Remote Sensing of Environment 232 (2019) 111341

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Estimating aboveground biomass in subtropical forests of China by integrating multisource remote sensing and ground data

Rong Zhang[a,b], Xuhui Zhou[a,b,c,*], Zutao Ouyang[d,e], Valerio Avitabile[f], Jiaguo Qi[d], Jiquan Chen[d], Vincenzo Giannico[g]

[a] Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, 200241, China
[b] Center for Global Change and Ecological Forecasting, East China Normal University, Shanghai 200062, China
[c] Shanghai Institute of Pollution Control and Ecological Security, 1515 North Zhongshan Rd, Shanghai 200437, China
[d] Center for Global Change & Earth Observations, Michigan State University, 218 Manly Miles Building, 1405 S. Harrison Road, East Lansing, MI 48823, USA
[e] Department of Earth System Science, Stanford University, Stanford, CA 94305, USA
[f] European Commission, Joint Research Centre, Via E. Fermi 2749, 21027 Ispra, Italy
[g] Department of Agricultural and Environmental Sciences, University of Bari "A. Moro", Bari, Italy, Via Amendola 165/A, 70126 Bari, Italy

ARTICLE INFO

Edited by Jing M. Chen

Keywords:
Aboveground biomass
Subtropical forest
Carbon stock
Remote sensing
Ground data

ABSTRACT

The biomass of the subtropical forests of China is an important component of the global carbon cycle. Recently, several above ground biomass (AGB) maps have been produced using a variety of approaches to assess the carbon stock of the subtropical forest in China. However, due to the lack of reliable ground observations and the limitations of AGB mapping methods at regional scales, estimates of the spatial distribution of AGB vary greatly, leading to large uncertainties in the carbon stock estimations. In this study, we produced a new 1-km spatial resolution AGB map by synthesizing an unprecedented number of ground AGB observations from published studies, and developed an AGB mapping method using a combination of ground observations, MODIS data, forest cover/gain/loss maps based on Landsat, GLAS forest canopy height, and climatic and terrain data. In addition, we validated our estimates using independent testing data and compared our estimates with three previous AGB maps. The results indicate that the total AGB stock in the subtropical forest of China is $(266 \pm 9.1) \times 10^6$ Mg, with an average AGB of 123.2 Mg/ha. Based on sixteen explanatory variables, our ensemble mean model explains 75% of the variance in forest AGB, with an RMSE of 45.5 Mg/ha. Comparison using all observation data shows that our map has a significantly lower RMSE and bias than previous maps, where the RMSE and bias tended to vary with forest type. This study not only improved the accuracy of AGB estimation for the subtropical forests but also highlighted the importance of forest type for regional AGB estimation.

1. Introduction

The above ground biomass (AGB) of forests is an important proxy for productivity, carbon stock and carbon sequestration strength in forest ecosystems and represents a key parameter for estimating carbon emissions and removals caused by land use and climate change (Baccini et al., 2017; Beer et al., 2010; Chen et al., 2014; Houghton et al., 2012; Le Toan et al., 2011). Therefore, accurate estimation of forest AGB is critical in assessing the carbon budget of terrestrial biomes (Houghton et al., 2009; Keith et al., 2009; Zarin et al., 2016). Currently, ground measurements are still the most accurate method for forest biomass data collection (Chave et al., 2015; Lu et al., 2016). However, acquiring AGB data over large areas is difficult because collecting ground measurements is a slow and laborious process. Thus, it is impossible to acquire a regional distribution of AGB using only ground measurements. An alternative method for regional AGB estimates is to integrate ground observations with remotely sensed satellite imagery (Timothy et al., 2016).

Remote sensing data have been widely utilized to map AGB, using the spectral signatures derived from satellite imagery to estimate parameters, such as vegetation indices, canopy cover, texture, shadow fraction, and leaf area, that have high correlations with AGB (Baccini et al., 2012; Badreldin and Sanchez-Azofeifa, 2015; Bouvet et al., 2018; Muukkonen and Heiskanen, 2007; Powell et al., 2010; Santoro et al., 2015). Images acquired by optical sensors have been widely applied to estimate biomass in different contexts because these data are freely

* Corresponding author at: School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200062, China.
E-mail address: xhzhou@des.ecnu.edu.cn (X. Zhou).

https://doi.org/10.1016/j.rse.2019.111341
Received 8 May 2018; Received in revised form 18 July 2019; Accepted 19 July 2019
Available online 29 July 2019
0034-4257/

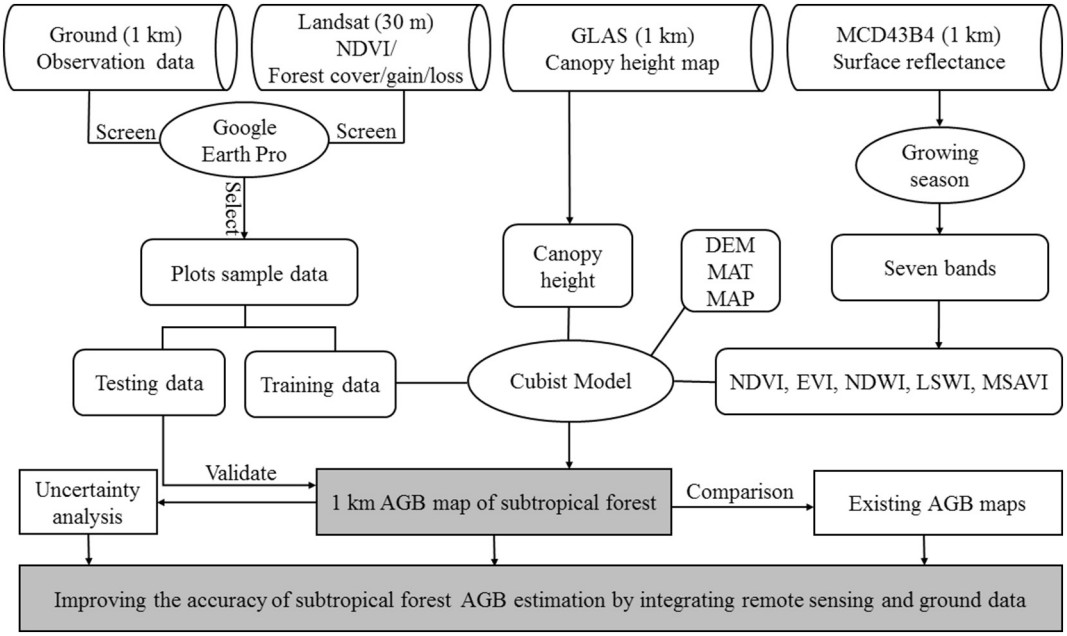

Fig. 1. A schematic diagram illustrating the methods for AGB estimation by integrating multisource remote sensing and ground data.

available and have various spatial, spectral and temporal resolutions (Blackard et al., 2008; Foody et al., 2003; Yan et al., 2015). For example, medium resolution data such as Landsat data are typically employed for AGB estimations at local scales (Hall et al., 2006; Powell et al., 2010; Zheng et al., 2004), whereas moderate and coarse resolution data (e.g., Moderate Resolution Imaging Spectra radiometer, MODIS) are used to estimate biomass at regional scales (Rodríguez-Veiga et al., 2016; Yan et al., 2015). Synthetic aperture radar data (e.g., the Advanced Land Observing Satellite (ALOS) Phased Array L-Band Synthetic Aperture Radar (PALSAR)) have also been used for estimating AGB at regional scales with medium spatial resolution (Cartus et al., 2014; Yu and Saatchi, 2016). However, both optical and radar data suffer from signal saturation at high AGB, limiting their ability to estimate AGB in dense tropical and subtropical forests (Lu et al., 2016; Zhao et al., 2016). Recently, LIDAR remote sensing data from spaceborne or airborne sensors have been used for AGB estimations in dense forests due to their sensitivity to vertical canopy structures and high AGB density (Asner et al., 2013; Baccini and Asner, 2013; Giannico et al., 2016; Lefsky et al., 2002). LIDAR data are typically acquired through a sampling approach and are often integrated with optical imagery for regional wall-to-wall AGB estimation. For example, Xu et al. (2017) mapped the AGB of the Democratic Republic of Congo by integrating Landsat, ALOS PALSAR, and SRTM data with airborne LIDAR data. Wang et al. (2016) mapped forest canopy heights globally via combined MODIS and GLAS data. Zarin et al. (2016) used a pantropical biomass map based on Landsat and GLAS to estimate carbon emissions due to tropical deforestation.

Although great progress has been made to improve spatially explicit AGB estimates by utilizing various remote sensing data, regional estimates of AGB still present large uncertainties at local scales (Mitchard et al., 2014; Su et al., 2016; Timothy et al., 2016). This is mainly because forests are very diverse (e.g., by type, climate, and site conditions) over a large spatial scale, i.e., a model trained with limited ground observation data can easily become over-fitted and fail to capture local features. The path forward to reducing uncertainties in AGB estimates therefore involves the acquisition of a sufficient amount of ground observations to capture AGB variability across a landscape where optimal integration of multiple remote sensing data can provide abundant spectral signals for diverse local features (Avitabile et al., 2011; Saatchi et al., 2011).

The subtropical forest area in China is one of the regions in the world with high uncertainty of AGB estimates (Santoro et al., 2015; Yu et al., 2014). Several AGB maps for this area have been recently produced at different spatial scales and resolutions using a variety of empirical modeling approaches based on remote sensing data calibrated by ground observations. For example, Piao et al. (2005) estimated carbon storage in China's forest based on the 8-km resolution of NOAA/AVHRR GIMMS and national forest inventory data. Du et al. (2014) mapped the forest biomass using the MODIS Land Cover Type product (aggregated to ~5.6-km resolution) for the whole of China. Saatchi et al. (2011) and Su et al. (2016) mapped AGB in global tropical and subtropical areas and China, respectively, through combined spaceborne LIDAR and optical imagery (1-km resolution). Avitabile et al. (2016) integrated existing regional AGB maps using a fusion model into a 1-km AGB pantropical map. However, AGB estimates from all these studies differ substantially for the subtropical forests of China. For example, the average biomass density and carbon stocks for the subtropical forest area of China derived from these products differ greatly, which hinders our understanding of the role of subtropical forests in the global carbon cycle (Yu et al., 2014), highlighting the need for a more accurate AGB map for the subtropical forest area of China. Although previous studies have tried to improve biomass estimates using higher resolution satellite imagery or more complex predictive models, the lack of sufficient and reliable ground observations had limited model development and thus performance. A large number of ground AGB observations in this region were collected by the national forest inventories of China and local studies between 2000 and 2014 (Luo et al., 2012; Zhang et al., 2015), but previous studies did not include all these data. For example, in studies of China, Piao et al. (2005), Du et al. (2014), and Su et al. (2016) used ground observations only from 1984–1998, 2004–2008 and 2000–2008, respectively, whereas in global studies, ground observations of the subtropical forest area of China were not available (Avitabile et al., 2016; Timothy et al., 2016; Chave et al., 2015; Saatchi et al., 2011).

