1 **Spatial distribution of forest aboveground biomass in China: estimation**
2 **through combination of spaceborne lidar, optical imagery, and forest**
3 **inventory data**

4 Yanjun Su^{a,b}, Qinghua Guo^{a,b*}, Baolin Xue^a, Tianyu Hu^a, Otto Alvarez^b, Shengli Tao^{a,c},
5 Jingyun Fang^{a,c}

6 ^aState Key Laboratory of Vegetation and Environmental Change, Institute of Botany,
7 Chinese Academy of Sciences, Beijing 100093, China

8 ^bSierra Nevada Research Institute, School of Engineering, University of California at
9 Merced, Merced, CA 95343, USA

10 ^cDepartment of Ecology, College of Urban and Environmental Sciences, and Key
11 Laboratory for Earth Surface Processes of the Ministry of Education, Peking University,
12 Beijing 100871, China

13 **Corresponding Author:**

14 Qinghua Guo

15 School of Engineering

16 University of California at Merced

17 5200 N. Lake Road, Merced, CA 95348, USA

18 Phone: +1(209)201-5018

19 Email: guo.qinghua@gmail.com

Spatial distribution of forest aboveground biomass in China: estimation through combination of spaceborne lidar, optical imagery, and forest inventory data

Abstract

The global forest ecosystem, which acts as a large carbon sink, plays an important role in modeling the global carbon balance. An accurate estimation of the total forest carbon stock in the aboveground biomass (AGB) is therefore necessary for improving our understanding of carbon dynamics, especially against the background of global climate change. The forest area of China is among the top five globally. However, because of limitations in forest AGB mapping methods and the availability of ground inventory data, there is still a lack in nationwide wall-to-wall forest AGB estimation map for China. In this study, we collected over 8000 ground inventory records from published literatures, and developed an AGB mapping method using a combination of these ground inventory data, Geoscience Laser Altimeter System (GLAS)/Ice, Cloud, and Land Elevation Satellite (ICESat) data, optical imagery, climate surfaces, and topographic data. An uncertainty field model was introduced into the forest AGB mapping procedure to minimize the influence of plot location uncertainty. Our nationwide wall-to-wall forest AGB mapping results show that the forest AGB density in China is 120 Mg/ha on average, with a standard deviation of 61 Mg/ha. Evaluation with an independent ground inventory dataset showed that our proposed method can accurately map wall-to-wall forest AGB across a large landscape. The adjusted coefficient of determination (R^2) and root-mean-square error between our predicted results and the validation dataset were 0.75 and 42.39 Mg/ha, respectively. This new method and the resulting nationwide wall-to-

43 wall forest AGB map will help to improve the accuracy of carbon dynamic predictions in
44 China.

45 *Keywords: forest aboveground biomass, GLAS/ICESat, lidar, ground inventory, China*

1. Introduction

Global forest ecosystems provide a large carbon sink, which plays an important role in the global carbon balance (Pan et al., 2011). Overall, forest ecosystems cover ~30% of the land surface, accounting for ~75% of terrestrial gross primary production and ~80% of global plant biomass (Beer et al., 2010; Bonan, 2008; Kindermann et al., 2008). It is therefore necessary to accurately estimate the current distribution of, and temporal variations in, the forest carbon stock in aboveground biomass (AGB) to obtain a clearer understanding of carbon dynamics against the background of global climate change (Galbraith et al., 2010; Keith et al., 2009; Pan et al., 2011).

The forest area of China is among the top five globally and covers 20.36% of the country's total area (State Forestry Administration of China, 2013). In addition, because of persistent afforestation and reforestation efforts (e.g., Grain for Green Program, GGP) since the 1950s, the country's forest area has increased by $0.73 \times 10^6 \text{ km}^2$ (~60%) since the 1970s (State Forestry Administration of China, 2013). The forest terrestrial ecosystem of China is a large carbon sink and contributes significantly to national and global carbon storage (Fang et al., 2001; Piao et al., 2009). Recent research has highlighted the substantial increase in carbon stock as a result of afforestation and reforestation (Fang et al., 2014; Liu et al., 2014; Xiao, 2014). Modeling results show that sequestration of an additional 110.45 Tg of carbon is expected by the 2020s, as a result of the GGP (Liu et al., 2014). Substantial changes in forest areas also result in large spatial and temporal variations in nationwide carbon stocks (Guo et al., 2013). However, it is still a challenge to estimate the carbon stocks in AGB, because of the lack of efficient and large-scale practical methods.

Generally, forest AGB can be estimated through three available methods: model-based simulations ([Bergh et al., 1998](#); [White et al., 2000](#)), measurements from traditional ground inventories ([Botkin & Simpson, 1990](#); [Botkin et al., 1993](#); [Fang et al., 1998](#)), and retrievals from remote-sensing datasets ([Ghasemi et al., 2011](#); [Mitchard et al., 2011a](#)). Model-based simulation methods usually provide forest AGB estimations from local to global scales based on model inputs (e.g., radiation, climate surfaces, and elevations) instead of the actual forest AGB distribution ([Iverson et al., 1994](#); [Lu, 2006](#)). Traditional forest inventory methods (e.g., direct harvest methods and indirect allometric modelling methods) can provide reliable information on biomass at local or regional scales ([Fang et al., 2001](#); [Malhi et al., 2002](#)). There have been many national forest inventory programs conducted across the world to provide accurate information for forest management, such as the National Forest Inventories program conducted in the Nordic countries ([Tomppo et al., 2010](#)). However, taking ground measurements is labor intensive and expensive when used for large areas, and is time consuming for nationwide forest survey ([Houghton, 2005](#)). For example, in China, nationwide forest inventories provide reliable information on forest AGB, but these inventories are usually conducted on a five-year cycle at the nationwide scale, and require extensive human and economic resources.

Compared with the forest inventory approach, remote-sensing techniques significantly improve the efficiency of forest AGB mapping in large areas and areas that are difficult to access ([Lu et al., 2005](#); [Powell et al., 2010](#)). By linking with ground inventory data, forest AGB can be estimated from remote sensing datasets using statistical models. Typically, passive optical remote sensing [e.g., Moderate Resolution Imaging Spectroradiometer (MODIS) and Landsat Thematic Mapper (TM)] and radar techniques

92 [e.g., phased array L-band synthetic aperture radar (PALSAR) and Shuttle Radar
93 Topography Mission (SRTM)] have become primary data sources for estimating forest
94 AGB, because of their availability ([Dong et al., 2003](#); [Ghasemi et al., 2011](#); [Koch, 2010](#);
95 [Mitchard et al., 2011a](#); [Rauste, 2005](#); [Soenen et al., 2010](#)). However, the retrieved AGB
96 values using these approaches are usually fraught with saturation effects, because of their
97 limited penetration in vegetated areas ([Baccini et al., 2008](#); [Luckman et al., 1997](#)). The
98 saturation points for optical remote sensing range from 15 to 70 Mg/ha, and those for
99 radar range from 30 to over 300 Mg/ha based on the use of different wavelengths ([Lu,](#)
100 [2006](#); [Mitchard et al., 2009](#), [Myneni et al., 2001](#); [Sader et al., 1989](#); [Woodhouse et al.,](#)
101 [2012](#)).

102 Another alternative method, light detection and ranging (lidar), an active remote sensing
103 technique, can penetrate the forest canopy effectively, because it uses a focused short-
104 wavelength laser pulse ([Su & Guo, 2014](#)). This technique has shown great potential in
105 mapping forest AGB by providing accurate estimates of tree metrics such as tree height,
106 which are closely linked to AGB ([Boudreau et al., 2008](#); [Clark et al., 2004](#); [Lefsky et al.,](#)
107 [2005](#); [Popescu et al., 2011](#)). Lidar does not suffer from saturation effects, even at high
108 AGB levels, and therefore far exceeds the capabilities of optical passive and radar remote
109 sensing in mapping forest AGB ([Clark et al., 2011](#); [Nelson et al., 2009](#); [Swatantran et al.,](#)
110 [2011](#)). However, neither of the two major lidar platforms, airborne lidar and spaceborne
111 lidar, provides complete global coverage of land surfaces. Currently, airborne lidar data
112 are only available for certain small areas across the world. The Geoscience Laser
113 Altimeter System (GLAS), onboard the Ice, Cloud, and Land Elevation Satellite (ICESat)

is the only available spaceborne lidar system, and its footprints (with a nominal diameter of ~65 m) are spaced at 170 m along tracks and tens of kilometers across tracks.

By taking advantages of each type of data, reliable regional or even global AGB products can be obtained by combining multisource datasets. For example, [Boudreau et al. \(2008\)](#) explored the possibility of using GLAS data, Landsat-derived land-cover maps, and SRTM data to map the forest AGB distribution in Quebec, Canada at the patch level. [Baccini et al. \(2008\)](#) mapped the forest AGB distribution in Africa at a resolution of 1 km by combining GLAS and MODIS data. [Mitchard et al. \(2011b\)](#) and [Saatchi et al. \(2011\)](#) mapped the forest AGB distributions in tropical areas using GLAS and PALSAR data, and GLAS, MODIS, and radar data, respectively.

