Open Access Earth System Science Data Discussions

A biomass equation dataset for common shrub species in China

Yang Wang[1], Wenting Xu[1], Zhiyao Tang[2],and Zongqiang Xie[1,3]

[1]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China
5 [2]Department of Ecology, College of Urban and Environmental Sciences and Key Laboratory for Earth Surface Processes, Peking University, Beijing, 100871, China
[3]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, 100190, China

Correspondence to: Zongqiang Xie (xie@ibcas.ac.cn)

10 **Abstract.** Shrub biomass equations provide an accurate, efficient and convenient method in estimating biomass of shrubland ecosystems and biomass of the shrub layer in forest ecosystems at various spatial and temporal scales. In recent decades, many shrub biomass equations have been reported mainly in journals, books and postgraduate's dissertations. However, these biomass equations are applicable for limited shrub species with respect to a large number of shrub species widely distributed in China, which severely restricted the study of terrestrial ecosystem structure and function, such as biomass, production, and 15 carbon budge. Therefore, we firstly carried out a critical review of published literature (from 1982 to 2019) on shrub biomass equations in China, and then developed biomass equations for the dominant shrub species using a unified method based on field measurements of 738 sites in shrubland ecosystems across China. Finally, we constructed the first comprehensive biomass equation dataset for China's common shrub species. This dataset consists of 822 biomass equations specific to 167 shrub species and has significant representativeness to the geographical, climatic and shrubland vegetation features across 20 China. The dataset is freely available at https://doi.org/10.11922/sciencedb.00641 (Wang et al., 2021) for non-commercial scientific applications, and this dataset fills a significant gap in woody biomass equations and provides key parameters for biomass estimation in studies on terrestrial ecosystem structure and function.

1 Introduction

Shrubland is widely distributed in China and containing a variety of shrub species and types. As well as forest, shrubland 25 plays an important role in the global carbon balance (Conti et al., 2013; Piao et al., 2010). Shrubs are the dominant plants in shrubland ecosystems and the shrub layer in forest ecosystems. Accurate estimation of shrub biomass and carbon storage is the basis and necessary prerequisite for the study of carbon dynamics, assessment of carbon sink potential, and understanding the structure and function of shrubland and forest ecosystems under global warming and projected drying trend (Estornell et al., 2011; Ketterings et al., 2001). However, in terrestrial ecosystem carbon accounting studies, shrubland 30 biomass is identified as the most important uncertain component in terrestrial carbon balance, due to limited research data and relative poorer estimating methods (Hu et al., 2006; Piao et al., 2010). Similarly, the uncertainty in forest biomass

estimation is also partly attributed to insufficient shrub biomass equations in estimating the shrub layer biomass (Estornell et al., 2011; Lin et al., 2010).

35 Shrub biomass equations refer to quantitative relationships between biomass of the whole individual or different components (such as stems, branches, leaves and roots) and one or several dendrometric variables (such as shrub height, basal diameter, crown projection area, etc.) (Chojnacky et al., 2013; Lambert et al., 2005; Whittaker and Woodwell, 1968). In recent decades, biomass equations of certain shrub species have been developed primarily consulted the methods in tree biomass equation researches, by harvesting samples in the growing season, establishing the optimal biomass equation through variable selection, model selection and precision evaluation (Chave et al., 2014; Jenkins et al., 2003; Ludwig et al.,

40 1975). Representative researches mainly focused on following shrubs species, such as *Ostryopsis davidiana*, *Spiraea pubescens*, *Rosa xanthina*, *Caragana korshinskii*, *Hedysarum scoparium* and several shrub species of *Artemisia* (Chen et al., 2002; Li et al., 2014; Zhang et al., 1993; Zhang, 1989).

However, due to the diversity of shrub species and the wide geographic distribution, there are still severe problems in shrub biomass estimation: first, biomass equations are mostly limited to certain research areas and species (Hounzandji et al.,

45 2015; Muukkonen, 2007; Zeng et al., 2015); second, equation forms are diverse and lack a relatively unified method (Foroughbakhch et al., 2005; Picard et al., 2015); third, only a few studies verified the accuracy of the equations with independent field measured data (Dong et al., 2012; Kozak and Kozak, 2003). Therefore, it is necessary to develop a biomass equation dataset for common shrub species in China using a relatively unified method based on a large number of field measurements (Cifuentes Jara et al., 2015), after collecting and screening biomass equations from published studies.

50 Therefore, in this research, we firstly scrutinized literature on shrub biomass equations published in recent decades and collected shrub biomass equations of high quality. Afterwards, we developed biomass equations for the dominant species of shrublands using a unified method, based on a large number of field measured biomass data obtained from a national scale shrubland ecosystem investigation. Consequently, we constructed a comprehensive biomass equation dataset for common shrub species in China. The dataset covers broad geographical and climatic gradients, represents the common shrub

55 communities across China, and facilitates terrestrial ecosystem woody biomass estimation on large scales.