In this study, we developed a framework to integrate multisource remote sensing and ground observation data to improve the AGB estimates for the subtropical forests of China (Fig. 1). The objectives of this study are to: (1) compile a large number of ground AGB observations collected from publications to represent AGB variability over the landscape; (2) produce an 1-km spatial resolution AGB map for the

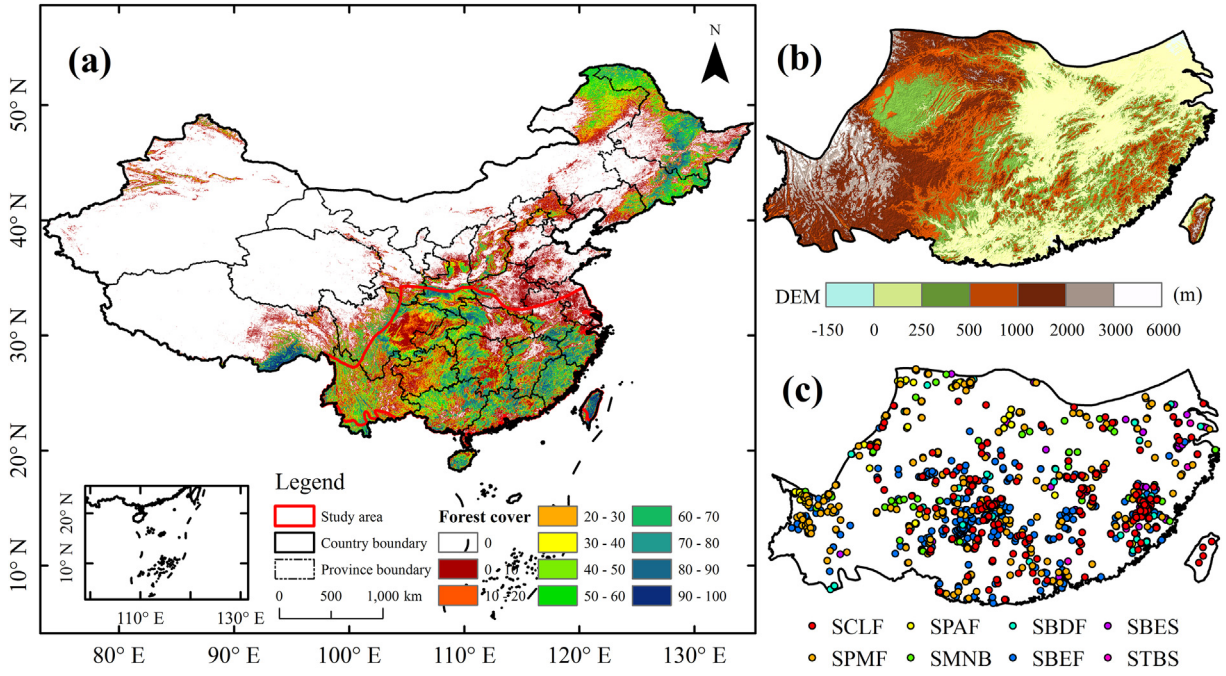

Fig. 2. Location of the study area and ground observation data sites. (a) 1-km resolution tree cover (%) map for 2000 based on Hansen et al. (2013); (b) Digital Elevation Model (DEM) of the study area at 1-km resolution; and (c) distribution of ground data for the eight forest types. Subtropical *Cunninghamia lanceolata* forest (SCLF), subtropical *Pinus massoniana* forest (SPMF), subtropical *Picea-Abies* forest (SPAF), subtropical mixed needleleaf and broadleaf forest (SMNB), subtropical broadleaf deciduous forest (SBDF), subtropical broadleaf evergreen forest (SBEF), subtropical broadleaf evergreen sclerphyllous forest (SBES), and subtropical and tropical bamboo and scrub forest (STBS).

subtropical forest area of China in 2000 using an advanced Cubist regression tree model that combines multiple remote sensing data (e.g., forest cover/gain/loss derived from Landsat, forest canopy height derived from GLAS data, and vegetation indices derived from MODIS) with a large ground dataset; (3) assess the reliability of our map by comparing with other AGB maps; and (4) validate our map and assess the uncertainty through cross-validation and bootstrapping.

2. Methods

2.1. Study area and existing AGB maps

We studied the subtropical forests of China located between latitude 21–35°N and longitude 97–123°E at an elevation ranging from −150 to 6000 m above sea level (Fig. 2a and b). The forest area is defined as all pixels with tree cover of > 10% based on the global tree cover dataset for the year 2000 from Hansen et al. (2013), which were then resampled to a resolution of 1 km to match MODIS and other spatial data. The climate of the study area is wetter than other places with similar latitudes on Earth (e.g., the Sahara Desert, the Sonoran Desert, and the North American Prairie) and gradually changes from marine climate in the east to continental climate in the west. The mean annual precipitation (MAP) is 1300 ± 300 mm, with 80% of precipitation falling in the growing season. The mean annual temperature (MAT) is 15.5 ± 3 °C, which normally declines from the south toward the north. The complex climate and geography have led to diverse forest communities. The northern part is primarily dominated by the mixed forests (e.g., mixed needleleaf and broadleaf forests, broadleaf deciduous and evergreen forests); the central part is mostly covered by the evergreen broadleaf forests with high biomass density; and the southern part is occupied by the broadleaf evergreen sclerophyllous forests.

In this study, we considered three existing AGB maps covering this area for the purpose of comparison: Saatchi et al. (2011), Avitabile et al. (2016) and Su et al. (2016). They are hereafter referred to as 'Saatchi', 'Avitabile', and 'Su', respectively. All of these maps were provided in a geographic projection (WGS-84) at an approximately 1-km (0.00833°) pixel size. These maps were developed based on similar input data layers (i.e., MODIS and GLAS), but presented different spatial AGB patterns for the subtropical forest of China (Fig. S1), partially because different ground datasets were used for calibration and partially due to the differences in modeling methodologies. The Avitabile map used the Saatchi map as an input layer, therefore, these two maps presented more similar spatial patterns than that of the Su.

2.2. Ground data collection and screening

The ground data used to train our AGB prediction model were synthesized from published literature and national forest inventories (Table S1). Altogether, AGB measurements of 1988 plots were collected from > 500 references. These plots covered eight main forest types (Fig. 2c and Table 1), including subtropical *Cunninghamia lanceolata* forest (SCLF), subtropical *Pinus massoniana* forest (SPMF), subtropical *Picea-Abies* forest (SPAF), subtropical mixed needleleaf and broadleaf forest (SMNB), subtropical broadleaf deciduous forest (SBDF), subtropical broadleaf evergreen forest (SBEF), subtropical broadleaf evergreen sclerphyllous forest (SBES), and subtropical and tropical bamboo and scrub forest (STBS). The detailed information for each plot and forest type are provided in Table S1. In this synthesized dataset, only 3.6% of AGB (Table S1) was estimated using destructive sampling, which is the most accurate approach for measuring ground AGB but is also time-consuming and labor-intensive (Houghton et al., 2009). Sixty-one percent of the ground AGB in our datasets was measured using the allometric models that calculate AGB based on linear or non-linear regression models from diameter at breast height (DBH), tree height, and woody density of the surveyed trees. The remaining ground AGB (35.4%) was estimated based on a conversion from the volume at the plot level to the AGB using a volume expansion factor, the average wood density, and the biomass expansion factor (Lu et al., 2016). The volume at the plot level was the sum of the volume of all individual trees, which were obtained from a volume table (Anonymous, 1987)

Table 1
Summary statistics of ground observed AGB and the number of plots selected after the screening process for eight forest type for the subtropical forests of China.

Forest type	AGB (Mg/ha)					Mean canopy height (m)	Collected plots	Selected plots
	Min	Max	Median	Mean	SD			
SCLF	6.5	691.5	97.8	130.1	105	16	498	240
SPMF	4.6	427.8	103.6	117.6	69.9	13.6	476	309
SPAF	2.9	562.5	123	164.5	105.1	17.7	63	45
SMNB	15.8	381.1	90.1	109.2	72.1	14.6	131	86
SBDF	4.7	397.7	77.5	101	84.3	14.8	135	95
SBEF	6.2	570.1	122.5	141.2	86.7	18.5	463	351
SBES	24.4	469.6	134.8	155.4	94.6	13.8	197	124
STBS	5.5	132.7	33.7	43.2	34	12.1	25	18
Total plots	2.9	691.5	106.4	129	88.7	15.7	1988	1268

that is based on DBH and tree height (Fang et al., 1998).

Quality control of the raw data was performed to remove unreliable observations. We first screened the collected data based on a previous method of Avitabile et al. (2016). Specifically, we only included data that satisfied all of the following criteria: (1) the ground plots were visited between the year 2000 and 2014; (2) all living trees with DBH of ≥ 2.0 cm were included in the measurement; (3) auxiliary information, including geographical coordinates, dominant tree species, forest type, and the methods used to calculate the plot AGB, were recorded.

To minimize the effects of forest cover change and errors due to the temporal mismatches between the ground measurements and satellite observations, ground observation data were further screened by comparing them with high resolution (30 m) Global Forest Change (2000–2014) maps derived from Landsat (Hansen et al., 2013) to verify that no forest cover changes occurred in the interim. Here, we used forest loss and gain imagery to determine whether significant forest gain or loss occurred from 2000 to 2014 (Fig. S2). When forest gain/loss was detected, the ground observation data were excluded. The forest cover, gain, and loss maps were downloaded from Google Earth Engine (Hansen Global Forest Change v1.2, 2000–2014). These exclusions were made because many field plots did not record the exact year of measurement but reported the sampling period (e.g. 2004–2008).

Finally, we considered the geolocation mismatch between ground and remote sensing data (i.e., spatial mismatch). Although satellite data have relatively accurate geolocations, most ground observation data used in the present study were collected from field campaigns that often recorded coarse geographic coordinates using portable GPS devices. This practice may cause a slight spatial shift of the recorded plot location, leading to a discrepancy between the recorded and true location. Additionally, most field sampling was based on a plot size of approximately 1000 m^2 (Supporting information 2), which is considerably smaller than the size of a MODIS pixel. To reduce the errors caused by geolocation errors and spatial mismatches between field plots and MODIS pixels, we have controlled our samples so that they are located in homogeneous MODIS pixels. A homogenous MODIS pixel is defined as the standard deviation of < 0.15 for 30-m resolution NDVI data for a 3×3 km area centered at that pixel. We selected the threshold of 0.15 after extensive visual interpretation and tests based on Google Earth's very high resolution images. After these screenings, we still found cases of multiple plots (two or more) that had the same GPS coordinates but different AGB estimates (Supporting information 2), which was probably due to inaccurate GPS readings, rounding of the coordinates, data entry mistakes, etc. For such cases of multiple ground observations within the same 1 km^2 pixel, we first identified and excluded the outliers by comparing with the AGB values of other observations of similar NDVI values, and then averaged the remaining data. The outliers were defined as the plots with AGB higher than the 80% quantile or lower than the 20% quantile of all other samples that had similar NDVI (i.e., all other samples with NDVI of < 0.15 deviated from the one under examination). We also visually compared each ground AGB data point with the high-resolution imagery from Google Earth Pro to determine whether there was a spatial mismatch. The suspicious data points with very high AGB (> 188.8 Mg/ha–the 80% quantile of all samples) that were located in nearly bare land (NDVI < 0.2), or with very low AGB (< 59.4 Mg/ha–the 20% quantile of all samples) but located in dense forest area (NDVI > 0.8) were excluded after comparisons, assuming that these situations were due to spatial mismatches. With the above quality controls, 1268 plots were included for this study (Table 1).