Studies have also been conducted in parts of China based on combining data from optical imagery, lidar, and other datasets. A recent study by [Zhang et al. \(2014c\)](#) mapped the AGB in northeastern China by integrating field data, GLAS data, and MODIS products using regression models; [Neigh et al. \(2013\)](#) estimated the total AGB for the entire circumboreal forest biome, which also included the northern China, by incorporating ground inventory, airborne lidar data and spaceborne lidar data. These pioneering studies prove the validity of GLAS-derived metrics in the regional-level estimation of forest AGB. However, to the best of our knowledge, few published studies have explored the nationwide forest AGB distribution of China using spaceborne lidar. One of the main obstacles is the lack of ground inventory data with which to build linkages with the GLAS-derived metrics across the country.

In this study, we developed a forest AGB estimation procedure using a combination of ground inventory data, spaceborne lidar, optical imagery, climate surfaces, and

topographic data. Over 8000 plot measurements across the country were collected from the published literature, and used to map the nationwide forest AGB at a 1 km resolution. The plot location was usually recorded at a 1–10 km accuracy, and therefore the influence of the plot location uncertainty was further considered in the forest AGB mapping procedure. The wall-to-wall map of forest AGB is downloadable from <http://faculty.ucmerced.edu/qguo/carbon>. The resulting forest AGB map of China was evaluated based on independent plot measurements, and compared with other forest AGB estimation results for different parts of China. This new method and the resulting AGB map will help to improve the accuracy of carbon dynamic predictions in China.

2. Data and methods

2.1 Forest inventory data

Forest ground inventory data are fundamental in retrieving AGB from remotely sensed datasets. In this study, we collected 8429 records of plot measurements (including both research plots and forest inventory plots) from previously published papers. The collected plot measurements covered both plantations and natural forests. The geolocation of each individual plot record and its corresponding attributes (including stand origin, measurement method, data published year, and AGB value) were included in the supplemental materials. Since these in-situ plot measurements were collected from various data sources and were taken under different standards, this study further used four filtering criteria to ensure their quality: (1) they should be georeferenced; (2) they should be larger than 0.05 ha; (3) they should have been measured after 2000; and (4) they should not have been surveyed using harvesting methods. Records with the same

geolocations but measured at different years were further averaged. Finally, 1065 plot samples were retained for the following forest AGB mapping procedures (Fig. 1).

-----Please insert Fig. 1 here-----

2.2 ICESat GLAS laser altimetry data

The NASA (National Aeronautics and Space Administration) GLAS instrument, onboard the ICESat satellite, was launched on January 12, 2003; it was designed to have a 183-day ground track repeating cycle. It was equipped with three laser sensors, L1–L3, and 1064 nm laser pulses operated at 20 Hz were used to record the returned full-waveform altimetry data within ~65 m ellipsoidal footprints. The footprints were spaced at 170 m along the track and tens of kilometers across tracks (Schutz et al., 2005). The returned waveform over land had 544 rang bins with 15 cm or 60 cm intervals, corresponding to a vertical range of 81.6 m or 150 m above the ground.

Since most of the collected plot measurements were obtained around 2004, we selected the GLAS data during the operating periods L2B, L2C, and L3A (from February 17, 2004 to November 09, 2004) for further AGB mapping procedures. Three GLAS products, GLA01, GLA06, and GLA14, were collected from the ICESat/GLAS data pool (<http://nsidc.org/data/icesat/order.html>); they provided full-waveform information, geolocation and data quality information, and surface elevation information, respectively. These three products were linked together using the unique ID and shot time of each laser pulse. For the GLAS footprints within the study area, four further filtering criteria were used to ensure the quality of GLAS measurements: 1) they should be taken under cloud free conditions; 2) they should have no saturation effects; 3) they should have high signal

to noise ratios (i.e., >50); and 4) they should not be significantly higher (i.e., < 100 m) than the land surface elevation denoted by the SRTM data.

The final GLAS dataset contained 629 075 full-waveform records across China. The sampling density was generally higher in flat and forested areas (Fig. 2). For each retained GLAS record, three parameters were derived from its full-waveform information, i.e., the waveform extent, leading edge extent, and trailing edge extent, which have been shown to be highly correlated with canopy height, canopy height variability, and terrain slope, respectively (Boudreau et al., 2008; Lefsky et al., 2007; Lefsky, 2010; Su et al., 2015). The definitions of these three parameters can be found in Lefsky et al. (2007), and will not be discussed in detail here.

-----Please insert Fig. 2 here-----

2.3 MODIS-Terra MOD13A2 NDVI data

The MODIS-Terra MOD13A2 data are calculated from the MODIS atmospherically corrected bi-directional surface reflectance product, and are provided globally at a 1 km resolution every 16 d. Water bodies, clouds, aerosols, and cloud shadows are masked in the product. In this study, we collected time-series MOD13A2 data during the growing season of the year 2004 (from May to September). The use of cumulative normalized difference vegetation index (NDVI) from time-series NDVI can increase the AGB estimation accuracy compared with the use of NDVI data from a single time period (Li et al., 2015). Therefore, we computed the cumulative NDVI from the sum of all collected MOD13A2 data, and used it as a predictor in the AGB mapping procedure.

2.4 Topographic data

SRTM, a joint mission conducted by NASA and the National Geospatial-Intelligence Agency, provides a digital elevation model (DEM) product that covers 99.97% of the Earth's land surface, from 56° S to 60° N (U.S. Geological Survey, 2013). Its designed accuracy is 20 m horizontally and 16 m vertically. In this study, we used the second version of the SRTM data for China at a resolution of 3 arcsec (often quoted as 90 m), which show well-defined water bodies and coastlines, and the absence of spikes and wells (U.S. Geological Survey, 2013). To be consistent with other datasets, the SRTM DEM was resampled to 1 km resolution for further AGB mapping procedures. The slope product (denoted by tangent values of slope) was also calculated from the resampled SRTM DEM.

2.5 Climate surfaces

Four climate surfaces, namely annual mean temperature, annual temperature seasonality, annual total precipitation, and annual precipitation seasonality between 1950 and 2000 were calculated at 1 km resolution using data from 2888 weather stations (1512 stations with temperature records and 1376 stations with precipitation records) across China. These weather station data were obtained from the Food and Agriculture Organization FAOCLIM 2.0 (http://www.fao.org/nr/climpag/pub/en1102_en.asp) and the Global Historical Climate Network Dataset (GHCN) version 2 (Peterson & Vose, 1997). All the weather station data were manually checked to remove as many outliers or human errors as possible.

To generate these four climate surfaces, monthly mean temperature and monthly total precipitation layers from 1950 to 2000 were interpolated from the weather station data using the thin plate spline algorithm, with the resampled 1 km resolution SRTM DEM as

the covariable (Alvarez et al., 2014). Then, for each year, the mean temperature and total precipitation were directly calculated from the monthly layers, and the temperature seasonality and precipitation seasonality were computed from the following equation (Xu & Hutchinson, 2011):

$$Seasonality = 100 \times SD_{monthly} / Mean_{monthly} \quad (1)$$

where *Seasonality* represents the temperature seasonality or precipitation seasonality of the corresponding year, and *SD_{monthly}* and *Mean_{monthly}* are the standard deviation and mean of the monthly temperature (in Kelvin) or monthly precipitation (in millimeters) for the corresponding year. Finally, the annual mean temperature, annual temperature seasonality, annual total precipitation, and annual precipitation seasonality were derived from the average of the corresponding yearly products.

2.6 Auxiliary data

In addition to the above-mentioned data, two auxiliary datasets, i.e., a 1:1 000 000 vegetation map of China (Hou, 2001) and a 1 km land-use map of China at a 1:100 000 scale from the year 2000 land-use database (Liu et al., 2002), were used in this study. The vegetation map divided China into eight vegetation zones, based on the dominant vegetation type (Fig. 1). The land-use map classified the land cover into six major categories (i.e., cropland, forest, grass land, water bodies, artificial areas, and bare earth) and 23 subgroups, based on the interpretation of Landsat TM images. Within each subgroup, the percentage of corresponding land-use type aggregated from the 30 m Landsat TM images was provided. In this study, we were particularly interested in two

subgroups in the forest category, the closed-forested area group and the open-forested area group.