2 Materials and methods

2.1 Literature retrieval

We retrieved the available literature (journals, books and postgraduate's dissertations) between 1982 and 2019 (excluding Taiwan Province in this study) from several comprehensive online literature databases (China Knowledge Resource

60 Integrated Database, China Science and Technology Journal Database, Wanfang Data Knowledge Service Platform, and Web of Science), using the most relevant keywords (shrub, biomass, model, equation, allometry, allometric, etc.) with logical operators. No a priori criteria (shrub species, site condition or equation form, etc.) were applied in the literature survey.

2.2 Equation collection and screening

65 Using the following criteria, we critically scrutinized the collected literature to obtain reliable biomass equations.

2.2.1 Scope

Natural and planted shrublands and those formed due to human disturbance (deforestation, over cutting of arbor species and forest fire, etc.) were all investigated in this study. Biomass equations including both total biomass and biomass of different components were developed mainly for the dominant shrub species, but also include a few tree species with shrub-like
70 architecture (dwarfed by long term disturbance) (Zhang et al., 2013). In addition, biomass equations for understory shrub species in forest ecosystems were collected and compiled as well.

2.2.2 Measurement method

Procedures of field investigation and biomass measurement were in accordance with a robust and unified method proposed in a technical specification for field investigation and laboratory analysis of carbon sequestration in shrub ecosystem (Xie
75 and Tang, 2015), including plot setting, sample shrubs selection, morphological form classification and biomass (the oven-dried mass) measurements. Generally, plot areas were not smaller than $25m^2$, and at least ten sample shrubs were harvested and weighed to determine the biomass of each component (leaf, stem, branch and root, etc.). The division of shrub components can be summarized as shown in Fig. 1. Aboveground biomass mainly including three components (leaf, stem and branch), but in some cases, such as during the florescence and fruit period, flower and fruit were also included.
80 Belowground biomass was determined by a full excavation of the entire root system to avoid significant underestimation, although the loss of fine roots was always inevitable during excavation.

2.2.3 Equation building

Predictor variables were not limited, but equations should be developed with robust regressions, explicit equation forms (*e.g.*, power, linear, and quadratic functions) and validation evaluations. If the differences (<0.05) in goodness-of-fit of biomass
85 equations, such as coefficients of determination (R^2) were small (*e.g.*, $|\Delta R^2|\leq0.1$) among all equation forms, the priority order for selection was power, linear and quadratic equations. Higher-degree polynomial functions were excluded for the lack of biological significance (*i.e.* the representativeness of plant growth and development processes). Besides, equations developed based on larger sample size were preferred. In addition, equations developed with fewer and easy-to-measure predictor variables were selected with priority in this study.

90 ### 2.2.4 Quality checking

There are great differences between shrub biomass equations due to the large time ranges, different investigation methods, and various methods used in equation creation. In some studies artificial mistakes were involved, such as printing errors or

wrong records in figures and charts. Therefore, biomass equations being considered for inclusion were checked or corrected

with the following steps. First, if original data were available, dendrometric variables (*e.g.* height, basal diameter and crown

95   projection area) of sample shrubs were used to verify the biomass equations. Second, relative growth relationships between

different components and biomass allocations (the percentages of biomass allocated to leave, stem, branch and root, and the

root-shoot ratio) were important references (if they were in reasonable ranges, considering the divergences among species

and habitats) in equation collection.

2.3 Equation creation and evaluation

100   #### 2.3.1 Classification of shrub types

All shrubs were classified into three types according to the morphological characteristics (Xie and Tang, 2015). Shrubs with

explicit and dispersed branch structure were defined as "Type A" shrubs (*e.g. Cotinus coggygria*). Shrubs with implicit and

unkempt branch structure were defined as "Type B" shrubs (*e.g. Potentilla fruticosa*). Shrubs with implicit and clustered

branch structure were defined as "Type C" shrubs (*e.g. Sophora moorcroftiana*).

105   In general, for both "Type A" and "Type C" shrubs, biomass could be estimated using biomass equations developed

based on the law of allometric growth. However, for "Type B" shrubs the dendrometric variables were hard to be accurately

measured, thus biomass was measured by destructive harvesting and weighing. Therefore, in this study we focused on

developing biomass equations for "Type A" and "Type C" shrubs.

2.3.2 Equation creation

110   For "Type A" shrubs, a compound variable D^2H (D, basal diameter in cm; H, shrub height in m) was used as the predictor,

while for "Type C" shrubs, A_c (crown projection area, $A_c = \pi （L_1×L_2）/4$, L_1 and L_2 are the longest axis of shrub crown and

the shorter axis perpendicular to it respectively, both in m) or V_c (crown projection volume, $V_c = A_c×H$) was used.

Power equation was preferred for its interpretation of the natural law, i.e. the allometric growth relationships between

related variables in plant growth and development process. The growth relationship can be expressed as Eq. (1):

115   $$Y = aX^b \qquad (1)$$

Generally, its linear form was more commonly used through natural logarithmic transformation, Eq. (2) (Baskerville, 1972):

$$\ln Y = \ln a + b \ln X \qquad (2)$$

Y is the dry weight of different shrub components to be estimated; X is the corresponding predictor; ln denotes natural

logarithm (base e); a is the constant in regression equation; b is the scaling coefficient of relative growth relationship.

