allometric models developed from large datasets, consistent with comparable models developed for trees across forest ecosystems. More accurate estimates of forest biomass and carbon results from the inclusion of small, multistemmed woody individuals, a biomass compartment previously underestimated. We further proposed an alternative model when BD cannot be adequately sampled, including CD as an important predictive variable, as well as H.

Depending on which allometric variable can be more precisely sampled at individual level, we suggest the following general procedure when estimating AGB in woody ecosystems where no regional models exist and destructive sampling is unfeasible (see schematic protocol in Figure 6):

1. If DBH can be accurately sampled (e.g., tall woody individuals branching off at stem height >130 cm), apply Chave et al.'s (2014) tree model (Equation 3);
2. If DBH cannot be sampled adequately (e.g., woody individuals branching off below 130 cm but above 10 cm stem height), but BD and CD or H can be recorded, apply our proposed Models 1 or 2 (Table 1);
3. If only BD (but not CD or H) can be sampled (e.g., woody individuals branching off below 130 cm but above 10 cm height, but with a very open canopy and twisted stems) apply Paul's MULTI model (Equation 6). This option is also useful when researchers need to reduce the individual sampling effort to have a decreased site-level estimation error by increasing the number of sampled individuals (i.e., maximizing the accuracy-to-cost ratio by measuring more individuals rather than spending more time maximizing the accuracy at individual level).
4. Finally, if precise acquisition of BD is impossible and it is only possible to record CD and H (e.g., woody individuals branching off just above root collar), apply our proposed Model 4 (Table 1).
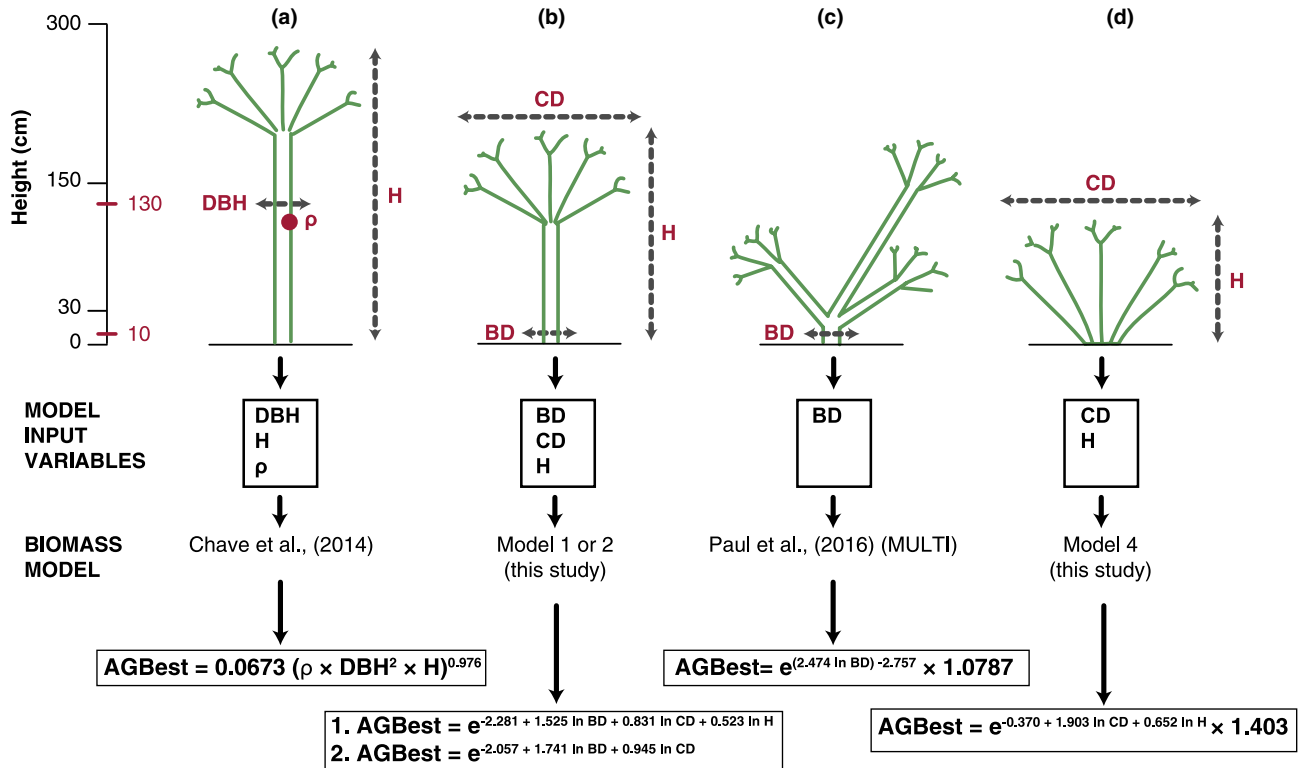
FIGURE 6 Schematic protocol recommended for applying general aboveground biomass (AGB) models for woody individuals. (a)–(d) Different individual woody physiognomy types sampled in the field. For each case, we suggest measuring different sets of allometric variables in order to apply the recommended biomass models. DBH = diameter at 130 cm stem height (cm); BD = stem basal diameter (cm); CD = mean crown diameter (m); H = height (m); ρ = wood density (g/cm^3); AGB_{est} = estimated AGB. See detailed explanation in the main text