2.7 Extrapolation of GLAS-derived parameters

The GLAS-derived parameters, i.e., waveform extent, leading edge extent, and trailing edge extent, normally do not have direct biological meanings (Lefsky et al., 2007). Additional forest parameter data derived from airborne lidar or field plot measurements are usually required to convert GLAS parameters to biologically meaningful parameters (e.g., canopy height and AGB) (Boudreau et al., 2008; Lefsky, 2010; Saatchi et al., 2011). However, airborne lidar data are only available for a few small areas, because of limitations in terms of flight mission cost and mission duration. Moreover, in our cases, the probability of the collected plot measurements overlapping with GLAS footprints was too small, because of the huge gap between two adjacent GLAS tracks (Fig. 2), and the considerable uncertainty within the plot location further increased the difficulty of matching the GLAS footprints with plot measurements.

In this study, to relate the GLAS parameters to plot measurements, we extrapolated the three GLAS parameters into spatial continuous layers using a regression method. GLAS footprints that did not fall in the closed-forested area group or the open-forested area group from the land-use map (Liu et al., 2002) were excluded to minimize the influence of non-forested areas on the extrapolation results. In total, 202 298 GLAS footprints were retained for the subsequent extrapolation process. Furthermore, because all the collected raster datasets were at (or resampled to) 1 km resolution, we aggregated the retained GLAS footprints into 87 893 1 km pixels, and the averages of the three GLAS-derived

parameters within each pixel were used in the final GLAS parameter extrapolation procedure.

The regression tree method Random Forest (RF) (Breiman, 2001) was used to extrapolate the three GLAS-derived parameters. RF, a formalized non-parametric machine-learning algorithm, has been successfully used in areas such as biomass mapping (Baccini et al., 2008) and niche modeling (Prasad et al., 2006). Simard et al. (2011) showed that RF was robust in extrapolating RH100, one of the GLAS-derived parameters, to a spatial continuous layer. One of the main advantages of RF is that it does not require an assumption to be made regarding the normality of covariables (Breiman, 2001), which is often violated when complex ecological systems and environmental variables are introduced (Saatchi et al., 2011). Also, RF can minimize the within-group variance at the expense of a small increase in the bias and overcome the overfitting habit of decision tree algorithms (e.g., Classification and Regression Tree algorithm) (Breiman, 2001; Friedman et al., 2001), because of the use of its unique tree “bagging” algorithm, which selects a random subset of covariables at each candidate split.

Seven ancillary predictors were used in the RF regression tree modeling process, namely cumulative NDVI, elevation, slope, and the four climate surfaces. Moreover, each vegetation zone from Hou (2001) was represented by a unique numerical identification number and fed into the RF regression tree model. The RF extrapolation method was implemented using the randomForest R package (Liaw & Wiener, 2002), which includes both classification and regression functions. In this study, 500 “RF trees” were included and four variables were tried at each split, based on manually iterative examination.

2.8 Forest AGB estimation

The 1065 plot measurements were randomly divided into a training dataset and validation dataset, at a ratio of 3:1 (Fig. 1). The 799 training plot measurements along with the three GLAS metrics and the seven collected predictors were used to estimate forest AGB using the RF regression tree algorithm (Fig. 3). As in the GLAS extrapolation procedure, the regression RF forests were built by relating the training plot measurements to the prediction variables. Similar to the GLAS parameter extrapolation process, 500 RF trees and four prediction variables tried at each split were selected based on manually iterative examination.

-----Please insert Fig. 3 here-----

To improve the computational efficiency and remove redundant information, we explored the importance of all prediction metrics and determined the optimal variables for forest AGB mapping. The percentage increase in the mean-squared error (%IncMSE) and the increase in node purity (IncNodePurity) were calculated to evaluate the variable importance (Liaw & Wiener, 2002). The larger the %IncMSE and the IncNodePurity of a variable are, the more important this variable is. As can be seen from Fig. 4, regardless of whether the evaluation was performed using the %IncMSE or the IncNodePurity, the absence of the three GLAS-derived metrics and topographic data significantly increased the mean-squared error and node purity, indicating a decrease in the prediction accuracy of the built RF regression tree model. The three GLAS-derived metrics and topographic data were therefore used as the prediction variables for mapping the AGB distribution. Note that considering the different characteristics of trees from different vegetation zones (e.g., tree species, tree size and tree density), the vegetation zone data was also used in the AGB mapping process.

-----Please insert Fig. 4 here-----

Additionally, when relating the ground inventory data to the predictors, the plot location uncertainties were too large to be neglected. Without considering the surveying accuracy, most of the latitudes and longitudes given by the inventory data were accurate to 0.01° (corresponding to ~ 1 km), and some of them were only accurate to 0.1° (corresponding to ~ 10 km). These huge plot location uncertainties could result in significant differences in the corresponding values of predictor variables, and therefore influence the forest AGB estimation result. To minimize the influence of plot location uncertainty, we introduced an uncertainty field model (Guo et al., 2008) into the RF AGB mapping procedure (Fig. 3). This method hypothesized that the real plot center was randomly located within a circular buffer zone of the provided plot location, and the radius of the buffer was determined by the corresponding plot location uncertainty. In this study, by assuming that the plot location cannot be more than 1 or 10 km (determined by the accuracy of the given plot location) away from the given location in any direction, we created a 1 or 10 km buffer around each plot. Then 100 sets of different ground inventory data were randomly generated within the buffers.

The AGB mapping procedure based on the RF regression tree algorithm was performed for each set of ground inventory data with a location uncertainty (Fig. 3). We generated 100 predicted forest AGB layers using this process, and the average of these forest AGB layers was taken as the initial forest AGB estimation result. The final forest AGB mapping result was computed from the initial forest AGB estimation result by setting the non-forested pixels at 0 Mg/ha. In this study, if a 1 km pixel was not covered by either of the two forested groups from the land-use map (i.e., the closed-forested area group and

the open-forested area group), it was treated as a non-forested pixel, and vice versa. It should be noted that the RF prediction process for each run was operated separately on each vegetation zone using the same RF regression tree model obtained from the corresponding run, considering the computation efficiency. The Qinghai-Tibet Plateau alpine vegetation and the temperate steppe vegetation zones were merged together and treated as one zone in this process, which were both dominated by grassland and shrubs and had relatively small forested areas.

2.9 Accuracy assessment

The accuracy of the estimated forest AGB of China was evaluated at the plot level, pixel level, and vegetation zone level. The plot-level accuracy was assessed by directly comparing the estimated result with an independent validation ground inventory dataset (Fig. 3). The adjusted coefficient of determination (R^2) and root-mean-square error (RMSE) were computed using the following equations:

$$R^2 = 1 - \frac{(n-1) \sum_{i=1}^n (B_i - \hat{B}_i)^2}{(n-2) \sum_{i=1}^n (B_i - \bar{B})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (B_i - \hat{B}_i)^2}{n-2}} \quad (3)$$

where B_i is the observed AGB from the validation plot, $\hat{B}_i$ is the modeled AGB at the plot location from the wall-to-wall AGB map, $\bar{B}$ is the average AGB of all validation plots, and n is the number of validation plots.

356 The pixel-level uncertainty was only evaluated by the uncertainty induced by the plot
 357 location ($\epsilon_{location}$), which was calculated from the standard deviation of 100 AGB
 358 estimation iterations:

$$359 \quad \epsilon_{location} = \sqrt{\frac{\sum_{j=1}^{100} (\hat{B}_j - \bar{\hat{B}})^2}{100}} \quad (4)$$

360 where $\hat{B}_j$ is the modeled AGB at the j^{th} iteration and $\bar{\hat{B}}$ is the average of all modeled
 361 AGBs. Finally, the uncertainty at the vegetation zone scale was estimated by the RMSE
 362 between the estimated AGB and the independent validation plot inventory data for each
 363 vegetation zone.

364 **3. Results**

365 *3.1 Extrapolated GLAS parameters*

366 The GLAS-derived waveform extent, leading edge extent, and trailing edge extent were
 367 extrapolated to spatially continuous layers using the RF regression tree method (Fig. 5).
 368 Overall, the built RF regression models can explain around 57%, 52%, and 46% of the
 369 variances in the waveform extent, leading edge extent, and trailing edge extent,
 370 respectively. The root-mean-square residuals for the extrapolated waveform extent,
 371 leading edge extent, and trailing edge extent are 24 m, 20 m, and 6 m, respectively. The
 372 waveform extent and leading edge extent share similar spatial patterns. Values in
 373 southern China are generally higher than those in northern China (Fig. 5a, b). In the
 374 tropical monsoon forest–rain forest located in southern Tibet, both the waveform extent
 375 and leading edge extent have significantly higher values than in other areas. Compared

with the waveform extent and the leading edge extent, the trailing edge extent is more evenly distributed across China.

-----Please insert Fig. 5 here-----

3.2 Forest AGB estimation in China

The wall-to-wall forest AGB map is shown in Fig. 6 and can be downloaded from <http://faculty.ucmerced.edu/qguo/carbon>. The RF regression tree model built from plot measurements without considering location uncertainties can explain 72% of variances in forest AGB and the root-mean-squared residual is 40.12 Mg/ha. Overall, the forest AGB density is higher in southern China than in northern China (Fig. 6). The average forest AGB density for the whole China (without considering non-forested areas) is about 120 Mg/ha, and the standard deviation is 61 Mg/ha.