2.3. MODIS data and pre-processing

We used 1-km resolution MODIS reflectance product (MCD43B4.V005) to upscale the ground-based AGB measurements over the landscape and to generate the wall–to–wall AGB map. We acquired MODIS NBAR data during the growing season (April–October) of 2000; they were corrected for solar and view geometry and atmospheric attenuation, screened for cloud cover, and composited to a 16-day time interval (Ju et al., 2010; Schaaf et al., 2002). Several indices correlated with vegetation, water and soil characteristics were calculated based on spectral reflectance, including NDVI (Tucker, 1979) and EVI (Huete et al., 2002) that are correlated with leaf area index, NDWI (Gao, 1996; McFeeters, 1996; Xu, 2006) that is strongly related to plant water content, LSWI (Xiao et al., 2004) that is sensitive to land surface water content, and MSAVI (Qi et al., 1994) that reflects a portion of the soil background.

2.4. Explanatory variables from other data sources

Variables such as precipitation, temperature, elevation, and forest canopy height improved AGB estimates at the regional scale (Lu, 2006; Timothy et al., 2016). Based on our Generalized Additive Model (GAM) exploratory analysis, a significant correlation ($P < 0.001$) existed between ground observed AGB and AP (annual precipitation), MAT (mean annual temperature), DEM (digital elevation model), and forest canopy height (Fig. 3). Consequently, we included these explanatory variables in the AGB prediction model.

The AP, MAT and DEM dataset for 2000 was downloaded from the Data Center for Resources and Environmental Sciences, Chinese Academy of Science (RESDC, http://www.resdc.cn). The AP and MAT dataset with a spatial resolution of 500-m was developed based on climate observations from 1915 meteorological stations across China. The MAT dataset was also corrected based on the DEM dataset. AP, MAT and DEM were reproduced at a resolution of 1-km spatially using the bilinear interpolation to spatially match the MODIS and other datasets.

We used a 1-km resolution forest canopy height map produced using 2005 data from the GLAS sensor onboard the ICESat (Ice, Cloud, and land Elevation Satellite) (Simard et al., 2011). The canopy height map has been widely used for forest biomass mapping (Fayad et al., 2016; Su et al., 2016; Zhang et al., 2014).

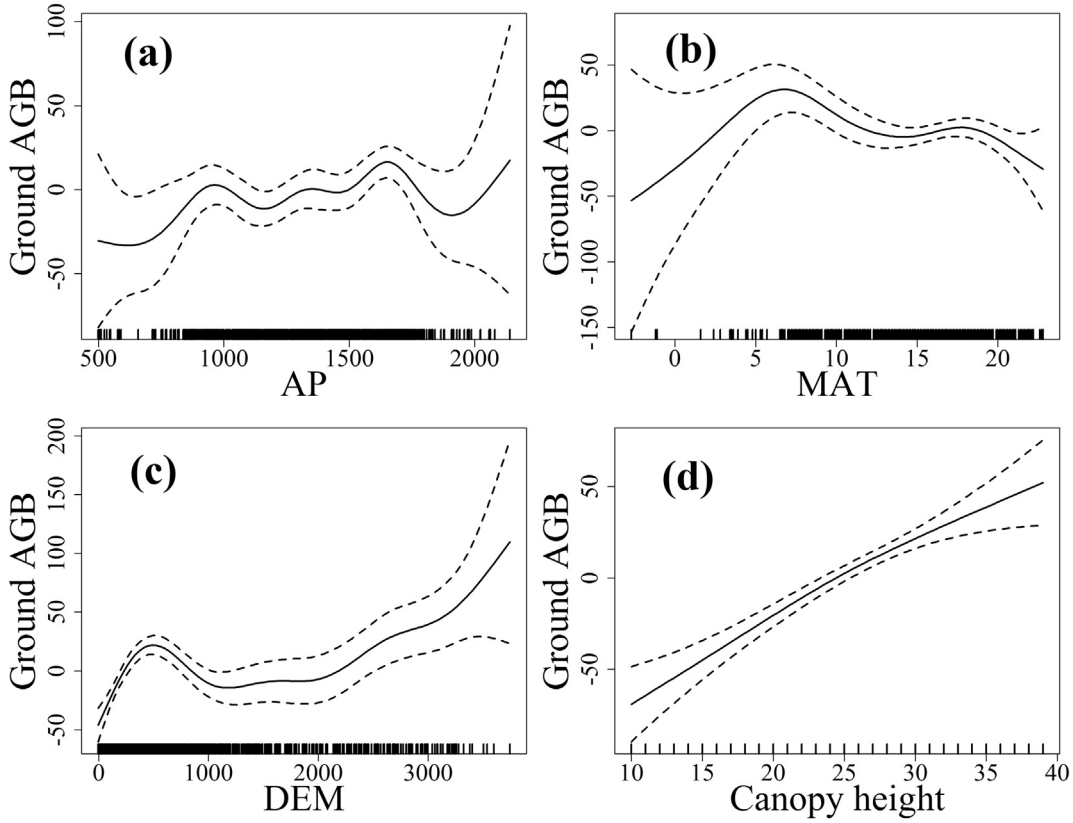

Fig. 3. Relationships between ground AGB (Mg/ha) and (a) AP (mm), (b) MAT (°C), (c) DEM (m), and (d) canopy height. All variables exhibit a significant relationship with ground AGB ($P < 0.001$). The short-dashed bars along the x-axis indicate sample site locations.

2.5. Regression tree modeling

We used an open source version of Cubist in the R package "Cubist" (Kuhn et al., 2018) to upscale the ground AGB to the regional scale, which is not only more effective than typical classification and regression trees (CART) but is also easier to understand than neural network models (Blackard et al., 2008; Gleason and Im, 2012; Walton, 2008). Cubist uses training data to produce a tree-based model that contains one or more rules, each rule with a set of conditions associated with a linear multivariate sub-model (Xiao et al., 2008, 2010). This feature differs from CART, which only contains a single variable at the end of each branch. Cubist allows multivariate linear models to overlap while not requiring them to be mutually exclusive (Gleason and Im, 2012). Cubist also differs from random forest method because predictions are produced using the linear regression model at the terminal rule of the tree but could be improved by combining the predictions from other similar training cases or nearest neighbor model results (Kuhn and Coulter, 2012; Quinlan, 1992). More importantly, Cubist has the ability to predict values beyond the range covered by the samples compared with the random forest method, which is important for forest biomass predictions because we do not know whether the lowest and highest biomass value in the study area had been sampled. Previously, Cubist has been successfully applied to scale-up net ecosystem exchange (NEE), gross primary production (GPP), and AGB data in various ecosystems (Xiao et al., 2008; Xiao et al., 2010; Ranjeet et al., 2018). Here, sixteen explanatory variables were used to train the Cubist models for AGB prediction, including the seven bands and five vegetation indices from MODIS, MAT, AP, DEM, and forest canopy height.

Bootstrapping, which constructs a new sample with the same sample size as the originals by randomly sampling with replacement was used to assist model development and validation. We bootstrapped 100 times to repeat the modeling process, with each Cubist model allowing 10 committee models (i.e., 10 regression trees). The 100 Cubist models were "bagged" in an ensemble, and the mean prediction derived from the 100 models (ensemble mean) is our final prediction. Altogether, 1000 regression trees were included in our ensemble model, which resembled the "bagging" process of a random forest algorithm.

2.6. Model evaluation

We used several statistical measures to quantify the performance of the model, including the coefficient of determination (R^2, Eq. (1)), the root mean square error (RMSE, Eq. (2)), and bias (Eq. (3)).

$$R^2 = 1 - \frac{(n-1)\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{(n-2)\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \quad (2)$$

$$Bias = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n} \quad (3)$$

where n is the number of samples, $\hat{y}_i$ is the predicted AGB value, and y_i is the reference AGB from ground measurements.

The cross-validation method was used in each bootstrap to estimate the model performance metrics, including R^2, and the accuracy and the uncertainties of the predicted AGB. The cross-validation method withholds a random sample of ground AGB observations and the correspondent predictors and uses this sample to test the validity of the predictive model. The ground observation data were randomly divided into a training set (70% of the total data) and a validation set (the remaining 30%) for each bootstrap. To determine the model performance, the R^2, RMSE and bias were computed using the training and validation data set to perform in-sample and out-of-sample tests. In

addition, we evaluated model performance using scatterplots of predicted vs. observed AGB, including both the training and validation sets.

The bootstrapping also allowed us to estimate the uncertainty of AGB estimates. Here, uncertainty refers to variation of the prediction (i.e., precision). Thus, the uncertainties of the AGB prediction can be readily estimated by calculating the standard deviations of multiple predictions from bootstrapping. We computed the mean and the standard deviation for the AGB estimates of each pixel based on our 100 bootstrapped Cubist models. The mean serves as our final aggregated prediction and the standard deviation is used for the uncertainty of the predictions.

Finally, we used all available field samples to compare the predictive accuracy between the map of the ensemble means based on 100 bootstraps and the Saatchi, Avitabile, and Su maps. The R^2, RMSE, and bias values were computed for comparison. In addition, we compared our ensemble mean prediction with the Saatchi, Avitabile, and Su map using scatterplots of predicted vs. observed AGB, including all ground observations.

We also computed the variable importance (VarImp) for each Cubist model, using the R Package Caret to assess the relative contributions of the predictor variables to the modeling process (Kuhn et al., 2018). Variables of importance are the linear combinations of the predictor variables used in both the rule conditions and the linear models (terminal) in a committee of Cubist regression trees.

3. Results

3.1. Model development and variable importance

Our ensemble model consisted of 100 predictive Cubist regression tree models through bootstrapping, where each Cubist regression tree model consisting of 10 committee models composed of 2–10 rule-based sub models (See Supporting information 1 for an example of a Cubist regression tree model). Among these models, the NDVI appeared the most important variable, followed by Band 1 and 4, the NDWI, Band 6 and 2, and the LSWI, which were also important variables (Table 2). Although other variables contributed to the predictive ability of the models, they were relatively less important (Table 2).

3.2. Model prediction

The AGB across the subtropical forest area in China was mapped at a 1-km spatial resolution and exhibited large spatial variability across the landscape (Fig. 4). Our Cubist ensemble mean model based on 16 explanatory variables explained 75% of the variance in forest AGB, with an RMSE of 45.5 Mg/ha. The estimated AGB in the region varied from 0 Mg/ha to 392 Mg/ha, and the average AGB for the whole study area (i.e., canopy cover > 10%, Fig. 2a) was 123.2 Mg/ha. Overall, the forest AGB was higher in the eastern part (110° E–125° E) than that in the western part (95° E–110° E) of the study area. The average forest AGB in the eastern part was 150–200 Mg/ha, with over half of the eastern area had an AGB of 200–250 Mg/ha. The average AGB for the western part was 40–80 Mg/ha and appeared more heterogeneously across the space. The highest AGB (> 300 Mg/ha) was concentrated in the southeastern part of subtropical China where the forest is dominated by SCLF, SBEF and SBES. The lowest AGB (< 50 Mg/ha) was distributed in the southwestern part of the study area where the forest cover was low (10–20%) and dominated by SPMF and SBDF. The majority of the study area had an AGB of 50–100 Mg/ha (27%), followed by areas with an AGB of 100–150 Mg/ha (22%). Less than 3% of the study area had an AGB of < 50 Mg/ha or > 300 Mg/ha (Fig. 4 inset).

