Remote Sensing of Environment 216 (2018) 514–528

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Quantification of uncertainty in aboveground biomass estimates derived from small-footprint airborne LiDAR

Qing Xu[a,*], Albert Man[b], Mark Fredrickson[b], Zhengyang Hou[a], Juho Pitkänen[c], Brian Wing[d,1], Carlos Ramirez[e], Bo Li[b], Jonathan A. Greenberg[a]

[a] Department of Natural Resources and Environmental Science, University of Nevada, Reno, NV 89557, USA
[b] Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA
[c] Natural Resources Institute Finland, Helsinki, Finland
[d] USDA Forest Service, Pacific Southwest Research Station, Redding, CA, USA
[e] USDA Forest Service, Region 5 Remote Sensing Laboratory, McClellan, CA, USA

ARTICLE INFO

Keywords:
California forests
Individual tree detection
Allometric equations
Uncertainty decomposition
Omission and commission errors

ABSTRACT

To address uncertainty in biomass estimates across spatial scales, we determined aboveground biomass (AGB) in Californian forests through the use of individual tree detection methods applied to small-footprint airborne LiDAR. We propagated errors originating from a generalized allometric equation, LiDAR measurements, and individual tree detection algorithms to AGB estimates at the tree and plot levels. Larger uncertainties than previously reported at both tree and plot levels were found when AGB was derived from remote sensing. On average, per-tree AGB error was 135% of the estimated AGB, and per-plot error was 214% of the estimated AGB. We found that from tree to plot level, the allometric equation constituted the largest proportion of the total AGB uncertainty. The proportion of the uncertainty associated with remote sensing errors was larger in lower AGB forests, and it decreased as AGB increased. The framework in which we performed the error propagation analysis can be used to address AGB uncertainties in other ecosystems and can be integrated with other analytical techniques.

1. Introduction

Forests play an important role in sequestering carbon by absorbing carbon dioxide from the atmosphere. Programs such as Reducing Emissions from Deforestation and Forest Degradation (REDD) have been developed to incentivize reducing forest carbon emissions which account for approximately 17% of all carbon emissions worldwide, just after the emissions from burning fossil fuels (Stocker et al., 2013). NASA Carbon Monitoring System (CMS) projects (NASA CMS, 2010) are designed to "monitor, report and verify" (MRV) carbon stocks and fluxes at local to global scales using a variety of terrestrial, airborne and spaceborne remote sensing systems. While carbon estimation using remote sensing technology is the central aspect of the NASA CMS, an important but understudied component is understanding the uncertainties in the estimation of carbon stocks in forested ecosystems at local, regional, national and international scales.

Assessing carbon sequestered in forests often relies first on the quantification of aboveground biomass (AGB), and then assuming that a fraction of that AGB is carbon (Brown and Lugo, 1982; Malhi et al., 2004). Originally, destructive sampling of trees was required for measuring individual tree AGB, which involved tree felling, cutting by components, drying and weighing. Tree-level structural parameters such as diameter at breast height (DBH) and height were found to have strong correlations with individual tree biomass, which led to the development and use of typically species-specific allometric equations that relate these structural attributes to biomass (Brown, 1997; Chave et al., 2001; Chave et al., 2005; Jenkins et al., 2003; Lambert et al., 2005; Clark and Kellner, 2012). The use of allometric equations has become the standard approach to estimating AGB in sample plots established for national forest inventory programs (Olson et al., 1983) as well as for calibrating remotely-sensed estimates of AGB. While the use of allometric equations is the dominant paradigm by which AGB is initially predicted, issues surrounding sampling design, field data quality and the proper application of statistical methods have led to a

* Corresponding author at: Department of Natural Resources and Environmental Science, University of Nevada, Mail Stop 186, 1664 N. Virginia Street, Reno, NV 89557, USA.
E-mail address: qingx@unr.edu (Q. Xu).
[1] Deceased.

https://doi.org/10.1016/j.rse.2018.07.022
Received 7 September 2017; Received in revised form 20 June 2018; Accepted 16 July 2018
Available online 23 July 2018
0034-4257/ Published by Elsevier Inc.

debate on the accuracy of field-based biomass estimates (Clark, 2002; Chave et al., 2003; Melson et al., 2011; Breidenbach et al., 2014; Duncanson et al., 2015; Marvin and Asner, 2016). Multiple sources of error in AGB estimates have been identified and quantified at both individual tree and plot levels. Chave et al. (2004) pointed out that in tropical forests, most of the AGB error comes from the selection of allometric equations. Duncanson et al. (2015) found that published allometric equations for temperate forests typically used small sample sizes that were insufficient to accurately parameterize the models. Tree biomass can also be obtained by multiplying tree volume with wood density. McRoberts and Westfall (2014) found that the effects of uncertainty in model predictions of individual tree volume on large area volume estimates and their uncertainties were small when at least 95% of variability was explained by the models.