-----Please insert Fig. 6 here-----

Northern China is mainly occupied by the cold temperate needleleaf forest, temperate needleleaf–broadleaf mixed forest, warm temperate deciduous broadleaf forest, temperate steppe, and temperate desert (Fig. 1). The mean AGB density for the cold temperate needleleaf forest is around 75 Mg/ha, and is more homogeneously distributed. Over 50% of such forested areas have an AGB density range of 70–80 Mg/ha (Fig. 7). The warm temperate needleleaf–broadleaf mixed forest has an average AGB density of around 90 Mg/ha, and over a half of such forested areas have an AGB density range of 80–120 Mg/ha. Although the warm temperate deciduous broadleaf forest is in the south of the temperate needleleaf–broadleaf mixed forest, its average AGB density is smaller, and close to that of the cold temperate needleleaf forest (Fig. 7). This may be due to the fact

that these areas have been heavily disturbed by human activities, and most of these forests are plantations instead of natural forests. The average AGB density for the temperate steppe is the lowest among all vegetation zones (~65 Mg/ha) (Fig. 7), and its proportion of forest is less than 9%. The average AGB density of the temperate desert is about 100 Mg/ha, and over 50% of this area is within the range 50 to 150 Mg/ha. Most forests in this vegetation zone are distributed within the Junggar Basin (Fig. 6).

Southern China is mainly covered by subtropical evergreen broadleaf forest, tropical monsoon forest–rain forest, and Qinghai-Tibet Plateau alpine vegetation. All these vegetation zones have significantly higher average AGB densities than those of the vegetation zones in northern China ($p < 0.0001$) (Fig. 7). Specifically, the average AGB density for the subtropical evergreen broadleaf forest is about 140 Mg/ha, and the maximum AGB density can reach over 330 Mg/ha. The tropical monsoon forest–rain forest has the highest mean forest AGB density among all vegetation zones, and has a broader range of AGB density distributions (Fig. 7). Its median AGB density is over 200 Mg/ha, and about half of the area is within the AGB density range 100–250 Mg/ha. The average AGB density for the Qinghai-Tibet Plateau alpine vegetation zone is about 150 Mg/ha, and is mainly contributed by forested areas in the most southeastern part of the Qinghai-Tibet Plateau (Fig. 6).

-----Please insert Fig. 7 here-----

The accuracy of the wall-to-wall forest AGB estimation in China was evaluated using 266 independent validation plot measurements. As shown in Fig. 8, the R^2 between the predicted and field-measured AGB is 0.75 and the RMSE is 42.39 Mg/ha. Although the slope and intercept for the correlation between estimated and field measured AGB are

larger than one and smaller than zero, the fitted line is still very close to the 1:1 line (Fig. 8). The negative intercept suggests that the proposed AGB estimation method tends to slightly overestimate AGB densities at areas with low values (<87 Mg/ha). After the forest AGB density reaching 87 Mg/ha, the proposed AGB estimation method tends to underestimate the forest AGB density, and this underestimation effect becomes more pronounced with the increase of forest AGB density.

-----Please insert Fig. 8 here-----

The estimated mean AGB density for each vegetation zone was further compared with that calculated from all 1065 plot measurements in each vegetation zone. As the data in Table 1 show, the differences between the predicted mean AGB density and plot-measured mean AGB density for most vegetation zones are smaller than 10 Mg/ha. The temperate desert has the largest difference, i.e., ~20 Mg/ha. The RMSE for each vegetation zone, calculated from the validation plot measurements, is shown in Table 1. As can be seen, the RMSEs for most vegetation zones are within the range 23–33 Mg/ha. The subtropical evergreen broadleaf forest and Qinghai-Tibet Plateau alpine vegetation have relatively large RMSEs (over 45 Mg/ha). The high RMSE for the Qinghai-Tibet alpine vegetation may be caused by the small number of validation plots (only three) located in this vegetation zone. The temperate desert has the smallest RMSE (~14 Mg/ha) among all vegetation zones.

-----Please insert Table 1 here-----

4. Discussion

4.1 Linking ground plots and GLAS data

In this study, we collected over 8000 records of forest AGB field measurements and developed a procedure to estimate wall-to-wall forest AGB distributions in China, using a combination of plot measurements, GLAS data, optical imagery, climate surfaces, and topographic data. Because the GLAS data are spatially discontinuous, how to link plot measurements to GLAS footprints is one of the key issues in estimating forest AGB distribution. Generally, there are three methods for addressing this issue: 1) direct link based on the geolocation (Saatchi et al., 2011; Zhang et al., 2014c); 2) use of airborne lidar data as a medium (Boudreau et al., 2008); and 3) extrapolation of GLAS parameters (Zhang et al., 2014b).

Direct linkage between the ground inventory and GLAS measurements at the same geolocation is the most intuitive matching method. However, in our cases, about 80% of the plots are over 1 km away from their closest GLAS footprint. The discrepancy between their locations means that they can hardly be matched. Instead of direct linkage, airborne lidar has been used as the medium between ground inventory and GLAS measurements. Airborne lidar has proved to be capable of estimating forest AGB based on plot measurements (Boudreau et al., 2008). The use of airborne lidar data obtained by selectively flying in portions of the study area enables the forest AGB within the footprints to be accurately estimated and therefore used as ground truth data to estimate the AGB within the GLAS footprints. However, currently, the availability of airborne lidar data in China is very limited because of the high flight mission costs.

Instead of trying to link the plot AGB and GLAS parameters at the GLAS footprint level, Zhang et al. (2014a) first extrapolated the GLAS-derived vegetation heights to continuous layers using a TM-derived leaf area index (LAI) product, and then estimated

the AGB distribution from the vegetation height in California, USA at a 30 m resolution. The TM-derived LAI was the only predictor used to generate the continuous vegetation height, therefore their final AGB estimation relied heavily on the accuracy of the LAI product. However, previous studies have shown that the TM-derived LAI data are fraught with uncertainty issues because of the spectral saturation effect (Lu et al., 2004; Wang et al., 2005). This was also indicated by their results, which showed that the estimated forest AGB had a large uncertainty, ranging from 40 to 150 Mg/ha (Zhang et al., 2014a). Simard et al. (2011) showed the feasibility of deriving vegetation height by extrapolating RH100, a GLAS-derived parameter, to a global continuous layer using an RF regression tree algorithm. This greatly improved the potential of GLAS data in the estimation of global AGB. However, it has been reported that RH100 underestimates the canopy height (Sun et al., 2008) and careful calibration may be needed for further forest AGB estimation (Zhang et al., 2014a).

In this study, we extrapolated the GLAS-derived waveform extent, leading edge extent, and trailing edge extent to spatially continuous products, using an RF regression tree model similar to that used by Simard et al. (2011). These three parameters have been successfully used to estimate the vegetation height and AGB at the GLAS footprint level (Boudreau et al., 2008; Lefsky et al., 2005). In this study, these three extrapolated GLAS metrics along with topographic data can explain over 72% variances of the forest AGB, and the forest AGB distribution can be mapped with high accuracy by linking the plot-measured AGB to them (Fig. 8). This shows the feasibility of using extrapolated GLAS full-waveform parameters to estimate forest AGB.

Moreover, the plot location uncertainty has rarely been considered when linking plot measurements to GLAS parameters. For the plot measurements collected from the literature in particular, the plot location uncertainty can be as high as 10 to 15 km, without considering the actual surveying uncertainty. This huge location uncertainty can result in significantly different correspondences between field-measured AGB and predictors. In this study, as a result of introducing plot location uncertainty into the forest AGB modeling procedure, R^2 between the predicted AGB and plot validation dataset increased from 0.64 to 0.75, and the RMSE decreased from 50 to 42 Mg/ha. At the pixel level, the absolute value of the uncertainty introduced by the plot location increased with forest AGB density, and contributed to around 10% of the final AGB estimation (Fig. 9). In the tropical monsoon forest–rain forest in the most southeastern part of Tibet, the uncertainty brought by the plot location can reach nearly 40 Mg/ha.

-----Please insert Fig. 9 here-----

4.2 Influence of different forest definitions on total forest AGB estimation

Usually, forest extent is defined as an area with a tree cover higher than a fractional cover threshold. However, the tree cover threshold used to define forest extent varies with time and place. Tree covers of 10%, 25%, and 30% are the most frequently used thresholds used by individual parties to the United Nations Framework Convention on Climate Change. This discrepancy in the definition of a forest may result in significantly different total forest areas, and therefore influence the estimation of total biomass carbon stock in the forest.