Considering the uncertainties derived from 100 bootstraps, the total amount of AGB in the subtropical forests of China was $(266 \pm 9.1) \times 10^6$ Mg, which is ~12% lower than the estimate of Saatchi (297×10^6 Mg), ~2% and ~11% higher than the estimates of Avitabile (263×10^6 Mg) and Su (237×10^6 Mg), respectively. Similarly, the predicted average forest AGB over the study area (123.2 Mg/ha) was lower than that of Saatchi (136.5 Mg/ha) and higher than those of Avitabile (116.8 Mg/ha) and Su (107.1 Mg/ha). Moreover, our AGB map presents spatial patterns that differ substantially from the three previous maps (Fig. 5). Specifically, 45% of the study area exhibited a higher AGB estimate than the Saatchi map, whereas 55% of the area exhibited lower AGB estimates than Saatchi (Fig. 5a inset). Compared with the Avitabile and Su maps, our AGB estimate was lower for 44% of the study area but higher for 56% the region (Fig. 5b and c inset). In the southwest, where the elevation is much higher than in the other parts, we estimated a lower AGB (> 100 Mg/ha difference) than the three previous maps. In the central and southeast areas, our AGB estimate was lower than that of Saatchi but higher than those of Avitabile and Su. In contrast, in the northwest, these differences were reversed.

Table 2
Variable importance (VarImp) of the 100 Cubist regression tree models used to predict the forest AGB in subtropical China. Band1 to band7 are the surface reflectance of the red (620–670 nm), near infrared (NIR, 841–876 nm), blue (459–479 nm), green (545–565 nm), thermal (1230–1250 nm), shortwave infrared (SWIR1, 1628–1652 nm), and SWIR2 (2105–2155 nm) of MODIS, respectively.

Variable	VarImp(%)		
	Mean	10% Quantile	90% Quantile
NDVI	92.47	87.50	98.20
Band1	45.42	38.52	50.00
Band4	42.23	37.50	51.20
NDWI	40.70	37.50	50.00
Band6	29.33	23.00	38.45
Band2	27.53	13.52	40.26
LSWI	25.43	15.00	35.75
Band7	23.13	18.50	31.50
AP	17.49	14.00	21.50
Band3	11.04	0	13.00
Band5	9.39	0	18.75
Height	8.72	0	18.52
DEM	6.52	0	10.57
MSAVI	6.00	0	10.88
MAT	2.64	0	0
EVI	0.98	0	0

3.3. Model performance and accuracy

Cross validation from independent testing data indicated that our map achieved high accuracy ($R^2 = 0.65 \pm 0.07$, RMSE = 54 ± 7.1 Mg/ha, Bias = −5.7 ± 3.1), but the accuracy varied by forest type (Table 3). The R^2 of the cross validation from 100 bootstraps ranged from 0.64 to 0.80 among the forest type, which was similar to the values of the training dataset (0.66 to 0.84), suggesting no significant overfitting or underfitting of our Cubist models. A high performance was achieved for both the training and testing datasets in most forest type, including the SMNB, SPMF, SBDF, SBEF, and SCLF. Among these forest type, SMNB had the lowest RMSE (31.9 ± 10 Mg/ha) and bias (0.8 ± 5.8 Mg/ha), with its median AGB of 90 Mg/ha (Fig. 4). That indicated approximately half of the ground observations within the AGB range of 50–100 Mg/ha. The SPAF had the highest RMSE (82.7 ± 28.2 Mg/ha) and bias (−36.8 ± 19.5 Mg/ha), which had the smallest number of reference data (16) and a very high AGB value (> 600 Mg/ha), which may not be representative.

Compared with the previous maps, our AGB map (ensemble mean based on 100 bootstraps) has higher accuracy when assessed using all available samples (Figs. 6 and 7). The RMSE and bias of the AGB estimate were 45.5 Mg/ha ($R^2 = 0.75$) and −5.8 Mg/ha, which were significantly lower than those of Saatchi ($R^2 = 0.20$, RMSE = 89 Mg/ha,

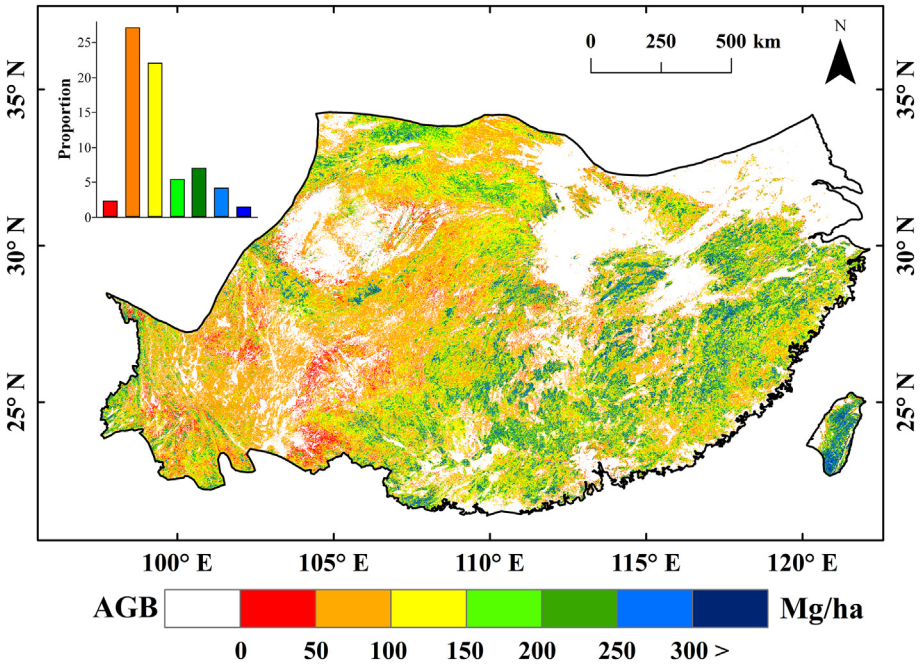

Fig. 4. The distribution of AGB at 1-km resolution for the subtropical forest in China. The white area represents areas with forest cover of < 10%. The histograms in the upper left represent the proportion of each AGB class. The map shown here is based on the mean prediction from 100 bootstrap estimates (i.e., the means of each pixel based on 100 estimated maps).

bias = 7.5 Mg/ha), Avitabile (R^2 = 0.26, RMSE = 84.9 Mg/ha, bias = −12.2 Mg/ha), and Su (R^2 = 0.14, RMSE = 92.3 Mg/ha, bias = −21.9 Mg/ha) (Table S2). The RMSE and bias of our map were also consistently lower than those of the previous maps among different forest type, except for SPAF (Fig. 6a). In SPAF, we produced a lower bias than Su but slightly higher values than Saatchi and Avitabile (Fig. 6b). Moreover, the AGB estimates for almost all forest type were underestimated in Avitabile and Su but overestimated in Saatchi.

Our AGB estimates were also correlated better with the field observations than the other AGB maps for all forest type except STBS, which had a small number of ground observations (Fig. 7). In addition, the predictive accuracy varied by AGB magnitude. Our estimates were

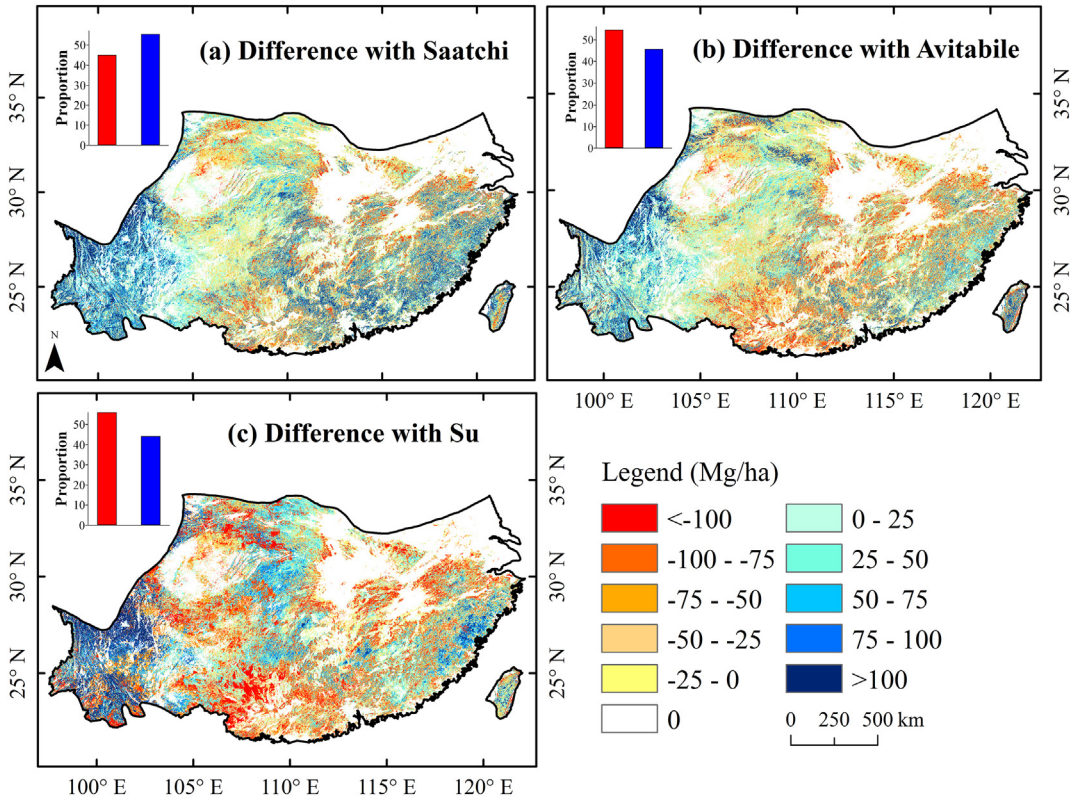

Fig. 5. Difference maps between estimated AGB in this study (the ensemble mean based on 100 bootstraps) and those of (a) Saatchi, (b) Avitabile, and (c) Su AGB maps. The histograms in the upper left of each map represent the distribution of the differences grouped into two classes: red indicates that the estimates of Saatchi, Avitabile or Su are lower than in this study, and blue indicates that the estimates of Saatchi, Avitabile or Su are higher than in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Model performance on the training, testing and total data for eight forest type. Note that STBS was not included due to the small amount of ground data.

Forest type	R^2		RMSE (Mg/ha)		Bias (Mg/ha)	
	Train	Test	Train	Test	Train	Test
SCLF	0.74 ± 0.08	0.67 ± 0.11	55.3 ± 10.4	62.7 ± 17.4	−9.1 ± 3.5	−10.1 ± 6.8
SPMF	0.77 ± 0.03	0.71 ± 0.08	34.0 ± 2.2	39.4 ± 7.7	0.04 ± 1.5	1.2 ± 4.3
SPAF	0.66 ± 0.08	0.64 ± 0.14	64.7 ± 12.6	82.7 ± 28.2	−28.7 ± 7.8	−36.8 ± 19.5
SMNB	0.84 ± 0.03	0.80 ± 0.09	28.8 ± 2.6	31.9 ± 10.0	−0.5 ± 2.6	0.8 ± 5.8
SBDF	0.73 ± 0.05	0.68 ± 0.13	43.9 ± 4.2	52.9 ± 23.1	1.7 ± 3.2	2.5 ± 10.8
SBEF	0.73 ± 0.04	0.66 ± 0.09	45.4 ± 4.0	53.3 ± 10.4	−7.1 ± 2.1	−6.6 ± 5.1
SBES	0.70 ± 0.05	0.65 ± 0.11	54.1 ± 5.5	58.2 ± 13.7	−14.7 ± 3.7	−15.2 ± 9.9
Total	0.74 ± 0.03	0.65 ± 0.07	46.0 ± 3.4	54.0 ± 7.1	−5.8 ± 1.3	−5.7 ± 3.1

mostly lower than the ground-observed AGB for values > 150 Mg/ha but closer to the ground observation for medium AGB (50–150 Mg/ha). This difference in predictive accuracy was observed in the distribution of AGB classes for each forest type (Fig. 8). For example, the SMNB with medium AGB ranging from 50 to 150 Mg/ha had the best correlation with observations (R^2 = 0.85, Fig. 7, Table S3), with no substantial over/under estimation. However, the SPAF (bias = −31.0 Mg/ha) and SBES (bias = −14.8 Mg/ha), which have a large proportion of AGB > 150 Mg/ ha, exhibited significantly underestimated AGBs.