120   In cases the goodness-of-fit or prediction accuracy of power equation was less effective to meet the requirements,

simple linear equation was fitted, Eq. (3):

$$Y = a + bX \qquad (3)$$

Y is the dry weight of different shrub components to be estimated; X is the corresponding predictor; a and b are the

125 intercept and slope in regression equation respectively.

Biomass equations were fitted with linear regression analysis in R statistical software (R version 3.3.0), using the ordinary least squares (OLS) method. For power equations, a standard error correction factor (cf) was applied (Snowdon, 1991; Sprugel, 1983).

2.3.3 Equation evaluation

130 Equation evaluation includes both the analysis of goodness-of-fit and the accuracy in future prediction. The regression equations and regression coefficients were tested for significance at first. Statistical parameters including adjusted-R^2 (R^2) or fitness index (FI) were used in the goodness-of-fit evaluation. A simple linear regression between predicted and field measured value was fitted without intercept, and the regression slope (b), R^2 and the relative error (RE) were used in evaluating the prediction accuracy (Table 1).

135 In this study, we split the data into two parts (Picard and Cook, 1984), 10% of the sample shrubs of each species were randomly sampled and used as independent test dataset, and the remaining 90% samples were used for equation creation and evaluation. With the resampling methods (bootstrap and cross-validation) samples for equation creation and evaluation were randomly allocated into two groups, and the sampling test iterated 1000 times for each shrub species. 75% of the samples were used for fitting biomass equation and obtaining relevant parameters of goodness-of-fit, and the remaining 25% samples

140 were used for analyzing the prediction accuracy.

Equation coefficients, parameters of the goodness-of-fit and prediction accuracy were the mean values of corresponding results in 1000 times random test. The optimal biomass equation was selected through a comprehensive analysis of the goodness-of-fit and prediction accuracy. Finally, the independent test dataset was used to test the accuracy of the species-specific biomass equations in future prediction.

145 **3 Results**

From 42 references (see full list in https://doi.org/10.11922/sciencedb.00641, Wang et al., 2021) during the period 1989-2019, we compiled 373 biomass equations specific to 125 shrub species. Through a national scale shrubland ecosystem investigation, 449 biomass equations specific to 87 shrub species were developed with field measured data. Consequently, a dataset consist of 822 biomass equations specific to 167 common shrub species in China were constructed and the detailed

150 variables and their descriptions in the dataset are summarized in Table 2. Temporal changes in the number of shrub biomass equations showed a continuously increasing trend, especially during 2011 to 2019. Equations developed in this period contributed 77.62% of the total equations. These studies were carried out in 851 sites (Fig. 2), showing broad geographical coverage (18.2-51.6°N, 79.8-132.8°E and -175-5220 m in altitude) across China and broad climatic ranges (-2.9-25.7 ℃ in MAT and 8.4-2156.6 mm in MAP).

155   These compiled studies and equations varied greatly with shrubland types, stem forms and shrub species (https://doi.org/10.11922/sciencedb.00641, Wang et al., 2021). The studied shrublands were categorized into five types: deciduous broadleaved shrubland, evergreen broadleaved shrubland, evergreen coniferous shrubland, open shrubland and the understory shrub layer in forest ecosystem (Fig. 3). It should be noted that, for easy retrieval and utilization of biomass equations in this dataset, the understory shrub layer was informally categorized into a group of shrublands. Among the five

160   types, deciduous broadleaved shrubland had the most equations (71.5% of the total equations), followed by evergreen broadleaved shrubland (14.4%), open shrubland (6.9%), understory shrub layer (6.3%) and evergreen coniferous shrubland (0.9%). The ten most commonly studied species contributed 18.2% of biomass equations and six of them were dominant species in xeromorphic shrublands.

        In previous studies root biomass was not always measured, therefore, equations for root were relatively few compared

165   with the aboveground sector. Equations for total biomass, aboveground biomass, stem biomass, current-year branch biomass, leaf biomass and root biomass, accounted for 22.3%, 21.2%, 17.2%, 6.2%, 15.5%, and 17.0% of the total 822 equations respectively (Fig. 4). However, only 0.7% of the equations were for other shrub organs, such as stem bark and fruit.

        Of the 822 equations, 76.5% were based on the commonly used compound variables (D^2H, 48.1%; V_c, 18.6%; A_c, 9.8%) and only 2.7% were based on single predictors (basal diameter, diameter at 10cm above the ground and shrub height) (Fig.

170   5a). In total, 41 equation forms were applied in describing the quantitative relationships between shrub biomass and dendrometric variables (https://doi.org/10.11922/sciencedb.00641), which were categorized into four types: power equation, linear equation, quadratic polynomial equation and exponential equation (Fig. 5b). Power equations were the most frequently used type (529 equations, accounting for 64.4% of the equations), followed by the linear equations (260, 31.6%), quadratic polynomial equations (26, 3.2%) and exponential equations (7, 0.9%).