Table 2 shows the total forest areas and total AGB estimations for different forest definitions. As can be seen, the total forest area for each vegetation zone decreases

significantly with increasing tree cover threshold used to define the forest extent. The total forest area from the original Landsat TM land-use map is about 7% larger than the total forest area when a 30% tree cover threshold is applied on the Landsat TM land-use map to define the forest extent. This difference in total forest area can lead to a drop of ~5% in the total AGB estimation for China, which equals about 1000 million tons of forest biomass. The drop effect of the total forest AGB is greater in climate zones with relatively small total forest areas. For example, the temperate desert and Qinghai-Tibet Plateau alpine vegetation have a very small proportion of forest cover, and differences in the definitions of forest extent can result in a decrease of over 16% in the total forest AGB estimation.

-----Please insert Table 2 here-----

4.3 Comparison of estimated forest AGB and published results

We compared our nationwide forest AGB map with previously published AGB mapping results covering different areas of China to further evaluate our result. For northeastern China, [Zhang et al. \(2014c\)](#) estimated that the average forest AGB density was 83.50 Mg/ha, and the total forest AGB carbon stock was 1.55 Pg. Our results indicate that the average forest AGB density in northeastern China is ~88.50 Mg/ha, and the total forest AGB carbon stock is 1.5 Pg (using a ratio of 50% to convert forest AGB to forest AGB carbon stock) ([Saatchi et al., 2011](#)). This total AGB carbon stock estimation is also consistent with the value of 1.4–1.6 Pg of carbon obtained by simulations using the TRIPLEX 1.0 model ([Peng et al., 2009](#)). The provincial average forest AGB density from our results is also close to that estimated by [Guo et al. \(2013\)](#) for northeastern China, which was estimated based on the national forest inventory data of China. The average

forest AGB densities for Inner Mongolia, Liaoning, Jilin, and Heilongjiang provinces from our wall-to-wall AGB map are 67, 77, 93, and 81 Mg/ha, respectively, and those from their results are 77, 63, 119, and 85 Mg/ha. However, these values are significantly higher than those reported by [Thurner et al. \(2014\)](#), namely 27, 25, 40 and 33 Mg/ha, respectively. This may be caused by the growing stock volume product retrieved from synthetic aperture radar data, which was used as the predictor in their AGB estimation process. This product was mainly calibrated based on plot measurements in Russia, which might make it less representative for forests in northern China.

For southern China, [Saatchi et al. \(2011\)](#) estimated the forest AGB distribution in tropical regions using a combination of GLAS, MODIS, and radar data, which included forests of China lower than 40° N. Their forest AGB estimation ranges match our nationwide wall-to-wall forest AGB map as well, except in the case of the Qinghai-Tibet Plateau. For the Qinghai-Tibet Plateau, the majority of our estimates are 0 Mg/ha, but their result ranges around 20-80 Mg/ha. This difference may be caused by the different land-use maps used by these two studies. In the land-use map we used, most of the Qinghai-Tibet Plateau was classified as non-forest groups, and our corresponding AGB values were set at 0 Mg/ha.

Besides the regional scale comparison, we also performed a pixel-level comparison between our estimated forest AGB map and that from [Saatchi et al. \(2011\)](#) in southern China (Fig. 10a). It should be noted that the forest AGB map from [Saatchi et al. \(2011\)](#) was masked by the land-use map from [Liu et al. \(2002\)](#) before the comparison. Similar to our forest AGB map, the non-forested areas were set at 0 Mg/ha. The average difference between our and their forest AGB results is -7.23 Mg/ha. The absolute values of differences for 60% of the pixels are smaller than 50 Mg/ha, and for 36% of the pixels are

smaller than 25 Mg/ha. As can be seen in Fig. 10a, our forest AGB result tends to be lower than their result in southeastern China, and higher than their result in Yunnan, Chongqing, Sichuan and Guizhou provinces. In the most southeastern part of the Qinghai-Tibet Plateau, the majority of their forest AGB densities are slightly higher (10-50 Mg/ha) than our estimates. These differences are possibly caused by the fact that their forest AGB result was estimated based on plots from tropical areas across the world and were less representative of forest conditions in China. As shown in Fig10b, we evaluated their AGB mapping result by our collected field measurements. Overall, their result shows a satisfying accuracy; the R^2 and the RMSE are around 0.4 and 68 Mg/ha respectively. However, their result inclines to underestimate forest AGB in areas with low values, and overestimate it in areas with high values (Fig. 10b). In southeastern China, although this region belongs to the subtropical evergreen broadleaf vegetation zone, it has been heavily disturbed by human activities, and the proportion of plantations is relatively higher. The possible lack of plot measurements from these plantation areas may lead to overestimations of forest AGB and therefore make their result higher than our forest AGB map.

-----Please insert Fig. 10 here-----

4.4 Limitations of the current study

Although the resulting nationwide wall-to-wall forest AGB map of this study shows good accuracy by comparing with both independent field measurements and other published products, there are still limitations in the methodology of the current study. The three extrapolated GLAS parameters along with topographic information can be used to explain over 72% of variations in forest AGB, and therefore accurately map the wall-to-

wall forest AGB distribution of China. However, the extrapolated GLAS parameters were imputed based on the regression analysis using selected ancillary datasets (i.e., cumulative NDVI, climate surfaces, and topographic data). Moreover, the spatial resolution of the original GLAS parameters and the extrapolated GLAS parameters are also different. These differences may lead the biophysical meanings of these three GLAS full-waveform parameters to become unclear. Although using spatial interpolation techniques (e.g. ordinary kriging and thin plate spline) to extrapolate GLAS parameters can help to preserve their biophysical meanings, the distribution of GLAS footprints is not random but concentrated around the ground tracks, which can result in significant strip effects to the extrapolated GLAS parameters.

Further, the extrapolation of GLAS parameters may even bring new error sources to the forest AGB estimation step. However, in the current study, the uncertainties induced by the extrapolated GLAS parameters as well as other prediction variables were not considered and evaluated in the forest AGB estimation procedure. The pixel level uncertainty can be evaluated in more detail by introducing a numerical error propagation model based on Monte Carlo simulation ([Zhang et al., 2014b](#)). A systematic evaluation of how uncertainties from different sources influence the final forest AGB distribution result will be conducted in a future study.

Additionally, the influence of forest growth on the AGB estimation result was not considered in this study. The plot measurements collected in this study were mainly taken between 2000 and 2010. The AGB increase brought by the forest growth might be non-negligible for certain plots, and result in mismatches between these plot measurements and remotely sensed covariables. If we only chose plot measurements from 2004 (the

same year with remotely sensed datasets), the number of plot measurements would be too small to build a reliable forest AGB estimation model. In this study, the average of plot measurements with the same geolocations but measured at different years were used to represent the forest AGB condition at that plot. This may be helpful to partially reduce the influence of forest growth on the final AGB estimation. However, collecting more ground inventory data for each calendar year is still needed to fully address this issue.

5. Conclusions

In this study, we developed a method to estimate the forest AGB distribution of China through the combination of multi-source data sets. Over 8000 ground inventory records were collected from the literature to address this mission. Moreover, to minimize the influence of plot location uncertainty, we introduced an uncertainty field model into the forest AGB mapping procedure. Our nationwide wall-to-wall forest AGB estimations show that the forest AGB density is generally higher in southern China than in northern China. The average forest AGB density across China is 120 Mg/ha with a standard deviation of 61 Mg/ha. Among all climate zones, the tropical monsoon forest–rain forest has the highest average forest AGB density. The nationwide wall-to-wall AGB map shows a good correspondence with an independent ground inventory dataset ($R^2=0.75$, RMSE=42.39 Mg/ha).

The plot location uncertainty can induce up to 10% uncertainty in the final forest AGB estimation, and generally the higher the forest AGB density is, the higher uncertainty induced by the plot location uncertainty. By considering the plot location uncertainty into the AGB estimation model, the forest AGB estimation accuracy can be significantly improved. Additionally, the definition of forest is an important factor influencing the

estimation of total forest biomass. The total forest biomass calculated from the forest definition with a 30% tree cover threshold can decrease the total forest biomass estimation by ~1000 million tons. A more reasonable and unified forest definition should be developed and adopted in the future so that equivalent comparisons can be made for studies on carbon dynamics from different study areas and different methods.

Acknowledgements

We thank the United States Geological Survey and National Aeronautics and Space Administration for providing the SRTM DEM products and Data Sharing Infrastructure of Earth System Science for providing the nationwide land-use map of China.

References

- Alvarez, O., Guo, Q., Klinger, R.C., Li, W. & Doherty, P. (2014) Comparison of elevation and remote sensing derived products as auxiliary data for climate surface interpolation. *International Journal of Climatology*, **34**, 2258-2268.
- Baccini, A., Laporte, N., Goetz, S., Sun, M. & Dong, H. (2008) A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environmental Research Letters*, **3**, 045011.
- Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M.A., Baldocchi, D. & Bonan, G.B. (2010) Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science*, **329**, 834-838.