4.4 | Including bioclimatic variables

A key requirement for developing more accurate allometric equations is to incorporate all of the appropriate variables that affect AGB, such as ρ, and the relationship between allometric variables subject to variation due to geographic constraints (Banin et al., 2012; Goodman et al., 2014). Our database comprised different bioforms and sites distributed across different biomes and climatic conditions, allowing us to test the hypothesis that changes in these scaling relationships could affect the predictive power of global biomass models. Results confirmed the fact that the scaling relationship between the allometric variables used to estimate AGB did vary with different bioclimatic variables. However, only bioform significantly strengthened the model. These results agree with Paul et al. (2016) and Jucker et al. (2017), where the inclusion of site-related factors (stand and climate characteristics) did not markedly improve the predictive ability of the allometric models, but the inclusion of plant architecture- or physiognomy-type did. Varying bioforms or species' architecture reflect different energy investment strategies likely to result in different crown–mass ratios among woody individuals with similar size (Ploton et al., 2016). Across our models, the effects of other bioclimatic variables (biome and GAI) significantly improved model fits particularly when height was not included in the model (Models 2 and 3). Although not conclusive, this may indicate that different allometric dimensions could be differentially limited across bioclimatic regions, and that accounting for more allometric variables may reduce this variation across sites, and so the associated uncertainty, as was found for trees (Blanchard et al., 2016). Future research is needed to explore the differences in shrubs' scaling relationships across varying bioclimatic conditions to accurately account for these differences. As a general conclusion, the inclusion of bioclimatic variables, especially bioform, contributes to more accurate estimates of individual shrub AGB. Despite this improvement, the original Model 1 remains strong enough on its own to be used across different bioforms, biomes and climatic regions, but decision on the final model used is at the discretion of the researcher.

In summary, information on stem basal diameter, crown diameter and height can be combined in different ways to provide a robust AGB estimate of individual shrubs, even more accurate than locally developed previous estimates. Our study supplements previous well-established models developed for trees, allowing more accurate biomass estimation of shrubs that are not usually accounted for when quantifying biomass and carbon stocks. We further provide tools for a methodological standardization of individual biomass quantification in shrub species worldwide. However, it is necessary to highlight that this method provides estimates—not direct measurements—and model errors should always be carefully examined

and analysed. With this work we expect not only to assist in improving the quality of biomass estimates across different ecosystems, but also to contribute to methodological consensus on field assessments of biomass in small and/or multistemmed woody species. This will benefit advancement towards a global mechanism to boost climate change mitigation projects.