175   A small proportion (4.4%) of the total 822 equations did not specify the sample size (*i.e.*, the number of sample shrubs used in developing biomass equations). The sample size varied from 5 and 312 shrubs, where the most common sample sizes were between 9 and 40 shrubs, accounting for 79.3% of the 786 equations with specified sample sizes. For the applicable ranges of equations, 724 out of the 822 equations had clear applicable ranges in this dataset. From the 680 equations with available shrub height ranges, the height varied between 3.0 and 600.0 cm, and 22.6% of the equations had maximums shrub

180   height lower than 100.0 cm. From the 481 equations with available basal diameter ranges, the basal diameter varied between 0.1 and 21.2 cm, and 8.1% of the equations had maximums basal diameter lower than 1.0 cm.

        The goodness-of-fit parameters such as R^2 and FI are relatively high that can meet the requirement. R^2 range from 0.13 to 0.99 with a mean valve of 0.82, and 68.3% of the equations have a R^2 larger than 0.8. FI range from 0.13 to 0.99 with a mean valve of 0.65, and 26.9% of the equations have a FI larger than 0.8.

185   Not all studies provided parameters of precision evaluation, thus in this dataset a small portion of the equations were not evaluated with the slope, R^2 or RE of simple linear regression between predicted and measured data. In this dataset slope ranges from 0.63 to 1.27 with a mean valve of 0.96, and 79.7% of the equations have a slope between 0.9 to 1.1. R^2 ranges from 0.47 to 0.99 with a mean valve of 0.85, and 80.2% of the equations have a R^2 larger than 0.8. RE ranges from -43.6% to

Earth System
Science
Data

Open Access / Discussions

190   23.2% with a mean value of -4.1%; moreover, 47.7% of the equations have a RE within ±5% in future prediction, and 73.9% of the equations have a RE within ±10%.

4 Data availability

This version of biomass equation dataset for common shrub species in China was developed with two steps; firstly, equations were collected from published studies during 1989 to 2019; secondly, equations were created through detailed field survey of 738 shrub sites across China. The dataset is freely available at https://doi.org/10.11922/sciencedb.00641

195   (Wang et al., 2021) for non-commercial scientific applications, but the free availability of the dataset does not constitute permission to reproduce or publish it.

5 Conclusion

In this study, we developed the first biomass equation dataset for common shrub species in China. This dataset contains comprehensive background information and covers broad geographical, climatic and shrub vegetation gradients, and

200   moreover, represents a significant expansion and supplement to the woody biomass equation datasets such as the biomass equation datasets for China's tree species (Luo et al., 2020), and thus fills an important gap in woody biomass estimation and terrestrial ecosystem carbon budget.

References

Baskerville, G. L.: Use of logarithmic regression in the estimation of plant biomass, Can J Forest Res, 2, 49-53, 1972.

205   Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C., Henry, M., Martínez-Yrízar, A., Mugasha, W. A., Muller-Landau, H. C., Mencuccini, M., Nelson, B. W., Ngomanda, A., Nogueira, E. M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C. M., Saldarriaga, J. G., and Vieilledent, G.: Improved allometric models to estimate the aboveground biomass of tropical trees, Global Change Biol, 20, 3177-3190, 2014.

210   Chen, X., Ma, Q., Kang, F., Cao, W., Zhang, G., and Feng, Z.: Studies on the biomass and productivity of typical shrubs in Taiyue Mountain , Shanxi Province, Forest Research, 15, 304-309, 2002 (in Chinese).

Chojnacky, D. C., Heath, L. S., and Jenkins, J. C.: Updated generalized biomass equations for North American tree species, Forestry, 0, 1-23, 2013.

Cifuentes Jara, M., Henry, M., Réjou-Méchain, M., Wayson, C., Zapata-Cuartas, M., Piotto, D., Alice Guier, F., Castañeda

215   Lombis, H., Castellanos López, E., Cuenca Lara, R., Cueva Rojas, K., Del Águila Pasquel, J., Duque Montoya, Á., Fernández Vega, J., Jiménez Galo, A., López, O., Marklund, L., Michel Fuentes, J., Milla, F., Návar Chaidez, J., Ortiz

Malavassi, E., Pérez, J., Ramírez Zea, C., Rangel García, L., Rubilar Pons, R., Saint-André, L., Sanquetta, C., Scott, C., and Westfall, J.: Guidelines for documenting and reporting tree allometric equations, Ann For Sci, 72, 763-768, 2015.

220   Conti, G., Enrico, L., Casanoves, F., and Díaz, S.: Shrub biomass estimation in the semiarid Chaco forest: a contribution to the quantification of an underrated carbon stock, Ann For Sci, 70, 515-524, 2013.

Dong, D., Li, X., Wan, H., and Lin, H.: Aboveground biomass estimation of *Tamarix ramosissima* shrub in the lower reaches of Tarim River, Acta Botanica Boreali-Occidentalia Sinica, 32, 384-390, 2012 (in Chinese).