3.4. Model uncertainty

The average uncertainty (expressed as standard deviation) of our AGB estimates over the entire study area at the pixel level was 11.9 Mg/ha, and uncertainty ranged between 0 Mg/ha and 115 Mg/ha for all pixels (Fig. 9a). The uncertainty for the majority of the study area (62%) was < 15 Mg/ha, with only a small part of the study area (8%) had uncertainty of > 45 Mg/ha. In addition, uncertainty was positively associated with the magnitude of the AGB estimates (Fig. 9b). The southeastern part of the study area with higher AGB estimates presented higher uncertainty (15–18%), whereas the western part with lower AGB estimates presented lower uncertainty (< 10%). Overall, the majority of the study area (88%) had an uncertainty lower than 18% (Fig. 9b inset). The largest uncertainty was observed in the areas with AGB > 300 Mg/ha.

4. Discussion

4.1. AGB estimation in subtropical forests

We improved the accuracy of AGB estimation for the subtropical forests in China through a synthesis of a large number of ground observations with MODIS, forest cover change, forest canopy height, climate, and topographical data for training machine learning algorithms (i.e., Cubist regression with bootstrapping). Since the majority of global AGB is stored in forests (Liu et al., 2015) and biomass has been identified as an important variable in climate models for understanding drivers and effects of climate change (https://gcos.wmo.int/en/essential-climate-variables), the compiled ground AGB measurements and the improved AGB map can help better understanding the role of the subtropical forest in global carbon cycling and climate change. Having a sufficient number of high-quality ground observations representative of the spatial distribution of biomass was reported as a key factor for improving AGB estimates (Avitabile et al., 2016; Timothy et al., 2016; Chave et al., 2015; Saatchi et al., 2011). A sufficient number of training data is especially critical for machine learning algorithms, which are mostly non-linear. We were able to compile a large amount of published ground AGB data for 2000–2014 (Tables 1 and S1) and carefully screen these data through quality control. This ground dataset covers the main forest type in the study area, a wide range of biomass values, and presents a wide spatial distribution (Fig. 4), which greatly facilitated the successful training of the Cubist regression tree model. As a result, we improved the forest AGB mapping for subtropical China from previous products, including the Saatchi, Avitabile, and Su maps (Figs. 6 and 7) where the authors did not include or only had a small amount of field observations for model training. The machine learning algorithm Cubist regression tree also contributed to the improved AGB map. The relationship between AGB and remote sensing measurements is often characterized by complexity and non-linearity (Twomey, 1997), which require more powerful algorithms (i.e., machine learning algorithms such as Cubist, ANN, random forest, and SVM) than linear regressions to process a large number of inputs and handle non-linearity. These machine learning algorithms have been widely used for AGB mapping, although the superiority of each

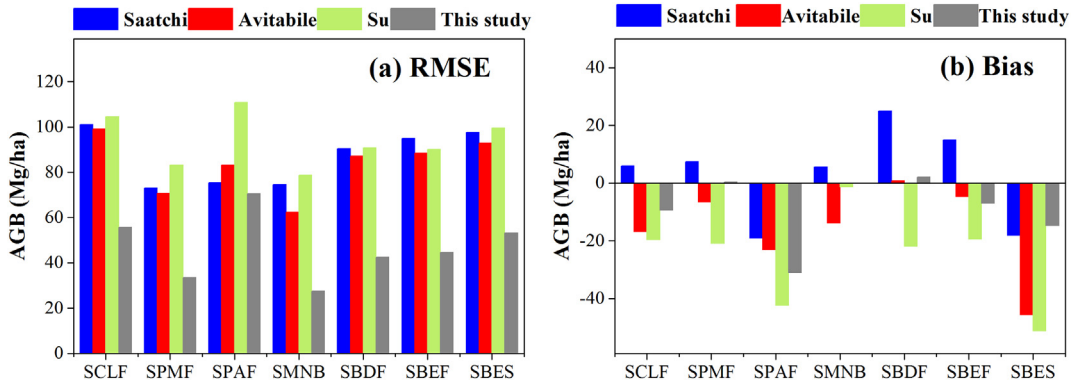

Fig. 6. (a) RMSE and (b) bias of the AGB estimate in this study (the ensemble mean based on 100 bootstraps) and three existing AGB maps by forest type. STBS was not included due to the small number of ground data points. Subtropical *Cunninghamia lanceolata* forest (SCLF), subtropical *Pinus massoniana* forest (SPMF), subtropical *Picea-Abies* forest (SPAF), subtropical mixed needleleaf and broadleaf forest (SMNB), subtropical broadleaf deciduous forest (SBDF), subtropical broadleaf evergreen forest (SBEF), subtropical broadleaf evergreen sclerphyllous forest (SBES), and subtropical and tropical bamboo and scrub forest (STBS).

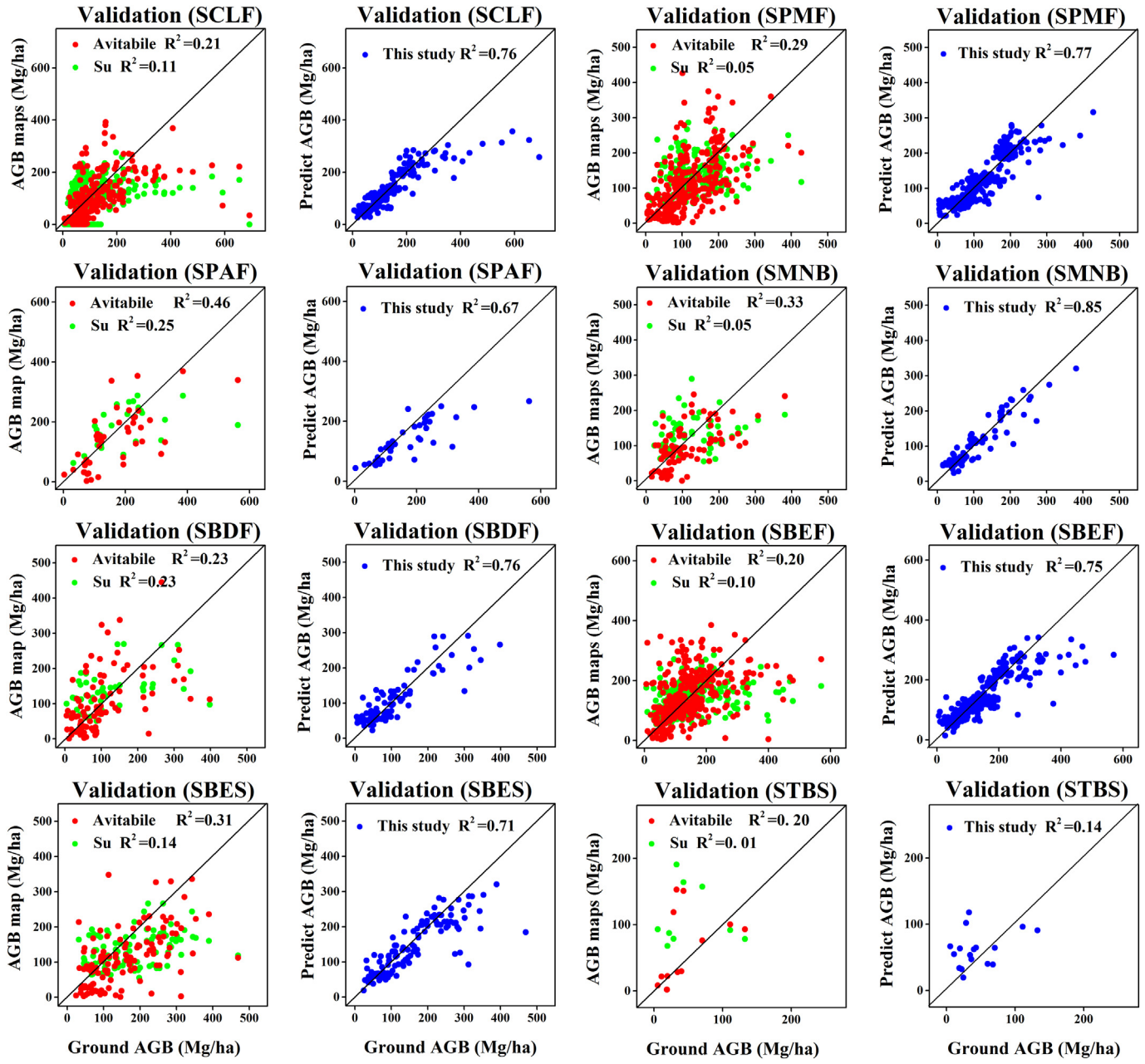

Fig. 7. Scatterplots of the validated ground-observed AGB (x-axis) and predictions (y-axis) of Avitabile (red) and Su (green) maps (left plots) and our ensemble mean estimates (blue) (right plots) by forest type. The 1:1 line is reported. The Saatchi map is not presented because it is similar to the Avitabile map. Subtropical *Cunninghamia lanceolata* forest (SCLF), subtropical *Pinus massoniana* forest (SPMF), subtropical *Picea-Abies* forest (SPAF), subtropical mixed needleleaf and broadleaf forest (SMNB), subtropical broadleaf deciduous forest (SBDF), subtropical broadleaf evergreen forest (SBEF), subtropical broadleaf evergreen sclerphyllous forest (SBES), and subtropical and tropical bamboo and scrub forest (STBS). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

algorithm is case dependent (Ali et al., 2017). We applied the Cubist regression tree model, and used a similar "bagging" approach of random forest through bootstrapping. Comparing different machine learning algorithms is beyond the scope of the current study, and the evaluation of the Cubist regression tree models shows good performance.

All predictors contributed to the ensemble model, but the vegetation indices and spectral bands appear to be the most important predictors (Table 2), which is consistent with previous studies (Foody et al., 2003; Ranjeet et al., 2018; Zheng et al., 2004). Water-related variables are also important, such as the NDWI, LSWI, and AP, which had a high score on variable importance. This finding was also expected because water content indices, including the NDWI and LSWI, have been consistently chosen as important variables for predicting AGB (Huete et al., 2002), and the ecosystem water deficit constrains the forest biomass (Stegen et al., 2011). Other variables (e.g., DEM, MAT, and canopy height) could contribute to the predictive ability of the models. However, excluding them would only slightly affect the results because many of the 100 Cubist regression tree models did not use them (as suggested by the zero score on variable importance). Canopy height is a good predictor for biomass in many cases (e.g., Zhang et al., 2014; Gwenzi et al., 2017), although it was not important in our model because the modeled 1-km forest canopy height map (Simard et al., 2011) was not detailed enough to represent the local variability of canopy height.