- 646 Bergh, J., McMurtrie, R.E. & Linder, S. (1998) Climatic factors controlling the
647 productivity of Norway spruce: a model-based analysis. *Forest Ecology and*
648 *Management*, **110**, 127-139.
- 649 Bonan, G.B. (2008) Forests and climate change: forcings, feedbacks, and the climate
650 benefits of forests. *Science*, **320**, 1444-1449.
- 651 Botkin, D.B. & Simpson, L.G. (1990) Biomass of the North American boreal forest: A
652 step toward accurate global measures. *Biogeochemistry*, **9**, 161-174.
- 653 Botkin, D.B., Simpson, L.G. & Nisbet, R.A. (1993) Biomass and carbon storage of the
654 North American deciduous forest. *Biogeochemistry*, **20**, 1-17.
- 655 Boudreau, J., Nelson, R.F., Margolis, H.A., Beaudoin, A., Guindon, L. & Kimes, D.S.
656 (2008) Regional aboveground forest biomass using airborne and spaceborne
657 LiDAR in Québec. *Remote Sensing of Environment*, **112**, 3876-3890.
- 658 Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5-32.
- 659 Clark, M.L., Clark, D.B. & Roberts, D.A. (2004) Small-footprint lidar estimation of sub-
660 canopy elevation and tree height in a tropical rain forest landscape. *Remote*
661 *Sensing of Environment*, **91**, 68-89.
- 662 Clark, M.L., Roberts, D.A., Ewel, J.J. & Clark, D.B. (2011) Estimation of tropical rain
663 forest aboveground biomass with small-footprint lidar and hyperspectral sensors.
664 *Remote Sensing of Environment*, **115**, 2931-2942.
- 665 Dong, J., Kaufmann, R.K., Myneni, R.B., Tucker, C.J., Kauppi, P.E., Liski, J.,
666 Buermann, W., Alexeyev, V. & Hughes, M.K. (2003) Remote sensing estimates

667 of boreal and temperate forest woody biomass: carbon pools, sources, and sinks.
668 *Remote Sensing of Environment*, **84**, 393-410.

669 Fang, J., Wang, G.G., Liu, G. & Xu, S. (1998) Forest biomass of China: an estimate
670 based on the biomass-volume relationship. *Ecological Applications*, **8**, 1084-
671 1091.

672 Fang, J., Chen, A., Peng, C., Zhao, S. & Ci, L. (2001) Changes in forest biomass carbon
673 storage in China between 1949 and 1998. *Science*, **292**, 2320-2322.

674 Fang, J., Guo, Z., Hu, H., Kato, T., Muraoka, H. & Son, Y. (2014) Forest biomass carbon
675 sinks in East Asia, with special reference to the relative contributions of forest
676 expansion and forest growth. *Global Change Biology*, **20**, 2019-2030.

677 Friedman, J., Hastie, T. & Tibshirani, R. (2001). *The elements of statistical learning*.
678 Springer, Berlin: Springer series in statistics.

679 Galbraith, D., Levy, P.E., Sitch, S., Huntingford, C., Cox, P., Williams, M. & Meir, P.
680 (2010) Multiple mechanisms of Amazonian forest biomass losses in three
681 dynamic global vegetation models under climate change. *New Phytologist*, **187**,
682 647-665.

683 Ghasemi, N., Sahebi, M.R. & Mohammadzadeh, A. (2011) A review on biomass
684 estimation methods using synthetic aperture radar data. *International Journal of*
685 *Geomatics and Geosciences*, **1**, 776-788.

- 686 Guo, Q., Liu, Y. & Wieczorek, J. (2008) Georeferencing locality descriptions and
 687 computing associated uncertainty using a probabilistic approach. *International*
 688 *Journal of Geographical Information Science*, **22**, 1067-1090.
- 689 Guo, Z., Hu, H., Li, P., Li, N. & Fang, J. (2013) Spatio-temporal changes in biomass
 690 carbon sinks in China's forests from 1977 to 2008. *Science China Life Sciences*,
 691 **56**, 661-671.
- 692 Hou, X. (2001) Vegetation atlas of China. *Chinese Academy of Science, The Editorial*
 693 *Board of Vegetation Map of China. Scientific Press, Beijing*, 113-124.
- 694 Houghton, R. (2005) Aboveground forest biomass and the global carbon balance. *Global*
 695 *Change Biology*, **11**, 945-958.
- 696 Iverson, L.R., Brown, S., Prasad, A., Mitsova, H., Gillespie, A.J. & Lugo, A.E. (1994)
 697 Use of GIS for estimating potential and actual forest biomass for continental
 698 South and Southeast Asia. *Effects of Land-Use Change on Atmospheric CO₂*
 699 *Concentrations* (ed. by V.H. Dale), pp. 67-116. Springer-Verlag, New York.
- 700 Keith, H., Mackey, B.G. & Lindenmayer, D.B. (2009) Re-evaluation of forest biomass
 701 carbon stocks and lessons from the world's most carbon-dense forests.
 702 *Proceedings of the National Academy of Sciences*, **106**, 11635-11640.
- 703 Kindermann, G.E., McCallum, I., Fritz, S. & Obersteiner, M. (2008) A global forest
 704 growing stock, biomass and carbon map based on FAO statistics. *Silva Fennica*,
 705 **42**, 387-396.

706 Koch, B. (2010) Status and future of laser scanning, synthetic aperture radar and
 707 hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal*
 708 *of Photogrammetry and Remote Sensing*, **65**, 581-590.

709 Lefsky, M.A. (2010) A global forest canopy height map from the Moderate Resolution
 710 Imaging Spectroradiometer and the Geoscience Laser Altimeter System.
 711 *Geophysical Research Letters*, **37**, L15401.

712 Lefsky, M.A., Keller, M., Pang, Y., De Camargo, P.B. & Hunter, M.O. (2007) Revised
 713 method for forest canopy height estimation from Geoscience Laser Altimeter
 714 System waveforms. *Journal of Applied Remote Sensing*, **1**, 013537.

715 Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom
 716 Espirito Santo, F., Hunter, M.O. & de Oliveira, R. (2005) Estimates of forest
 717 canopy height and aboveground biomass using ICESat. *Geophysical Research*
 718 *Letters*, **32**, L22S02.

719 Li, L., Guo, Q., Tao, S., Kelly, M. & Xu, G. (2015) Lidar with multi-temporal MODIS
 720 provide a means to upscale predictions of forest biomass. *ISPRS Journal of*
 721 *Photogrammetry and Remote Sensing*, **102**, 198-208.

722 Liaw, A. & Wiener, M. (2002) Classification and regression by randomForest. *R News*, **2**,
 723 18-22.

724 Liu, D., Chen, Y., Cai, W., Dong, W., Xiao, J., Chen, J., Zhang, H., Xia, J. & Yuan, W.
 725 (2014) The contribution of China's Grain to Green Program to carbon
 726 sequestration. *Landscape Ecology*, **29**, 1675-1688.

- 727 Liu, J., Liu, M., Deng, X., Zhuang, D., Zhang, Z. & Luo, D. (2002) The land use and land
728 cover change database and its relative studies in China. *Journal of Geographical*
729 *Sciences*, **12**, 275-282.
- 730 Lu, D. (2005) Aboveground biomass estimation using Landsat TM data in the Brazilian
731 Amazon. *International Journal of Remote Sensing*, **26**, 2509-2525.
- 732 Lu, D. (2006) The potential and challenge of remote sensing-based biomass estimation.
733 *International Journal of Remote Sensing*, **27**, 1297-1328.
- 734 Lu, D., Mausel, P., Brondizio, E. & Moran, E. (2004) Relationships between forest stand
735 parameters and Landsat TM spectral responses in the Brazilian Amazon Basin.
736 *Forest Ecology and Management*, **198**, 149-167.
- 737 Luckman, A., Baker, J., Kuplich, T.M., Yanasse, C.d.C.F. & Frery, A.C. (1997) A study
738 of the relationship between radar backscatter and regenerating tropical forest
739 biomass for spaceborne SAR instruments. *Remote Sensing of Environment*, **60**, 1-
740 13.
- 741 Malhi, Y., Phillips, O., Lloyd, J., Baker, T., Wright, J., Almeida, S., Arroyo, L.,
742 Frederiksen, T., Grace, J. & Higuchi, N. (2002) An international network to
743 monitor the structure, composition and dynamics of Amazonian forests
744 (RAINFOR). *Journal of Vegetation Science*, **13**, 439-450.
- 745 Mitchard, E., Saatchi, S., Lewis, S., Feldpausch, T., Woodhouse, I., Sonké, B., Rowland,
746 C. & Meir, P. (2011a) Measuring biomass changes due to woody encroachment
747 and deforestation/degradation in a forest-savanna boundary region of central

748 Africa using multi-temporal L-band radar backscatter. *Remote Sensing of*
749 *Environment*, **115**, 2861-2873.

750 Mitchard, E., Saatchi, S., Woodhouse, I., Nangendo, G., Ribeiro, N., Williams, M., Ryan,
751 C., Lewis, S., Feldpausch, T. & Meir, P. (2009) Using satellite radar backscatter
752 to predict above-ground woody biomass: A consistent relationship across four
753 different African landscapes. *Geophysical Research Letters*, **36**, L23401.