Estornell, J., Ruiz, L. A., Velázquez-Martí, B., and Fernández-Sarría, A.: Estimation of shrub biomass by airborne LiDAR data in small forest stands, Forest Ecol Manag, 262, 1697-1703, 2011.

225   Foroughbakhch, R., Reyes, G., Alvarado-Vázquez, M. A., Hernández-Piñero, J., and Rocha-Estrada, A.: Use of quantitative methods to determine leaf biomass on 15 woody shrub species in northeastern Mexico, Forest Ecol Manag, 216, 359-366, 2005.

Hounzandji, A., Jonard, M., Nys, C., Saint-André, L., and Ponette, Q.: Improving the robustness of biomass functions: from empirical to functional approaches, Ann For Sci, 72, 795-810, 2015.

230   Hu, H., Wang, Z., Liu, G., and Fu, B.: Vegetation carbon storage of major shrublands in China, Journal of Plant Ecology, 30, 539-544, 2006 (in Chinese).

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., and Birdsey, R. A.: National-scale biomass estimators for United States tree species, Forest Sci, 49, 12-35, 2003.

Ketterings, Q. M., Coe, R., van Noordwijk, M., Ambagau', Y., and Palm, C. A.: Reducing uncertainty in the use of allometric

235   biomass equations for predicting above-ground tree biomass in mixed secondary forests, Forest Ecol Manag, 146, 199-209, 2001.

Kozak, A. and Kozak, R.: Does cross validation provide additional information in the evaluation of regression models?, Can J Forest Res, 33, 976-987, 2003.

Lambert, M. C., Ung, C. H., and Raulier, F.: Canadian national tree aboveground biomass equations, Can J Forest Res, 35,

240   1996-2018, 2005.

Li, G., Zhao, X., and Liu, B.: The construction of predicting model of aboveground biomass of four types of shrubs in northern Shanxi Province, Forest Resources Management, 1, 71-76, 2014 (in Chinese).

Lin, W., Li, J., Zheng, B., Guo, J., and Hu, L.: Models for estimating biomass of twelve shrub species in Jinggang Mountain nature reserve, Journal of Wuhan Botanical Research, 28, 725-729, 2010 (in Chinese).

245   Ludwig, J. A., Reynolds, J. F., and Whitson, P. D.: Size-biomass Relationships of Several Chihuahuan Desert Shrubs, Am Midl Nat, 94, 451-461, 1975.

Luo, Y., Wang, X., Ouyang, Z., Lu, F., Feng, L., and Tao, J.: A review of biomass equations for China's tree species, Earth Syst. Sci. Data, 12, 21-40, 2020.

Muukkonen, P.: Generalized allometric volume and biomass equations for some tree species in Europe, EUR J FOREST

250   RES, 126, 157-166, 2007.

Piao, S., Fang, J., and Huang, Y.: The carbon balance of terrestrial ecosystems in China, Chinese Basic Science, 2, 20-23, 2010 (in Chinese).

Picard, N., Rutishauser, E., Ploton, P., Ngomanda, A., and Henry, M.: Should tree biomass allometry be restricted to power models?, Forest Ecol Manag, 353, 156-163, 2015.

255 Picard, R. R. and Cook, R. D.: Cross-validation of regression models, J Am Stat Assoc, 79, 575-583, 1984.

Snowdon, P.: A ratio estimator for bias correction in logarithmic regressions, Can J Forest Res, 21, 720-724, 1991.

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

Wang, Y., Xu, W., Tang, Z., Xie, Z.: A biomass equation dataset for common shrub species in China [Dataset], sciencedb, https://doi.org/ 10.11922/sciencedb.00641, 2021.

260 Whittaker, R. H. and Woodwell, G. M.: Dimension and production relations of trees and shrubs in the Brookhaven forest, New York, J Ecol, 56, 1-25, 1968.

Xie, Z. and Tang, Z.: Technical specification for field investigation and laboratory analysis of carbon sequestration in shrub ecosystem. // Technical manual writing group of ecosystem carbon sequestration project, Observation and investigation for carbon sequestration in terrestrial ecosystems., Science Press, Beijing, 145-191, 2015 (in Chinese).

265 Zeng, W., Bai, J., Song, L., Zhao, X., Wang, X., Xing, L., and Zhang, Z.: Establishment of individual biomass equations for *Caragana korshinkii* and *Armeniaca sibirica* in Inner Mongolia, Forest Research, 28, 311-316, 2015 (in Chinese).

Zhang, F., Shangguan, T., and Li, S.: Improvement on the modelling method of biomass of brush, Chinese Journal of Ecology, 12, 67-69, 1993 (in Chinese).

Zhang, S.: Selection of pre-estimating model for the aboveground biomass of artificial *Caragana Korshinskii* shrub, Journal 270 of Desert Research, 9, 52-61, 1989 (in Chinese).