DATA ACCESSIBILITY

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.282bm46. Data are fully public but should be appropriately referenced by citing this data paper. We additionally suggest that if you plan to use the data or want to contribute with new data, please contact the data compilers.

ORCID

G. Conti https://orcid.org/0000-0002-5420-6843

P. M. Fernandes http://orcid.org/0000-0003-0336-4398

REFERENCES

Alaback, P. B. (1986). Biomass regression equations for understory plants in coastal Alaska: Effects of species and sampling design on estimates. *Northwest Science*, 60, 90–103.

Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., ... Houghton, R. A. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, 2, 182–185. https://doi.org/10.1038/nclimate1354

Banin, L., Feldpausch, T. R., Phillips, O. L., Baker, T. R., Lloyd, J., Affum-Baffoe, K., ... Lewis, S. L. (2012). What controls tropical forest architecture? Testing environmental, structural and floristic drivers. *Global Ecology and Biogeography*, 21, 1179–1190. https://doi.org/10.1111/j.1466-8238.2012.00778.x

Barbier, N., Couteron, P., Proisy, C., Malhi, Y., & Gastellu-Etchegorry, J. (2010). The variation of apparent crown size and canopy heterogeneity across lowland Amazonian forests. *Global Ecology and Biogeography*, 19, 72–84. https://doi.org/10.1111/j.1466-8238.2009.00493.x

Barton, K. (2016). *MuMIn: Multi-model inference*. R package version 1.40. 4. CRAN. Retrieved from https://cran.r-project.org/package=MuMIn

Baskerville, G. L. (1972). Use of logarithmic regression in the estimation of plant biomass. *Canadian Journal of Forest Research*, 2(1), 49–53. https://doi.org/10.1139/x72-009

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823*.

Blanchard, E., Birnbaum, P., Ibanez, T., Boutreux, T., Antin, C., Ploton, P., ... Couteron, P. (2016). Contrasted allometries between stem diameter, crown area, and tree height in five tropical biogeographic areas. *Trees*, 30, 1953–1968. https://doi.org/10.1007/s00468-016-1424-3

Brown, S. (1997). *Estimating biomass and biomass change of tropical forests: A primer*. FAO Forestry Paper (Vol. 134, 57p). Rome, Italy: A Forest Resource Assessment Publication.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and inference. A practical information-theoretic approach* (2nd ed.). Berlin, Germany: Springer.

Cairns, M. A., Brown, S., Helmer, E. H., & Baumgardner, G. A. (1997). Root biomass allocation in the world's upland forests. *Oecologia*, 111(1), 1–11. https://doi.org/10.1007/s004420050201

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., ... Yamakura, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, 145, 87–99. https://doi.org/10.1007/s00442-005-0100-x

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., & Pérez, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 409–420. https://doi.org/10.1098/rstb.2003.1425

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. https://doi.org/10.1111/j.1461-0248.2009.01285.x

Chave, J., Rejou-Mechain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., ...Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20, 3177–3190. https://doi.org/10.1111/gcb.12629

Chojnacky, D. C., & Milton, M. (2008). Measuring carbon in shrubs. In C. M. Hoover (Ed.), *Field measurements for forest carbon monitoring* (pp. 45–72). New York, NY: Springer.