754 Mitchard, E., Saatchi, S., White, L., Abernethy, K., Jeffery, K., Lewis, S., Collins, M.,
755 Lefsky, M., Leal, M. & Woodhouse, I. (2011b) Mapping tropical forest biomass
756 with radar and spaceborne LiDAR: overcoming problems of high biomass and
757 persistent cloud. *Biogeosciences Discussions*, **8**, 8781.

758 Myneni, R., Dong, J., Tucker, C., Kaufmann, R., Kauppi, P., Liski, J., Zhou, L.,
759 Alexeyev, V. & Hughes, M. (2001) A large carbon sink in the woody biomass of
760 northern forests. *Proceedings of the National Academy of Sciences*, **98**, 14784-
761 14789.

762 Neigh, C.S., Nelson, R.F., Ranson, K.J., Margolis, H.A., Montesano, P.M., Sun, G.,
763 Kharuk, V., Næsset, E., Wulder, M.A., & Andersen, H.-E. (2013) Taking stock of
764 circumboreal forest carbon with ground measurements, airborne and spaceborne
765 LiDAR. *Remote Sensing of Environment*, **137**, 274-287.

766 Nelson, R., Ranson, K., Sun, G., Kimes, D., Kharuk, V., & Montesano, P. (2009)
767 Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote*
768 *Sensing of Environment*, **113**, 691-701.

- 769 Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L.,
770 Shvidenko, A., Lewis, S.L. & Canadell, J.G. (2011) A large and persistent carbon
771 sink in the world's forests. *Science*, **333**, 988-993.
- 772 Peng, C., Zhou, X., Zhao, S., Wang, X., Zhu, B., Piao, S. & Fang, J. (2009) Quantifying
773 the response of forest carbon balance to future climate change in Northeastern
774 China: model validation and prediction. *Global and Planetary Change*, **66**, 179-
775 194.
- 776 Peterson, T.C. & Vose, R.S. (1997) An overview of the Global Historical Climatology
777 Network temperature database. *Bulletin of the American Meteorological Society*,
778 **78**, 2837-2849.
- 779 Piao, S., Fang, J., Ciais, P., Peylin, P., Huang, Y., Sitch, S. & Wang, T. (2009) The
780 carbon balance of terrestrial ecosystems in China. *Nature*, **458**, 1009-1013.
- 781 Popescu, S.C., Zhao, K., Neuenschwander, A. & Lin, C. (2011) Satellite lidar vs. small
782 footprint airborne lidar: Comparing the accuracy of aboveground biomass
783 estimates and forest structure metrics at footprint level. *Remote Sensing of*
784 *Environment*, **115**, 2786-2797.
- 785 Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B. &
786 Ohmann, J.L. (2010) Quantification of live aboveground forest biomass dynamics
787 with Landsat time-series and field inventory data: A comparison of empirical
788 modeling approaches. *Remote Sensing of Environment*, **114**, 1053-1068.

789 Prasad, A.M., Iverson, L.R. & Liaw, A. (2006) Newer classification and regression tree
790 techniques: bagging and random forests for ecological prediction. *Ecosystems*, **9**,
791 181-199.

792 Rauste, Y. (2005) Multi-temporal JERS SAR data in boreal forest biomass mapping.
793 *Remote Sensing of Environment*, **97**, 263-275.

794 Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R.,
795 Buermann, W., Lewis, S.L. & Hagen, S. (2011) Benchmark map of forest carbon
796 stocks in tropical regions across three continents. *Proceedings of the National*
797 *Academy of Sciences*, **108**, 9899-9904.

798 Sader, S.A., Waide, R.B., Lawrence, W.T. & Joyce, A.T. (1989) Tropical forest biomass
799 and successional age class relationships to a vegetation index derived from
800 Landsat TM data. *Remote Sensing of Environment*, **28**, 143-198.

801 Schutz, B., Zwally, H., Shuman, C., Hancock, D. & DiMarzio, J. (2005) Overview of the
802 ICESat mission. *Geophysical Research Letters*, **32**, L21S01.

803 Simard, M., Pinto, N., Fisher, J.B. & Baccini, A. (2011) Mapping forest canopy height
804 globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences*
805 *(2005–2012)*, **116**, G04021.

806 Soenen, S.A., Peddle, D.R., Hall, R.J., Coburn, C.A. & Hall, F.G. (2010) Estimating
807 aboveground forest biomass from canopy reflectance model inversion in
808 mountainous terrain. *Remote Sensing of Environment*, **114**, 1325-1337.

- 809 State Forestry Administration of China (2013) *2012 China Forestry Statistical Yearbook*.
 810 China Forestry Publishing House, Beijing.
- 811 Su, Y. & Guo, Q. (2014) A practical method for SRTM DEM correction over vegetated
 812 mountain areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, **87**, 216-
 813 228.
- 814 Su, Y., Guo, Q., Ma, Q., & Li, W. (2015) SRTM DEM Correction in Vegetated Mountain
 815 Areas through the Integration of Spaceborne LiDAR, Airborne LiDAR, and
 816 Optical Imagery. *Remote Sensing*, **7**, 11202-11225.
- 817 Sun, G., Ranson, K., Masek, J., Guo, Z., Pang, Y., Fu, A. & Wang, D. (2008) Estimation
 818 of Tree Height and Forest Biomass from GLAS Data (< Special Issue>
 819 Silvilaser). *Journal of Forest Planning*, **13**, 157-164.
- 820 Swatantran, A., Dubayah, R., Roberts, D., Hofton, M. & Blair, J.B. (2011) Mapping
 821 biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion.
 822 *Remote Sensing of Environment*, **115**, 2917-2930.
- 823 Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D.,
 824 Shvidenko, A., Kompter, E., Ahrens, B. & Levick, S.R. (2014) Carbon stock and
 825 density of northern boreal and temperate forests. *Global Ecology and*
 826 *Biogeography*, **23**, 297-310.
- 827 Tomppo, E., Gschwantner, T., Lawrence, M., & McRoberts, R.E. (2010). *National forest*
 828 *inventories: Pathways for Common Reporting*. New York: Springer.

829 U.S. Geological Survey, 2013. Shuttle Radar Topography Mission:
830 <http://www2.jpl.nasa.gov/srtm/> (accessed 07 April, 2015).

831 Wang, Q., Adiku, S., Tenhunen, J. & Granier, A. (2005) On the relationship of NDVI
832 with leaf area index in a deciduous forest site. *Remote Sensing of Environment*,
833 **94**, 244-255.

834 White, J., Coops, N. & Scott, N. (2000) Estimates of New Zealand forest and scrub
835 biomass from the 3-PG model. *Ecological Modelling*, **131**, 175-190.

836 Woodhouse, I.H., Mitchard, E.T., Brolly, M., Maniatis, D. & Ryan, C.M. (2012) Radar
837 backscatter is not a'direct measure'of forest biomass. *Nature Climate Change*, **2**,
838 556-557.

839 Xiao, J. (2014) Satellite evidence for significant biophysical consequences of the “Grain
840 for Green” Program on the Loess Plateau in China. *Journal of Geophysical*
841 *Research: Biogeosciences*, **119**, 2261-2275.

842 Xu, T., & Hutchinson, M. (2011) ANUCLIM version 6.1 user guide. *The Australian*
843 *National University, Fenner School of Environment and Society, Canberra*.

844 Zhang, C., Ju, W., Chen, J.M., Li, D., Wang, X., Fan, W., Li, M. & Zan, M. (2014a)
845 Mapping forest stand age in China using remotely sensed forest height and
846 observation data. *Journal of Geophysical Research: Biogeosciences*, **119**, 1163-
847 1179.

848 Zhang, G., Ganguly, S., Nemani, R.R., White, M.A., Milesi, C., Hashimoto, H., Wang,
849 W., Saatchi, S., Yu, Y. & Myneni, R.B. (2014b) Estimation of forest aboveground

850 biomass in California using canopy height and leaf area index estimated from
851 satellite data. *Remote Sensing of Environment*, **151**, 44-56.

852 Zhang, Y., Liang, S. & Sun, G. (2014c) Forest biomass mapping of northeastern China
853 using GLAS and MODIS data. *IEEE Journal of Selected Topics in Applied Earth*
854 *Observations and Remote Sensing*, **7**, 140-152.

855 **List of Tables**

856 **Table 1** Comparison of mean forest AGB values from wall-to-wall predicted map and all
857 plot measurements in each vegetation zone. Note that the RMSE was only calculated
858 based on the validation plot dataset.