Zhang, Y., Ouyang, X., Li, Y., Liu, S., Zhang, D., and Zhou, G.: Shrub community characteristics and quantitative calculation of theirs biomass in southern China, Journal of Central South University of Forestry and Technology, 33, 71-79, 2013 (in Chinese).
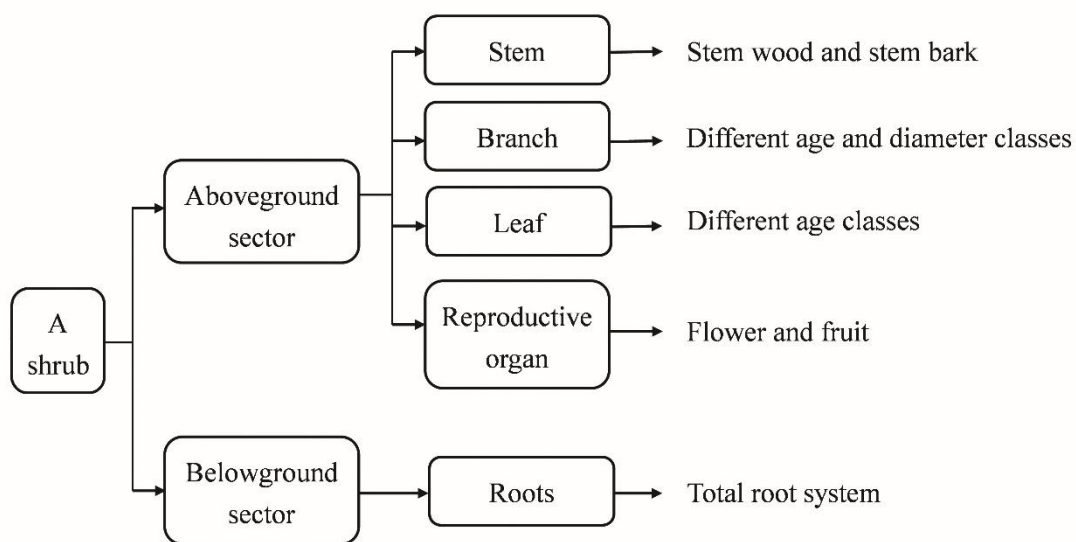
Author contributions.

ZX, WX and ZT originated, conceived and designed the work. YW, WX and ZX developed and analysed the equation dataset. All authors contributed to the writing of the manuscript.
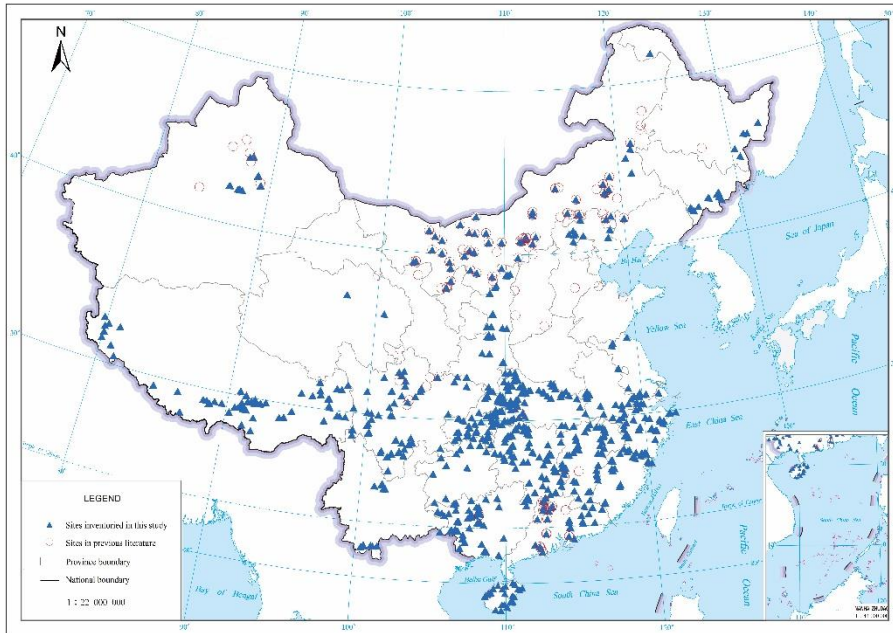
Competing interests.

The authors declare that they have no conflict of interest.

Figure 1: The division of shrub components. A shrub can be divided into aboveground and belowground sector; aboveground sector is often subdivided into stem, branch, leaf and reproductive organ.