Clifford, D., Cressie, N., England, J. R., Roxburgh, S. H., & Paul, K. I. (2013). Correction factors for unbiased, efficient estimation and prediction of biomass from log–log allometric models. *Forest Ecology and Management*, 310, 375–381. https://doi.org/10.1016/j.foreco.2013.08.041

Conti, G., Enrico, L., Casanoves, F., & Díaz, S. (2013). Shrub biomass estimation in the semiarid Chaco forest: A contribution to the quantification of an underrated carbon stock. *Annals of Forest Science*, 70, 515–524. https://doi.org/10.1007/s13595-013-0285-9

El-Shaarawi, A. H., & Viveros, R. (1997). Inference about the mean in log-regression with environmental applications. *Environmetrics: the Official Journal of the International Environmetrics Society*, 8(5), 569–582. https://doi.org/10.1002/(SICI)1099-095X(199709/10)8:5<569:AID-ENV274>3.0.CO;2-I

Feldpausch, T. R., Lloyd, J., Lewis, S. L., Brienen, R. J. W., Gloor, M., Monteagudo Mendoza, A., ... Affum-Baffoe, K. (2012). Tree height integrated into pantropical forest biomass estimates. *Biogeosciences*, 9, 3381–3403. https://doi.org/10.5194/bg-9-3381-2012

Goodman, R. C., Phillips, O. L., & Baker, T. R. (2014). The importance of crown dimensions to improve tropical tree biomass estimates. *Ecological Applications*, 24, 680–698. https://doi.org/10.1890/13-0070.1

Haase, R., & Haase, P. (1995). Above-ground biomass estimates for invasive trees and shrubs in the Pantanal of Mato Grosso, Brazil. *Forest Ecology and Management*, 73, 29–35. https://doi.org/10.1016/0378-1127(94)03509-U

Hierro, J. L., Branch, L. C., Villarreal, D., & Clark, K. L. (2000). Predictive equations for biomass and fuel characteristics of Argentine shrubs. *Journal of Range Management*, *53*, 617–621. https://doi.org/10.2307/4003156

Hijmans, R. J., & van Etten, J. (2017). *raster: Geographic data analysis and modeling*. R package version, 2 Retrieved from https://cran.r-project.org/web/packages/raster/index.html

Hofstad, O. (2005). Review of biomass and volume functions for individual trees and shrubs in Southeast Africa. *Journal of Tropical Forest Science*, *17*, 151–162.

Huang, J., Yu, H., Guan, X., Wang, G., & Guo, R. (2016). Accelerated dryland expansion under climate change. *Nature Climate Change*, *6*, 166–171.

Ishihara, M. I., Utsugi, H., Tanouchi, H., Aiba, M., Kurokawa, H., Onoda, Y., ... Hiura, T. (2015). Efficacy of generic allometric equations for estimating biomass: A test in Japanese natural forests. *Ecological Applications*, *25*, 1433–1446. https://doi.org/10.1890/14-0175.1

Jucker, T., Caspersen, J., Chave, J., Antin, C., Barbier, N., Bongers, F., ... Coomes, D. A. (2017). Allometric equations for integrating remote sensing imagery into forest monitoring programmes. *Global Change Biology*, *23*, 177–190. https://doi.org/10.1111/gcb.13388

Ketterings, Q. M., Coe, R., van Noordwijk, M., Ambagau, Y., & Palm, C. A. (2001). Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. *Forest Ecology and Management*, *146*, 199–209. https://doi.org/10.1016/S0378-1127(00)00460-6

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A. C., ... Bakker, D. C. E. (2018). Global carbon budget 2017. *Earth System Science Data*, *10*, 405–448. https://doi.org/10.5194/essd-10-405-2018

Le Toan, T., Quegan, S., Davidson, M. W. J., Balzter, H., Paillou, P., Papathanassiou, K., ... Ulander, L. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, *115*, 2850–2860. https://doi.org/10.1016/j.rse.2011.03.020

Ledo, A., Paul, K. I., Burslem, D. F. R. P., Ewel, J. J., Barton, C., Battaglia, M., ... England, J. R. (2018). Tree size and climatic water deficit control root to shoot ratio in individual trees globally. *New Phytologist*, *217*, 8–11. https://doi.org/10.1111/nph.14863

Ludwig, J. A., Reynolds, J. F., & Whitson, P. D. (1975). Size-biomass relationships of several Chihuahuan desert shrubs. *The American Midland Naturalist*, *94*, 451–461. https://doi.org/10.2307/2424437

Lutz, J. A., Furniss, T. J., Johnson, D. J., Davies, S. J., Allen, D., Alonso, A., ... Zimmerman, J. K. (2018). Global importance of large-diameter trees. *Global Ecology and Biogeography*, *27*, 849–864. https://doi.org/10.1111/geb.12747

Middleton, N. J., & Thomas, D. S. (1992). *World atlas of desertification* (2nd ed.). London, UK: United Nations Environment Programme.