Our estimate of the total AGB stock in the subtropical forests of

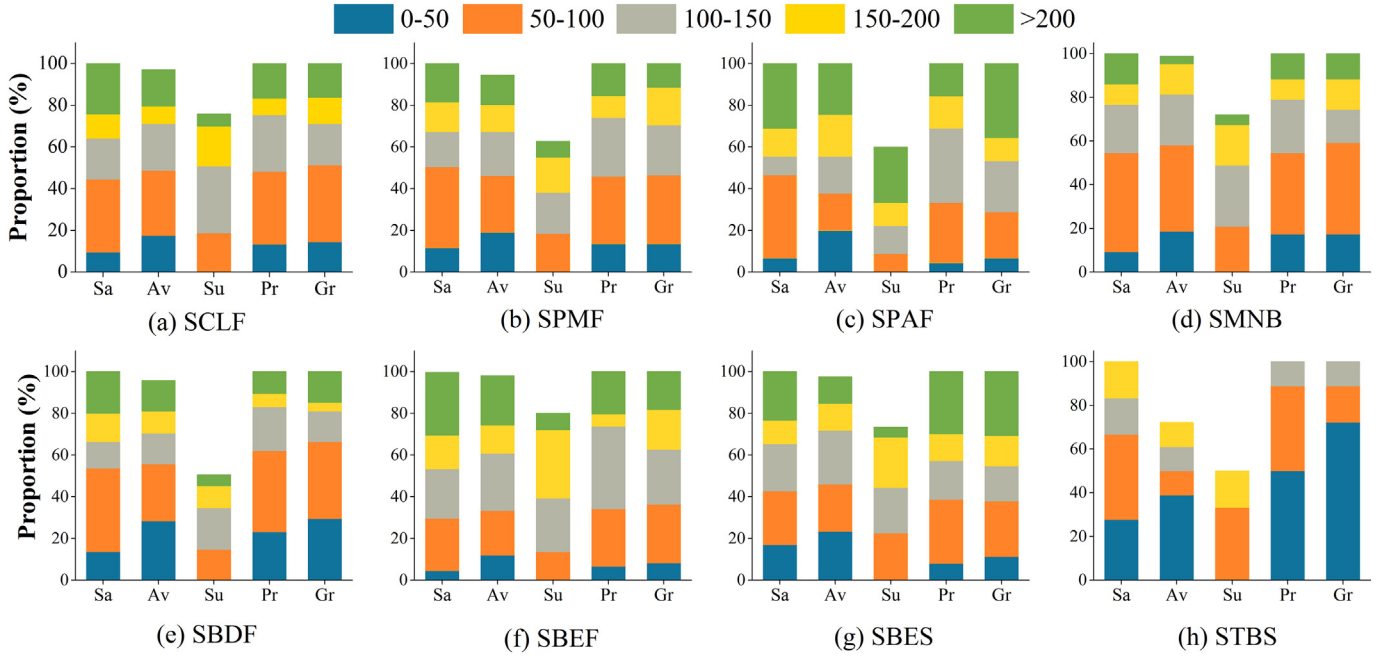

Fig. 8. Frequency distribution of AGB (Mg/ha) classes by forest type. Due to the scarcity of high AGB values, all data with AGB of > 200 Mg/ha were grouped into one class. Sa = Saatchi, Av = Avitabile, Su = Su, Pr = The ensemble mean estimates in this study, and Gr = Ground data. Subtropical *Cunninghamia lanceolata* forest (SCLF), subtropical *Pinus massoniana* forest (SPMF), subtropical *Picea-Abies* forest (SPAF), subtropical mixed needleleaf and broadleaf forest (SMNB), subtropical broadleaf deciduous forest (SBDF), subtropical broadleaf evergreen forest (SBEF), subtropical broadleaf evergreen sclerphyllous forest (SBES), and subtropical and tropical bamboo and scrub forest (STBS).

China is 11% lower than that of Saatchi (Fig. 5a), which is consistent with the overestimation of 23–42% of the total stocks in the Saatchi map reported by Mitchard et al. (2013). In particular, a comparison between our estimates and ground observations showed that the Saatchi map overestimated AGB for values > 200 Mg/ha, in SCLF, SBDF and SBEF (Figs. 7 and 8 and S3). Our estimates of AGB stocks are higher than those of Su (Fig. 5c), mainly because Su trained his predictive model using ground observations from northern China where the forest AGB is lower than subtropical China. Su's map was produced only for the forest cover identified in the land-use map of China at a 1:1,000,000 scale for the year 2000, which did not include many forest areas with low canopy coverage (Liu et al., 2010; Su et al., 2016). This was also confirmed by our visual comparison of Su's forest map with Hansen et al.'s (2013) global forest cover map (Fig. S1). Consequently, Su's map tends to miss areas with AGBs < 50 Mg/ ha (Figs. 8 and S3). Both the underestimation of pixel-level AGB and forest cover area caused the underestimation of total AGB estimates. Our estimate of the total AGB stocks is similar to that of Avitabile (Fig. 5b), but this convergence seems to be mostly due to compensation of contrasting estimates when averaging over large areas. In addition, neither the Saatchi nor Avitabile maps contain any ground data in our study area, which likely caused the low accuracy of their AGB estimates compared to that of our estimates.

4.2. Limitations of the current study

Although our AGB map of subtropical forest in China achieved higher accuracy than previous maps, both our data and methodology suffer from limitations similar to those in previous studies.

The first limitation is the uncertainty inherent in ground measurements of AGB uncertainty caused by both the AGB estimation procedure (e.g., errors in measurement and in the allometric models) and geolocation errors (Mitchard et al., 2013, 2014; Santoro et al., 2015). By selecting only samples located in homogeneous MODIS pixels, we reduced the uncertainties caused by inaccurate GPS values and the scale mismatch between plot size and MODIS pixel size. However, the errors could not be completely eliminated and their impact on the results is difficult to determine. These estimate uncertainties caused by errors are especially relevant when using generalized allometric models or conversion factors, which do not account for the variability in species composition and environmental/climatic conditions (Guo et al., 2010; Halme and Tomppo, 2001). The uncertainties inherited from ground AGB estimates may be further reduced using a local or species-specific AGB estimation method and using a plot design that allows for better integration with remote sensing data (Muukkonen and Heiskanen, 2007).

The second limitation is the temporal difference between ground measurement and remote sensing data, which may introduce some uncertainties into the prediction model (Avitabile et al., 2016; Saatchi et al., 2007, 2011). Our compiled ground dataset was collected from 2000 to 2014, whereas the MODIS data were acquired in 2000. Because many plots did not record the exact year when they were sampled, we could not match the year of all plot measurements with the year of satellite data acquisition for model fitting. However, if we used only the plots that were measured in 2000, the number of plots (~100) would be too small to train the Cubist model. We reduced the temporal mismatch error by excluding ground measurements where forest canopy cover had significantly changed (Fig. S2), but we did not account for small forest growth or degradation effects due to the lack of reliable growth factors or degradation maps.

The third limitation is the saturation of the remote sensing data at high biomass density (> 150 Mg/ha) (Baccini et al., 2008; Houghton et al., 2009). We used spectral bands, vegetation indices including NDVI, EVI, LSWI, NDWI and MSAVI, as well as forest canopy height, climate (temperature and precipitation), and topography variables to generate the prediction model. The use of multiple predictive variables may offset the underestimation caused by spectral saturation to some extent (Saatchi et al., 2007). However, the saturation effect remains present in our product, as indicated by the underestimation of AGB > 150 Mg/ha compared with the ground data (Fig. 7). The pattern was also observed in other biomass maps when compared to forest inventory data (Avitabile and Camia, 2018). The signal saturation cannot

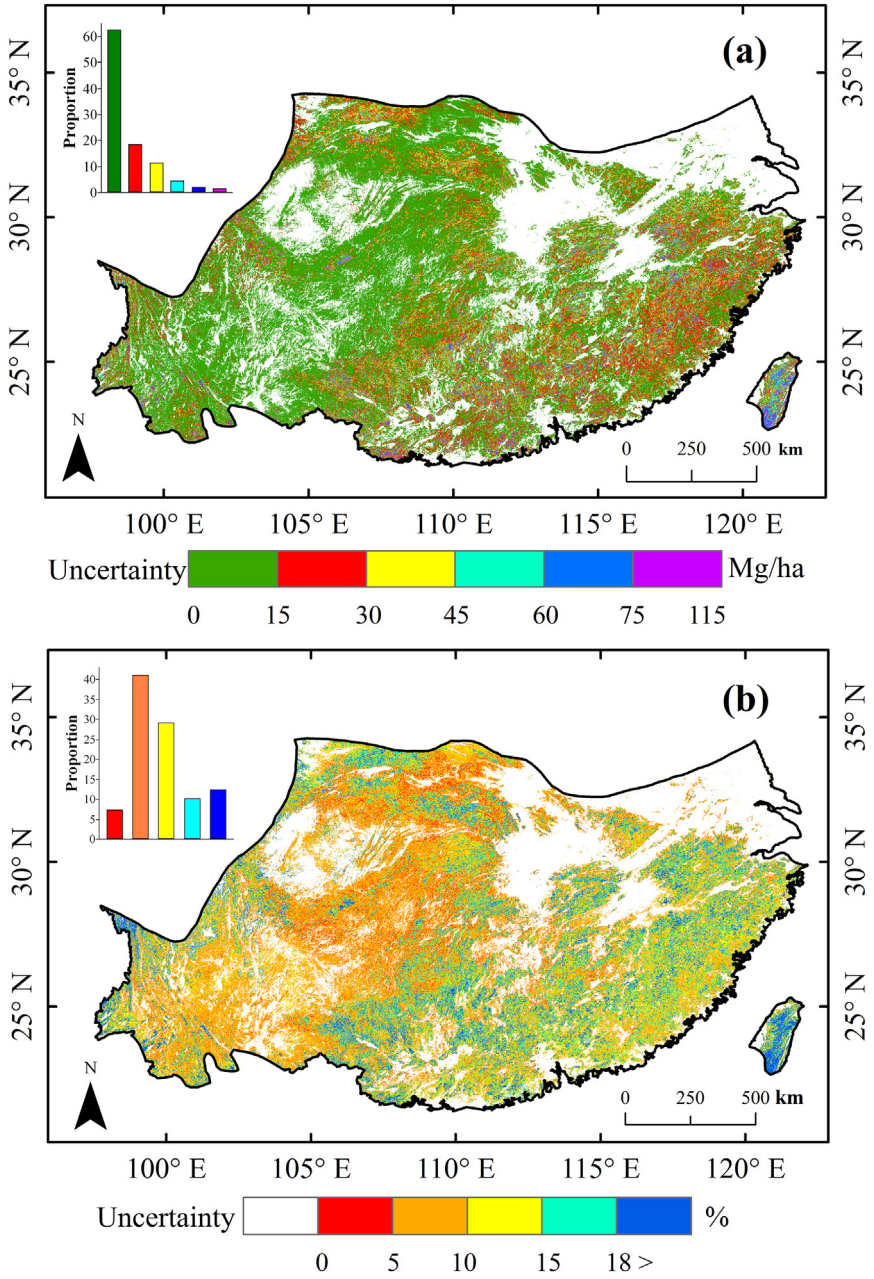

Fig. 9. Uncertainty of the AGB map (a) in absolute values (reported as standard deviation) and (b) relative to that of the ensemble mean AGB estimates based on 100 bootstraps.

be completed solved by regression tree-based models, which inherently tend to aggregate their prediction at extreme values, resulting in the prediction of lower AGB for regions with high AGB and vice versa (Baccini et al., 2008). However, the saturation problem could be reduced by the integration of optical, radar, and spaceborne and airborne LIDAR data (Cutler et al., 2012) for further studies.