290

Figure 2: Spatial distribution of study sites. Circle dots denote study sites in previous literature; solid dots denote shrubland sites
295 **inventoried in this study.**
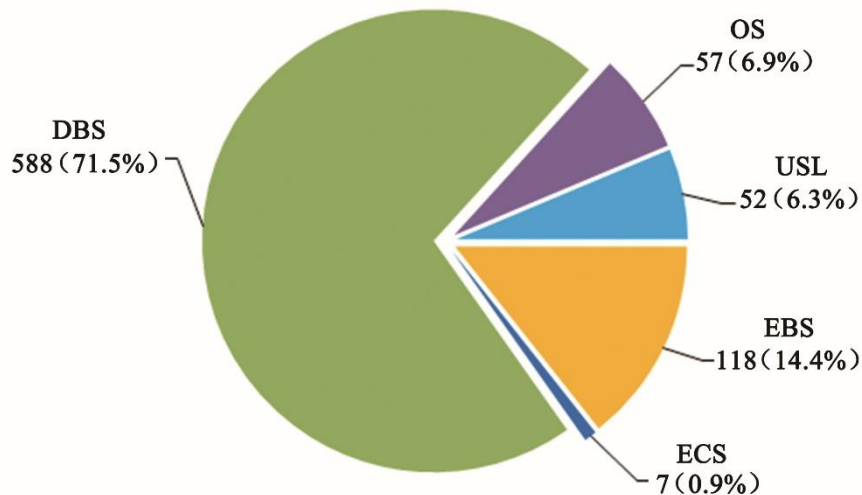
Figure 3: Distribution of biomass equations by shrubland type. Shrublands are categorized into deciduous broadleaved shrubland (DBS), evergreen broadleaved shrubland (EBS), evergreen coniferous shrubland (ECS), open shrubland (OS) and understory shrub layer (USL).
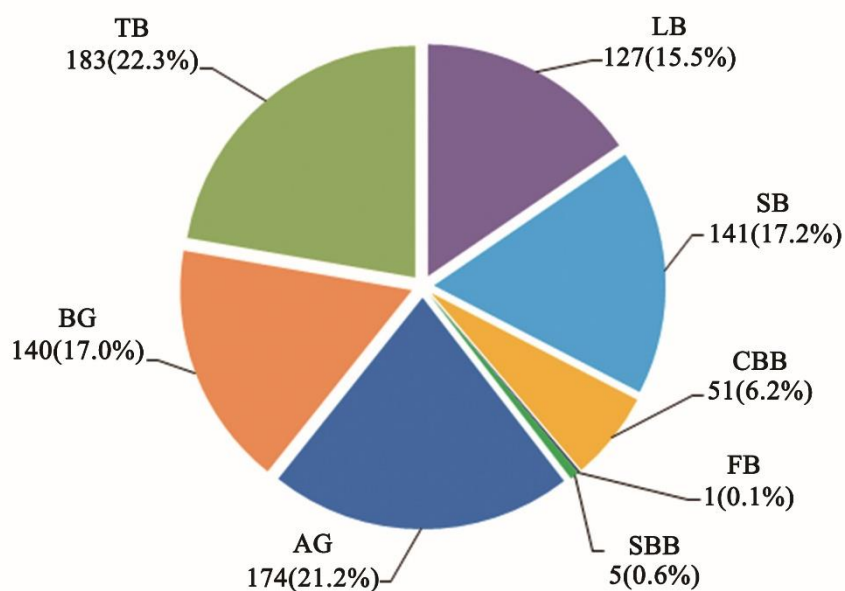
Figure 4: Distribution of biomass equations by biomass component. Leaf biomass (LB), stem biomass (SB), current-year branch biomass (CBB), fruit biomass (FB), stem bark biomass (SBB), aboveground biomass (AGB), belowground biomass (BGB) and total biomass (TB).
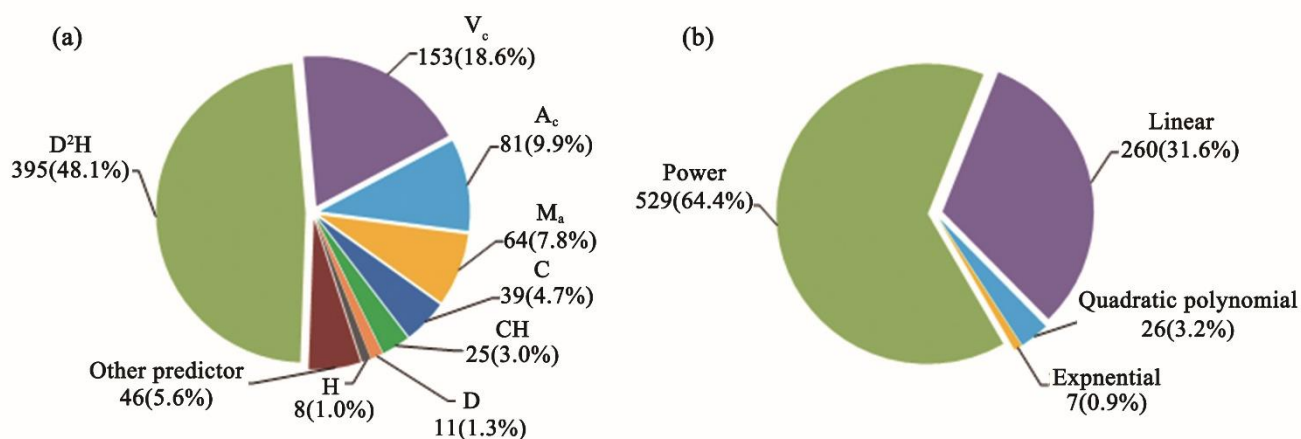
305

Figure 5: Distribution of biomass equations by (a) predictor variable and (b) equation form. D is basal diameter, H is shrub height, C is mean value of the crown diameter, A_c is crown projection area, V_c is crown projection volume, M_a is aboveground biomass, other predictors include some uncommonly used variables. Equation forms are categorized into power equation, linear equation, quadratic polynomial equation and exponential equation.