859 **Table 2** Comparison of forest area and total forest AGB for each climate zone under
860 different forest definitions. Vegetation zones 1–8 represent cold temperate needleleaf
861 forest, temperate needleleaf–broadleaf mixed forest, warm temperate deciduous broadleaf
862 forest, subtropical evergreen broadleaf forest, tropical monsoon forest–rain forest,
863 temperate steppe, temperate desert, and Qinghai-Tibet Plateau alpine vegetation,
864 respectively.

Table 1 Comparison of mean forest AGB values from wall-to-wall predicted map and all plot measurements in each vegetation zone. Note that the RMSE was only calculated based on the validation plot dataset.

Vegetation zone	Predicted mean AGB (Mg/ha)	Plot mean AGB (Mg/ha)	RMSE (Mg/ha)
Cold temperate needleleaf forest	72.75	74.48	32.81
Temperate needleleaf–broadleaf mixed forest	104.02	92.79	26.08
Warm temperate deciduous broadleaf forest	72.29	82.12	32.56
Subtropical evergreen broadleaf forest	133.22	139.43	45.86
Tropical monsoon forest–rain forest	191.43	180.90	23.30
Temperate steppe	56.11	64.70	26.20
Temperate desert	81.45	101.63	13.86
Qinghai-Tibet Plateau alpine vegetation	168.20	160.42	50.12

Table 2 Comparison of forest area and total forest AGB for each climate zone under different forest definitions. Vegetation zones 1–8 represent cold temperate needleleaf forest, temperate needleleaf–broadleaf mixed forest, warm temperate deciduous broadleaf forest, subtropical evergreen broadleaf forest, tropical monsoon forest–rain forest, temperate steppe, temperate desert, and Qinghai-Tibet Plateau alpine vegetation, respectively.

Vegetation zone	0% ^a		10% ^b		25% ^b		30% ^c	
	Forest area (M km ²)	Total AGB (M ton)	Forest area (M km ²)	Total AGB (M ton)	Forest area (M km ²)	Total AGB (M ton)	Forest area (M km ²)	Total AGB (M ton)
1	0.14	926.54	0.14	923.17	0.13	907.54	0.13	899.02
2	0.22	2053.42	0.22	2047.99	0.22	2024.97	0.21	2012.64
3	0.11	796.07	0.11	783.83	0.10	743.50	0.10	725.25
4	1.00	11416.54	0.99	11322.78	0.96	10944.46	0.94	10750.21
5	0.12	2123.52	0.12	2112.93	0.12	2074.80	0.12	2056.13
6	0.08	472.55	0.08	461.35	0.07	426.29	0.07	411.64
7	0.03	205.42	0.03	199.73	0.02	179.61	0.02	170.59
8	0.01	127.97	0.01	124.52	0.01	112.05	0.01	106.39
Total	1.71	18122.03	1.70	17976.30	1.63	17413.21	1.60	17131.86

^aThe value 0% represents forest areas directly coming from the closed-forested area and open-forested area of the Landsat TM land-use map.

^bThe values of 10%, 25%, and 30% represent the three forest definitions using fractional cover thresholds of 10%, 25%, and 30% tree cover, respectively.

List of Figures

Fig. 1 Geolocations of collected ground inventory data and corresponding vegetation zones from Hou (2001). The collected ground inventory data were divided into validation samples (green dots) and training samples (blue dots), at a ratio of 1:3.

Fig. 2 Densities of filtered GLAS shots (per 10^4 ha) used in AGB mapping procedure.

Fig. 3 Scheme for estimating forest AGB distribution in China from multisource remotely sensed data and ground inventory data.

Fig. 4 Importance of variable, denoted by percentage increase of mean-squared error (%IncMSE) (left) and the increase in node purity (IncNodePurity) (right), for AGB estimation RF model built from training dataset without considering plot location uncertainty. LE, WE, TE, PTotal, PSeason, TMean, TSeason, AccNDVI, and VegZone represent the leading edge extent, waveform extent, trailing edge extent, annual total precipitation, annual precipitation seasonality, annual mean temperature, annual temperature seasonality, cumulative NDVI, and vegetation zone, respectively.

Fig. 5 Distributions of extrapolated GLAS-derived full-waveform parameters, namely (a) waveform extent, (b) leading edge extent, and (c) trailing edge extent. Note that areas without closed-forested area and open-forested area coverage have been excluded from the maps.

Fig. 6 Wall-to-wall map of forest AGB distribution in China.

Fig. 7 Boxplot of estimated forest AGB in each vegetation zone. The blue “+” indicates the mean forest AGB of corresponding vegetation zone. Numbers 1–8 on x axis represent cold temperate needleleaf forest, temperate needleleaf–broadleaf mixed forest, warm

901 temperate deciduous broadleaf forest, subtropical evergreen broadleaf forest, tropical
902 monsoon forest–rain forest, temperate steppe, temperate desert, and Qinghai-Tibet
903 Plateau alpine vegetation, respectively.

904 **Fig. 8** Evaluation of modeled forest AGB using validation ground inventory dataset at
905 plot level. Note that R^2 represents the adjusted coefficient of determination, $RMSE$
906 represents the root-mean-square error, B represents the field measured AGB, and $\hat{B}$
907 represents the estimated AGB.

908 **Fig. 9** Distribution of modeled forest AGB uncertainty introduced by plot location
909 uncertainty at pixel level. White areas indicate areas with neither closed-forested area nor
910 open-forested area land-use cover.

911 **Fig. 10** (a) Pixel-level differences between the estimated forest AGB from this study and
912 the forest AGB from Saatchi et al. (2011) [our estimated AGB minus the forest AGB
913 from Saatchi et al. (2011)]. (b) Scatter plot between the field measured AGB and the
914 forest AGB from Saatchi et al. (2011). Note that R^2 represents the adjusted coefficient of
915 determination, $RMSE$ represents the root-mean-square error, B represents the field
916 measured AGB, and $\hat{B}$ represents the estimated AGB from Saatchi et al. (2011).

Fig. 1 Geolocations of collected ground inventory data and corresponding vegetation zones from Hou (2001). The collected ground inventory data were divided into validation samples (green dots) and training samples (blue dots), at a ratio of 1:3.

Fig. 2 Densities of filtered GLAS shots (per 10^4 ha) used in AGB mapping procedure.

Fig. 3 Scheme for estimating forest AGB distribution in China from multisource remotely sensed data and ground inventory data.

Fig. 4 Importance of variables, denoted by percentage increase of mean-squared error (%IncMSE) (left) and the increase in node purity (IncNodePurity) (right), for AGB estimation RF model built from training dataset without considering plot location uncertainty. LE, WE, TE, PTotal, PSeason, TMean, TSeason, AccNDVI, and VegZone represent the leading edge extent, waveform extent, trailing edge extent, annual total precipitation, annual precipitation seasonality, annual mean temperature, annual temperature seasonality, cumulative NDVI, and vegetation zone, respectively.

Fig. 5 Distributions of extrapolated GLAS-derived full-waveform parameters, namely (a) waveform extent, (b) leading edge extent, and (c) trailing edge extent. Note that areas without closed-forested area and open-forested area coverage have been excluded from the maps.

Fig. 6 Wall-to-wall map of forest AGB distribution in China.

941

942 **Fig. 7** Boxplot of estimated forest AGB in each vegetation zone. The blue “+” indicates
 943 the mean forest AGB of corresponding vegetation zone. Numbers 1–8 on x axis represent
 944 cold temperate needleleaf forest, temperate needleleaf–broadleaf mixed forest, warm
 945 temperate deciduous broadleaf forest, subtropical evergreen broadleaf forest, tropical
 946 monsoon forest–rain forest, temperate steppe, temperate desert, and Qinghai-Tibet
 947 Plateau alpine vegetation, respectively.

948

949 **Fig. 8** Evaluation of modeled forest AGB using validation ground inventory dataset at
 950 plot level. Note that R^2 represents the adjusted coefficient of determination, $RMSE$
 951 represents the root-mean-square error, B represents the field measured AGB, and $\hat{B}$
 952 represents the estimated AGB.

Fig. 9 Distribution of modeled forest AGB uncertainty introduced by plot location uncertainty at pixel level. White areas indicate areas with neither closed-forested area nor open-forested area land-use cover.

Fig. 10 (a) Pixel-level differences between the estimated forest AGB from this study and the forest AGB from Saatchi et al. (2011) [our estimated AGB minus the forest AGB from Saatchi et al. (2011)]. (b) Scatter plot between the field measured AGB and the forest AGB from Saatchi et al. (2011). Note that R^2 represents the adjusted coefficient of determination, $RMSE$ represents the root-mean-square error, B represents the field measured AGB, and $\hat{B}$ represents the estimated AGB from Saatchi et al. (2011).