4.3. Toward further improvements of forest AGB estimations

One method to improve the forest AGB estimates is to include forest age as a predictor. When building our Cubist model, we found that including the observed forest age as a predictor significantly increased the performance of the model and decreased the uncertainty in the AGB estimation (Table S4). This notion is consistent with previous studies that also emphasized that forest stand age is an important factor for improving the accuracy of AGB estimation (Du et al., 2014; Guo et al., 2010; Pan et al., 2004). However, a reliable, spatially explicit forest age dataset is not available for our study area. Zhang et al. (2017) provided a 1-km forest stand age map for China based on the national forest inventory data, climate data, and LIDAR-based forest height. However, our ground forest stand age data indicated that this dataset has large uncertainties and bias for our study area (Fig. S4). As a result, we excluded forest age from our model, but we note that an accurate forest age map may substantially improve future forest AGB estimates.

Another effective method to improve AGB estimation is through hierarchical models developed for different forest type groups. In this study, we pooled all samples from all type of forests and established a single universal model. This model performed differently for different forest type (Table 3), suggesting that samples from different forest type could not be regarded as true replicates because they are governed by similar but not identical processes (e.g., different canopy structure, soil background, and climate). Although the same regression model may

hold for different forest type, the specific parameters may vary according to the local characteristics of each forest type. Recent studies have suggested that when among-group heterogeneity is caused by spatial, temporal, or other organizational factors, hierarchical models (especially Bayesian hierarchical models) can significantly improve the predictive accuracy (Asner et al., 2013; Qian et al., 2015; Yun and Qian, 2015), as hierarchical models not only incorporate information from all samples but also consider the unique situation of each group simultaneously through partially pooling samples, thus improving the power of prediction. However, upscaling a hierarchical model based on different forest type groups requires a land cover map identifying these forest type, which was not available for our study area. However, we note that integrating advanced forest mapping techniques for accurate forest type mapping may substantially improve future AGB estimations.

5. Conclusions

The major contribution of this study is the improved accuracy of above-ground biomass (AGB) estimates in the subtropical forest of China achieved by synthesizing a large number of ground AGB observations and integrating multi-sourced remote sensing data. Based on this improved estimate, we found the total AGB stock for the subtropical forest in China is $(266 \pm 9.1) \times 10^6$ Mg. In addition, to better understand the reliability of our predicted AGB map, we validated our estimates using independent testing data and compared them with three previous AGB maps. The results indicate that our new predicted map not only captured the amount and spatial distribution of AGB but also showed lower RMSE and bias than the previous maps. Moreover, the study also highlights the importance of considering forest type and stand age for reducing uncertainties in AGB prediction. To improve the accuracy of forest biomass estimation at the regional scale, our future work will focus on collecting more ground data for all relevant forest type and building hierarchical models that consider the local characteristics of different forest type.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2019.111341.

Acknowledgements

We would like to thank Yunjian Luo, Hao Zhang, Shunlei Peng, and Nianpeng He for sharing the AGB ground data with us. We acknowledge Sassan S. Saatchi and Yanjun Su for providing a free subtropical forest AGB map. This study was supported by the National Natural Science Foundation of China (Grant No. 41801025, 31770559, 31600352, 31600387), the Fundamental Research Funds for the Central Universities, and “Thousand Young Talents” Program in China.

References

Ali, I., Greigeneder, F., Stamenkovic, J., Neumann, M., Notarnicola, C., 2017. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. Remote Sens. 7, 16398–16421.

Anonymous, 1987. Stem Volume Table. Chinese Forestry Publisher, Beijing, China.

Asner, G.P., Mascaro, J., Anderson, C., Knapp, D.E., Martin, R.E., Kennedy-Bowdoin, T., van Breugel, M., Davies, S., Hall, J.S., Muller-Landau, H.C., Potvin, C., Sousa, W., Wright, J., Bermingham, E., 2013. High-fidelity national carbon mapping for resource management and REDD+. Carbon Balance Manag. 8, 7.

Avitabile, V., Camia, A., 2018. An assessment of forest biomass maps in Europe using harmonized national statistics and inventory plots. For. Ecol. Manag. 409, 489–498.

Avitabile, V., Herold, M., Henry, M., Schmullius, C., 2011. Mapping biomass with remote sensing: a comparison of methods for the case study of Uganda. Carbon Balance Manag. 6, 7.

Avitabile, V., Herold, M., Heuvelink, G.B.M., Lewis, S.L., Phillips, O.L., Asner, G.P., Armston, J., Ashton, P.S., Banin, L., Bayol, N., Berry, N.J., Boeckx, P., de Jong, B.H.J., DeVries, B., Girardin, C.A.J., Kearsley, E., Lindsell, J.A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E.T.A., Nagy, L., Qie, L., Quinones, M.J., Ryan, C.M., Ferry, S.J.W., Sunderland, T., Laurin, G.V., Gatti, R.C., Valentini, R., Verbeeck, H., Wijaya, A., Willcock, S., 2016. An integrated pan-tropical biomass map using multiple reference datasets. Glob. Chang. Biol. 22, 1406–1420.

Baccini, A., Asner, G.P., 2013. Improving pantropical forest carbon maps with airborne LiDAR sampling. Carbon Manage. 4, 591–600.

Baccini, A., Laporte, N., Goetz, S.J., Sun, M., Dong, H., 2008. A first map of tropical Africa’s above-ground biomass derived from satellite imagery. Environ. Res. Lett. 3, 045011.

Baccini, A., Goetz, S.J., Walker, W.S., Laporte, N.T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P.S.A., Dubayah, R., Friedl, M.A., Samanta, S., Houghton, R.A., 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nat. Clim. Chang. 2, 182–185.

Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., Houghton, R.A., 2017. Tropical forests are a net carbon source based on aboveground measurements of gain and loss. Science 358, 230–234.

Badreldin, N., Sanchez-Azofeifa, A., 2015. Estimating forest biomass dynamics by integrating multi-temporal landsat satellite images with ground and airborne LiDAR data in the Coal Valley Mine, Alberta, Canada. Remote Sens. 7, 2832–2849.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rodenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K.W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F.I., Papale, D., 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science 329, 834–838.

Blackard, J., Finco, M., Helmer, E., Holden, G., Hoppus, M., Jacobs, D., Lister, A., Moisen, G., Nelson, M., Riemann, R., 2008. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. Remote Sens. Environ. 112, 1658–1677.

Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L., Asner, G.P., 2018. An above-ground biomass map of African savannahs and woodlands at 25 m resolution derived from ALOS PALSAR. Remote Sens. Environ. 206, 156–173.

Cartus, O., Kellndorfer, J., Walker, W., Franco, C., Bishop, J., Santos, L., Fuentes, J., 2014. A national, detailed map of forest aboveground carbon stocks in Mexico. Remote Sens. 6, 5559–5588.

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., Vieilledent, G., 2015. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Chang. Biol. 20, 3177–3190.

Chen, J., John, R., Sun, G., Mcnulty, S., Noormets, A., Xiao, J., Turner, M.G., Franklin, J.F., 2014. Carbon fluxes and storage in forests and landscapes. In: Azevedo, J.C., Perera, A.H., Pinto, M.A. (Eds.), Forest Landscapes and Global Change. Springer, New York, NY, pp. 139–166.

Cutler, M.E.J., Boyd, D.S., Foody, G.M., Vetrivel, A., 2012. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: an assessment of predictions between regions. ISPRS J. Photogramm. Remote Sens. 70, 66–77.

Du, L., Zhou, T., Zou, Z., Zhao, X., Huang, K., Wu, H., 2014. Mapping forest biomass using remote sensing and national forest inventory in China. Forests 5, 1267–1283.

Fang, J., Wang, G., Liu, G., Xu, S., 1998. Forest biomass of China: an estimate based on the biomass-volume relationship. Ecol. Appl. 8, 1084–1091.

Fayad, I., Baghdadi, N., Guitet, S., Bailly, J.-S., Hérault, B., Gond, V., El Hajj, M., Minh, D.H.T., 2016. Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data. Int. J. Appl. Earth Obs. Geoinf. 52, 502–514.

Foody, G.M., Boyd, D.S., Cutler, M.E.J., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. Remote Sens. Environ. 85, 463–474.

Gao, B.C., 1996. NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266.

Giannico, V., Lafortezza, R., John, R., Saneso, G., Pesola, L., Chen, J., 2016. Estimating stand volume and above-ground biomass of urban forests using LIDAR. Remote Sens. 8, 339.

Gleason, C.J., Im, J., 2012. Forest biomass estimation from airborne LiDAR data using machine learning approaches. Remote Sens. Environ. 125, 80–91.

Guo, Z., Fang, J., Pan, Y., Birdsey, R., 2010. Inventory-based estimates of forest biomass carbon stocks in China: a comparison of three methods. For. Ecol. Manag. 259, 1225–1231.

Gwenzi, D., Helmer, E.H., Zhu, X., Lefsky, M.A., Marcano-Vega, H., 2017. Predictions of tropical forest biomass and biomass growth based on stand height or canopy area are improved by landsat-scale phenology across Puerto Rico and the U.S. Virgin Islands. Remote Sens. 9, 123.

Hall, R.J., Skakun, R.S., Arsenault, E.J., Case, B.S., 2006. Modeling forest stand structure attributes using Landsat ETM+ data: application to mapping of aboveground biomass and stand volume. For. Ecol. Manag. 225, 378–390.

Halme, M., Tomppo, E., 2001. Improving the accuracy of multisource forest inventory estimates to reducing plot location error — a multicriteria approach. Remote Sens. Environ. 78, 321–327.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853.

Houghton, R.A., Hall, F., Goetz, S.J., 2009. Importance of biomass in the global carbon cycle. J. Geophys. Res. Biogeosci. 114, 1–13.

Houghton, R.A., House, J.I., Pongratz, J., van der Werf, G.R., DeFries, R.S., Hansen, M.C., Le Quéré, C., Ramankutty, N., 2012. Carbon emissions from land use and land-cover change. Biogeosciences 9, 5125–5142.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.