Earth System
Science
Data

Open Access

Discussions

Table 1: Statistics used in evaluating the goodness-of-fit and prediction precision of shrub biomass equations.

Statistics	Calculation formulas
R^2 in power equation	$R^2_{(P)} = 1 - \sum(\ln Y_i - \ln \acute{Y}_i)^2 / \sum(\ln Y_i - \ln \bar{Y})^2$
Fitness Index (FI)	$FI = 1 - \sum(Y_i - \acute{Y}_i)^2 / \sum(Y_i - \bar{Y})^2$
Standard Error of Estimate (SEE)	$SEE = [\sum(\ln Y_i - \ln \bar{Y}_i)^2 / (n-2)]^{1/2}$
Correction factor (cf)	$cf = \exp（SEE^2 / 2）$
R^2 in linear equation	$R^2_{(L)} = 1 - \sum(Y_i - \acute{Y}_i)^2 / \sum(Y_i - \bar{Y})^2$
adjusted-R^2	$adj\text{-}R^2_{(P/L)} = R^2_{(P/L)} - (1 - R^2_{(P/L)}) \times (p/n - p - 1)$
Root Mean Squared Error (RMSE)	$RMSE = (\sum(Y_i - \acute{Y}_i)^2 / n)^{1/2}$
Relative Error (RE)	$RE = \sum(\acute{Y}_i - Y_i) / \sum Y_i$

315 **Table 2: Summary of variable information in the dataset.**

Variable	Description	Data origin	Unit	Type
1. General sheet				
Shrub species	Dominant shrub species of a shrubland or dominant shrub species of the shrub layer of a forest.	Original studies	Unitless	String
Information number	Identification number of each set of equations for a shrub species.	Author defined	Unitless	String
Study site	Locality name of study site or name of the province where field inventory data were obtained.	Original studies	Unitless	String
Shrub density	The number of shrubs per unit area. Shrub density is given as the mean values or ranges.	Original studies	shrubs/ha	String
Equation number	Identification number of each equation.	Author defined	Unitless	String
Height range	Applicable ranges of shrub height.	Original studies	cm	String
Crown diameter range	Applicable ranges of crown diameter.	Original studies	cm	String
Crown projection area range	Applicable ranges of shrub crown projection area.	Original studies	m^2	String
Basal diameter range	Applicable ranges of basal diameter.	Original studies	cm	String
Biomass range	Applicable ranges of total [aboveground] biomass.	Original studies	g	String
Shrub type	Shrub community characterized by the same leaf habit, leaf shape, or physiognomy.	Original studies	Unitless	String
Stem form	Shrubs are classified by stem form into Type a and Type c (Type a: shrubs with explicit branch structure and disperse and countable stems; Type c: shrubs with implicit branch structure and densely tufted stems)	Original studies	Unitless	String
Sources	Source of the data.	Original studies	Unitless	String

Table 2: Continued.

Variable	Description	Data origin	Unit	Type
2. Equation sheet				
Shrub species	Dominant shrub species of a shrubland or dominant shrub species of the shrub layer of a forest.	Original studies	Unitless	String
Equation number	Identification number of each equation.	Author defined	Unitless	String
Shrub component	A tree component divided in a certain way.	Original studies	Unitless	String
Predictor variable	One or more dendrometric variables, i.e., shrub height in cm, crown diameter in cm, crown projection area in m^2, basal diameter (diameter at 5 cm above soil surface) in cm, and component biomass in g.	Original studies	cm; m^2; g	String
Equation form	Quantitative relationship between the biomass (M in g) and one or more predictor variables.	Original studies	Unitless	String
Coeff. a	Equation coefficients.	Original studies	Unitless	Float
Coeff. b	Equation coefficients.	Original studies	Unitless	Float
Coeff. c	Equation coefficients.	Original studies	Unitless	Float
Coeff. d	Equation coefficients.	Original studies	Unitless	Float
CF	Goodness-of-fit statistics.	Original studies	Unitless	Float
n	Number of shrub samples used in equation creation.	Original studies	Unitless	Integer or String
$R^2[R]$	Goodness-of-fit statistics.	Original studies	Unitless	Float
FI	Goodness-of-fit statistics.	Original studies	Unitless	Float
n	Number of shrub samples used in equation evaluation.	Original studies	Unitless	Integer or String
b	Equation evaluation statistics, slope of linear regression between predicted and field measured values.	Original studies	Unitless	Float
$R^2[R]$	Equation evaluation statistics.	Original studies	Unitless	Float
RE	Equation evaluation statistics.	Original studies	Unitless	Float