Murray, R. B., & Jacobson, M. Q. (1982). An evaluation of dimension analysis for predicting shrub biomass. *Journal of Range Management*, *35*, 451–454. https://doi.org/10.2307/3898603

Návar, J., Méndez, E., Nájera, A., Graciano, J., Dale, V., & Parresol, B. (2004). Biomass equations for shrub species of Tamaulipan thornscrub of North-eastern Mexico. *Journal of Arid Environments*, *59*, 657–674. https://doi.org/10.1016/j.jaridenv.2004.02.010

Northup, B. K., Zitzer, S. F., Archer, S., McMurtry, C. R., & Boutton, T. W. (2005). Above-ground biomass and carbon and nitrogen content of woody species in a subtropical thornscrub parkland. *Journal of Arid Environments*, *62*, 23–43. https://doi.org/10.1016/j.jaridenv.2004.09.019

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., ... Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on earth. *BioScience*, *51*, 933. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Oñatibia, G. R., Aguiar, M. R., Cipriotti, P. A., & Troiano, F. (2010). Individual plant and population biomass of dominant shrubs in Patagonian grazed fields. *Ecologia Austral*, *20*, 269–279.

Paton, D., Nuñez, J., Bao, D., & Muñoz, A. (2002). Forage biomass of 22 shrub species from Monfragüe Natural Park (SW Spain) assessed by log–log regression models. *Journal of Arid Environments*, *52*, 223–231. https://doi.org/10.1006/jare.2001.0993

Paul, K. I., Larmour, J. S., Roxburgh, S. H., England, J. R., Davies, M. J., & Luck, H. D. (2017). Measurements of stem diameter: Implications for individual- and stand-level errors. *Environmental Monitoring and Assessment*, *189*, 189–416. https://doi.org/10.1007/s10661-017-6109-x

Paul, K. I., Larmour, J., Specht, A., Zerihun, A., Ritson, P., Roxburgh, S. H., ... England, J. R. (2019). Testing the generality of below-ground biomass allometry across plant functional types. *Forest Ecology and Management*, *432*, 102–114. https://doi.org/10.1016/j.foreco.2018.08.043

Paul, K. I., Radtke, P. J., Roxburgh, S. H., Larmour, J., Waterworth, R., Butler, D., ... Ximenes, F. (2018). Validation of allometric biomass models: How to have confidence in the application of existing models. *Forest Ecology and Management*, *412*, 70–79. https://doi.org/10.1016/j.foreco.2018.01.016

Paul, K. I., Roxburgh, S. H., Chave, J., England, J. R., Zerihun, A., Specht, A., ... Sinclair, J. (2016). Testing the generality of above-ground biomass allometry across plant functional types at the continent scale. *Global Change Biology*, *22*, 2106–2124. https://doi.org/10.1111/gcb.13201

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). *Nlme: Linear and nonlinear mixed effects models*. R package version 3.1-128. 2016. R software.