Ju, J., Roy, D.P., Shuai, Y., Schaaf, C., 2010. Development of an approach for generation

of temporally complete daily nadir MODIS reflectance time series. Remote Sens. Environ. 114, 1–20.
Keith, H., Mackey, B.G., Lindenmayer, D.B., 2009. Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. Proc. Natl. Acad. Sci. U. S. A. 106, 11635–11640.
Kuhn, M., Coulter, N., 2012. Cubist Models For Regression. (R package Vignette R package version 0.0).
Kuhn, M., Weston, S., Keefer, C., Coulter, N., Quinlan, R., 2018. Package Package 'Cubist' Version 0.0.13.
Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., Ulander, L., 2011. The BIOMASS mission: mapping global forest biomass to better understand the terrestrial carbon cycle. Remote Sens. Environ. 115, 2850–2860.
Lefsky, M.A., Cohen, W.B., Harding, D.J., Parker, G.G., Acker, S.A., Gower, S.T., 2002. Lidar remote sensing of above-ground biomass in three biomes. Glob. Ecol. Biogeogr. 11, 393–399.
Liu, J., Zhang, Z., Xu, X., Kuang, W., Zhou, W., Zhang, S., Li, R., Yan, C., Yu, D., Wu, S., Jiang, N., 2010. Spatial patterns and driving forces of land use change in China during the early 21st century. J. Geogr. Sci. 20, 483–494.
Liu, Y., van Dijk, A.I.J.M., de Jeu, R.A.M., Canadell, J.G., McCabe, M.F., Evans, J.P., Wang, G., 2015. Recent reversal in loss of global terrestrial biomass. Nat. Clim. Chang. 5, 470–474.
Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. Int. J. Remote Sens. 27, 1297–1328.
Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., Moran, E., 2016. A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. Int. J. Digital Earth 9, 63–105.
Luo, Y., Wang, X., Zhang, X., Booth, T.H., Lu, F., 2012. Root: shoot ratios across China's forests: forest type and climatic effects. For. Ecol. Manag. 269, 19–25.
McFeeters, S.K., 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.
Mitchard, E.T.A., Saatchi, S.S., Baccini, A., Asner, G.P., Goetz, S.J., Harris, N.L., Brown, S., 2013. Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps. Carbon Balance Manag. 8, 10.
Mitchard, E.T.A., Feldpausch, T.R., Brienen, R.J.W., Lopez-Gonzalez, G., Monteagudo, A., Baker, T.R., Lewis, S.L., Lloyd, J., Quesada, C.A., Gloor, M., ter Steege, H., Meir, P., Alvarez, E., Araujo-Murakami, A., Aragão, L.E.O.C., Arroyo, L., Aymard, G., Banki, O., Bonal, D., Brown, S., Brown, F.I., Cerón, C.E., Moscoso, V.C., Chave, J., Comiskey, J.A., Cornejo, F., Medina, M.C., Da Costa, L., Costa, F.R.C., Di Fiore, A., Domingues, T.F., Erwin, T.L., Frederickson, T., Higuchi, N., Coronado, E.N.H., Killeen, T.J., Laurance, W.F., Levis, C., Magnusson, W.E., Marimon, B.S., Marimon, B.H., Mendoza Polo, I., Mishra, P., Nascimento, M.T., Neill, D., Vargas, M.P.N., Palacios, W.A., Parada, A., Pardo-Molina, G., Peña-Claros, M., Pitman, N., Peres, C.A., Poorter, L., Prieto, A., Ramirez-Angulo, H., Correa, Z.R., Roopsind, A., Roucoux, K.H., Rudas, A., Salomão, R.P., Schietti, J., Silveira, M., de Souza, P.F., Steininger, M.K., Stropp, J., Terborgh, J., Thomas, R., Toledo, M., Torres-Lezama, A., van Andel, T.R., van der Heijden, G.M.F., Vieira, I.C.G., Vieira, S., Vilanova-Torre, E., Vos, V.A., Wang, O., Zartman, C.E., Malhi, Y., Phillips, O.L., 2014. Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. Glob. Ecol. Biogeogr. 23, 935–946.
Muukkonen, P., Heiskanen, J., 2007. Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: a possibility to verify carbon inventories. Remote Sens. Environ. 107, 617–624.
Pan, Y., Luo, T., Birdsey, R., Hom, J., Melillo, J., 2004. New estimates of carbon storage and sequestration in China's forests: effects of age-class and method on inventory-based carbon estimation. Clim. Chang. 67, 211–236.
Piao, S., Fang, J., Zhu, B., Tan, K., 2005. Forest biomass carbon stocks in China over the past 2 decades: estimation based on integrated inventory and satellite data. J. Geophys. Res. Biogeosci. 110, G01006.
Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., Ohmann, J.L., 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches. Remote Sens. Environ. 114, 1053–1068.
Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126.
Qian, S.S., Chaffin, J.D., DuFour, M.R., Sherman, J.J., Golnick, P.C., Collier, C.D., Nummer, S.A., Margida, M.G., 2015. Quantifying and reducing uncertainty in estimated microcystin concentrations from the ELISA method. Environ. Sci. Technol. 49, 14221–14229.
Quinlan, J.R., 1992. Learning with continuous classes. Mach. Learn. 92, 343–348.
Ranjeet, J., Chen, J., Giannico, V., Park, H., Xiao, J., Shirkey, G., Ouyang, Z., Shao, C., Lafortezza, R., Qi, J., 2018. Grassland canopy cover and aboveground biomass in Mongolia and Inner Mongolia: spatiotemporal estimates and controlling factors. Remote Sens. Environ. 213, 34–48.
Rodríguez-Veiga, P., Saatchi, S., Tansey, K., Balzter, H., 2016. Magnitude, spatial distribution and uncertainty of forest biomass stocks in Mexico. Remote Sens. Environ. 183, 265–281.
Saatchi, S.S., Houghton, R.A., Alvalá, R.C.D.S., Soares, J.V., Yu, Y., 2007. Distribution of aboveground live biomass in the Amazon basin. Glob. Chang. Biol. 13, 816–837.
Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T.A., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M., Morel, A., 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. Proc. Natl. Acad. Sci. U. S. A. 108, 9899–9904.
Santoro, M., Beaudoin, A., Beer, C., Cartus, O., Fransson, J.E.S., Hall, R.J., Pathe, C., Schmullius, C., Schepaschenko, D., Shvidenko, A., Thurner, M., Wegmüller, U., 2015. Forest growing stock volume of the northern hemisphere: spatially explicit estimates for 2010 derived from Envisat ASAR. Remote Sens. Environ. 168, 316–334.
Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X., Tsang, T., Strugnell, N.C., Zhang, X., Jin, Y., Muller, J.-P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., D'Entremont, R.P., Hu, B., Liang, S., Privette, J.L., Roy, D., 2002. First operational BRDF, albedo nadir reflectance products from MODIS. Remote Sens. Environ. 83, 135–148.
Simard, M., Pinto, N., Fisher, J.B., Baccini, A., 2011. Mapping forest canopy height globally with spaceborne lidar. J. Geophys. Res. Biogeosci. 116, G04021.
Stegen, J., Swenson, N.G., Enquist, B.J., White, E.P., Phillips, O.L., Jorgensen, P.M., Weiser, M.D., Mendoza, A., Vargas, P., 2011. Variation in above-ground forest biomass across broad climatic gradients. Glob. Ecol. Biogeogr. 20, 744–754.
Su, Y., Guo, Q., Xue, B., Hu, T., Alvarez, O., Tao, S., Fang, J., 2016. Spatial distribution of forest aboveground biomass in China: estimation through combination of spaceborne lidar, optical imagery, and forest inventory data. Remote Sens. Environ. 173, 187–199.
Timothy, D., Onisimo, M., Cletah, S., Adelabu, S., Tsitsi, B., 2016. Remote sensing of aboveground forest biomass: a review. Trop. Ecol. 57, 125–132.
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.
Twomey, S., 1997. Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements. Dover Publications, Mineola, NY.
Walton, J.T., 2008. Subpixel urban land cover estimation: comparing cubist, random forests and support vector regression. Photogramm. Eng. Remote. Sens. 74, 1213–1222.
Wang, Y., Li, G., Ding, J., Guo, Z., Tang, S., Wang, C., Huang, Q., Liu, R., Chen, J.M., 2016. A combined GLAS and MODIS estimation of the global distribution of mean forest canopy height. Remote Sens. Environ. 174, 24–43.
Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore, B., Ojima, D., 2004. Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. Remote Sens. Environ. 91, 256–270.
Xiao, J., Zhuang, Q., Baldocchi, D.D., Law, B.E., Richardson, A.D., Chen, J., Oren, R., Starr, G., Noormets, A., Ma, S., Verma, S.B., Wharton, S., Wofsy, S.C., Bolstad, P.V., Burns, S.P., Cook, D.R., Curtis, P.S., Drake, B.G., Falk, M., Fischer, M.L., Foster, D.R., Gu, L., Hadley, J.L., Hollinger, D.Y., Katul, G.G., Litvak, M., Martin, T.A., Matamala, R., McNulty, S., Meyers, T.P., Monson, R.K., Munger, J.W., Oechel, W.C., Paw, U., T, K., Schmid, H.P., Scott, R.L., Sun, G., Suyker, A.E., Torn, M.S., 2008. Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. Agric. For. Meteorol. 148, 1827–1847.
Xiao, J., Zhuang, Q., Law, B.E., Chen, J., Baldocchi, D.D., Cook, D.R., Oren, R., Richardson, A.D., Wharton, S., Ma, S., 2010. A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data. Remote Sens. Environ. 114, 576–591.
Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033.
Xu, L., Saatchi, S.S., Shapiro, A., Meyer, V., Ferraz, A., Yang, Y., Bastin, J.-F., Banks, N., Boeckx, P., Verbeeck, H., Lewis, S.L., Muanza, E.T., Bongwele, E., Kayembe, F., Mbenza, D., Kalau, L., Mukendi, F., Ilunga, F., Ebuta, D., 2017. Spatial distribution of carbon stored in forests of the democratic Republic of Congo. Sci. Rep. 7, 15030.
Yan, F., Wu, B., Wang, Y., 2015. Estimating spatiotemporal patterns of aboveground biomass using Landsat TM and MODIS images in the Mu Us Sandy Land, China. Agric. For. Meteorol. 200, 119–128.
Yu, Y., Saatchi, S., 2016. Sensitivity of L-band SAR backscatter to aboveground biomass of global forests. Remote Sens. 8, 522.
Yu, G., Chen, Z., Piao, S., Peng, C., Ciais, P., Wang, Q., Li, X., Zhu, X., 2014. High carbon dioxide uptake by subtropical forest ecosystems in the East Asian monsoon region. Proc. Natl. Acad. Sci. U. S. A. 111, 4910–4915.
Yun, J., Qian, S.S., 2015. A hierarchical model for estimating long-term trend of atrazine concentration in the surface water of the contiguous U.S. J. Am. Water Resour. Assoc. 51, 1128–1137.
Zarin, D.J., Harris, N.L., Baccini, A., Aksenov, D., Hansen, M.C., Azevedo-Ramos, C., Azevedo, T., Margono, B.A., Alencar, A.C., Gabris, C., Allegretti, A., Potapov, P., Farina, M., Walker, W.S., Shevade, V.S., Loboda, T.V., Turubanova, S., Tyukavina, A., 2016. Can carbon emissions from tropical deforestation drop by 50% in 5 years? Glob. Chang. Biol. 22, 1336–1347.
Zhang, G., Ganguly, S., Nemani, R.R., White, M.A., Milesi, C., Hashimoto, H., Wang, W., Saatchi, S., Yu, Y., Myneni, R.B., 2014. Estimation of forest aboveground biomass in California using canopy height and leaf area index estimated from satellite data. Remote Sens. Environ. 151, 44–56.
Zhang, H., Song, T., Wang, K., Wang, G., Liao, J., Xu, G., Zeng, F., 2015. Biogeographical patterns of forest biomass allocation vary by climate, soil and forest characteristics in China. Environ. Res. Lett. 10, 044014.
Zhang, Y., Yao, Y., Wang, X., Liu, Y., Piao, S., 2017. Mapping spatial distribution of forest age in China. Earth Space Sci. 4, 108–116.
Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., Yu, S., 2016. Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. Remote Sens. 8, 469.
Zheng, D., Rademacher, J., Chen, J., Crow, T., Bresee, M., Le Moine, J., Ryu, S.-R., 2004. Estimating aboveground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA. Remote Sens. Environ. 93, 402–411.