Ploton, P., Barbier, N., Momo, S. T., Rejou-Mechain, M., Boyemba Bosela, F., Chuyong, G., ... Pélissier, R. (2016). Closing a gap in tropical forest biomass estimation: Taking crown mass variation into account in pantropical allometries. *Biogeosciences*, *13*, 1571–1585. https://doi.org/10.5194/bg-13-1571-2016

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., ... van der Werf, G. R. (2014). Contribution of semi-arid ecosystems to inter-annual variability of the global carbon cycle. *Nature*, *509*, 600–603. https://doi.org/10.1038/nature13376

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Richardson, D. M., & Rejmánek, M. (2011). Trees and shrubs as invasive alien species—A global review. *Diversity and Distributions*, *17*, 788–809. https://doi.org/10.1111/j.1472-4642.2011.00782.x

Roxburgh, S. H., Paul, K. I., Clifford, D., England, J. R., & Raison, R. J. (2015). Guidelines for constructing allometric models for the prediction of woody biomass: How many individuals to harvest? *Ecosphere*, *6*(3), 1–27.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., ... Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences USA*, *108*, 9899–9904. https://doi.org/10.1073/pnas.1019576108

Sampaio, E. (2005). Biomass equations for Brazilian semiarid caatinga plants. *Acta Botanica Brasilica*, *19*, 935–943. https://doi.org/10.1590/S0102-33062005000400028

Shen, H., & Zhu, Z. (2008). Efficient mean estimation in log-normal linear models. *Journal of Statistical Planning and Inference*, *138*(3), 552–567. https://doi.org/10.1016/j.jspi.2006.10.016

Skole, D. L., Samek, J. H., & Smalligan, M. J. (2011). Implications of allometry. *Proceedings of the National Academy of Sciences USA*, *108*, E12–E12. https://doi.org/10.1073/pnas.1015854108

Slik, J. W. F., Paoli, G., Mcguire, K., Amaral, I., Barroso, J., Bastian, M., ... Zweifel, N. (2013). Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics. *Global Ecology and Biogeography*, *22*, 1261–1271. https://doi.org/10.1111/geb.12092

Swenson, N. G., & Enquist, B. J. (2007). Ecological and evolutionary determinants of a key plant functional trait: Wood density and its

community-wide variation across latitude and elevation. *American Journal of Botany*, *94*, 451–459. https://doi.org/10.3732/ajb.94.3.451

Tummers, B. (2006). *Data thief III*. Retrieved from https://datathief.org/

Vesk, P. A., Warton, D. I., & Westoby, M. (2004). Sprouting by semi-arid plants: Testing a dichotomy and predictive traits. *Oikos*, *107*, 72–89. https://doi.org/10.1111/j.0030-1299.2004.13122.x

Vieilledent, G., Vaudry, R., Andriamanohisoa, S. F. D., Rakotonarivo, O. S., Randrianasolo, H. Z., Razafindrabe, H. N., … Rasamoelina, M. (2012). A universal approach to estimate biomass and carbon stock in tropical forests using generic allometric models. *Ecological Applications*, *22*, 572–583. https://doi.org/10.1890/11-0039.1

Zeng, H. Q., Liu, Q. J., Feng, Z. W, & Ma, Z. Q. (2010). Biomass equations for four shrub species in subtropical China. *Journal of Forest Research*, *15(2)*, 83–90. https://doi.org/10.1007/s10310-009-0150-8

Zizka, A., Govender, N., & Higgins, S. I. (2014). How to tell a shrub from a tree: A life-history perspective from a South African savanna. *Austral Ecology*, *39*, 767–778. https://doi.org/10.1111/aec.12142

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

BIOSKETCH

From the beginning, Conti´s scientific interests were in the area of plant–soil ecology. She has worked on research questions regarding comparative plant functional ecology and its relationship with the ecosystem services provided by semi-arid forests. Particularly, she is interested in the carbon storage potential of drylands and semi-arid ecosystems, quantifying biomass and C stocks in the dry Chaco of South America. Currently her research concerns the processes controlling the fertility and sequestration of soil carbon, mediated by the rhizosphere.

How to cite this article: Conti G, Gorné LD, Zeballos SR, et al. Developing allometric models to predict the individual aboveground biomass of shrubs worldwide. *Global Ecol Biogeogr*. 2019;00:1–15. https://doi.org/10.1111/geb.12907