Biomass estimates at the plot level are typically obtained by aggregating predicted biomass of individual trees within a plot and multiplying the result by an expansion factor. These plot-level AGB estimates are used as both calibration and validation data in most remote sensing-based AGB maps. Therefore, remote sensing derived AGB suffers not only from the remote sensing models, methodologies applied, and sensor limitations, but also from the uncertainty stemming from the predicted AGB used to calibrate these models in the first place. Recent remote sensing studies have investigated these error sources from different perspectives. Gregoire et al. (2011) presented a model-assisted approach that used airborne laser scanning (ALS) metrics to estimate mean and total biomass for a Norwegian county. They also derived variance estimators for these population parameters in a design-based framework. Gonzalez et al. (2010) analyzed data from California and assessed empirical uncertainty in carbon derived from field measurements, LiDAR and QuickBird respectively using simulations. They found that at the population level LiDAR achieved the lowest uncertainty. Chen et al. (2015) quantified the uncertainty of airborne LiDAR derived biomass at the pixel level for tropical forests, and they concluded that errors from residuals of the remote sensing regression model contributed the most to the pixel-level biomass uncertainty. These studies are based on area-based approaches that usually rely on regression functions to relate LiDAR metrics to forest mean attributes at the plot level. Area-based approaches have been found to be more accurate than individual tree detection approaches in generating plot-level estimates of forest attributes (Næsset, 2004; Xu et al., 2014). However, previous studies that are based on LiDAR metrics are not sufficient for fully assessing biomass uncertainties due to the lack of insight into tree-level details and forest structure and dynamics. Furthermore, using area-based approaches makes it difficult to understand biomass uncertainties at different spatial scales. Instead of area-based approaches, individual tree detection (ITD) approaches allow estimation of plot-level biomass and propagation of errors in an up-scaling framework.

Similar to how biomass inventories are carried out using field data, the ITD approach applied to hyperspatial remote sensing data can identify and estimate structural parameters of single trees but at the scale of entire landscape. Per-tree biomass can be predicted using tree-level allometric models and then be aggregated to the plot level. Thereafter, a model-based or model-assisted approach can be used for estimating the population mean, $\hat{\mu}$, and total, τ, as well as their variances $\hat{V}(\hat{\mu})$ and $\hat{V}(\hat{\tau})$ (Kangas, 1994; Gregoire, 1998; Hou et al., 2017, 2018). However, limitations on current approaches to analyzing ALS data preclude the direct use of most existing allometric equations. Specifically, ALS has not been demonstrated to be highly effective in discriminating tree species (Brandtberg, 2007; Korpela et al., 2010), nor can it measure DBH since tree trunks are not typically resolved by ALS (Vauhkonen et al., 2010; Hou et al., 2016). Popescu and Wynne (2004) and Salas et al. (2010) predicted DBH using LiDAR-derived height and crown variables, but the underexposed uncertainty in DBH estimates has the potential to intensify the uncertainty in LiDAR-derived biomass. Thus, the development and use of species-agnostic height-based allometric equations are required to predict biomass of individual trees using current LiDAR methodologies. Uncertainties associated with these equations and their impacts on larger scale estimations of biomass have not been fully studied. Furthermore, ITD approaches can contain significant omission and commission errors (Brandtberg, 1999; Hyyppä and Inkinen, 1999; Kaartinen and Hyyppä, 2008; Breidenbach et al., 2010) that must be addressed when considering the uncertainty in AGB estimation.

In this analysis, our primary goal is to develop a methodological framework to study the uncertainty (variance and bias) of LiDAR remote sensing based AGB estimation by propagating four different error sources at both tree and plot level and to quantify the relative contribution of each of these four components to the total uncertainty in AGB. We consider four sources of error that are responsible for the variability in the estimated biomass: 1) measurement errors in the LiDAR-derived tree height (variability in the LiDAR height measurement), 2) errors associated with parameters of the allometric equation (variability in model parameters between sites), 3) errors associated with residuals of the allometric equation (variability in the AGB that cannot be explained by the model), 4) errors associated with tree detection performance using individual tree crown detection techniques (omission and commission errors). In other words, our goal is not to pursue the best model that generates the most accurate point estimate of AGB, but to model and understand uncertainty when a common form of allometric equation is used. Subsequently, we quantify the spatially explicit AGB and their uncertainties by applying this approach to a large collection of LiDAR acquisitions over California, to understand the landscape-scale biomass uncertainty of various forest ecosystems in California.

2. Materials

2.1. Study area

California, a state that features a variety of climate zones has approximately 133,546 km^2 of forests characterized by a rich diversity in species composition, size distribution and management objectives. In 2006, the California Global Warming Solutions Act (California Assembly Bill 32, 2006) was passed. This legislation allows the California Air Resources Board (ARB) to collect revenues from the state's largest greenhouse gas emitters for the Greenhouse Gas Reduction Fund (GGRF). One of the uses of the GGRF is to fund CAL FIRE‘s California Climate Investments (CCI) Forest Health Grant Program, with a goal of restoring forests to “protect upper watersheds where the state's water supply originates, promote the long-term storage of carbon in forest trees and soils, minimize the loss of forest carbon from large, intense wildfires, and to further the goals of the California Global Warming Solutions Act of 2006” (CAL FIRE, 2017). Projects submitted under the CCI program have a requirement to perform carbon accounting (CARB, 2018) using forest growth and yield models, FIA's Carbon OnLine Estimator (COLE) or forest fire models, in the case of projects with a fuels reduction emphasis. Improved methods for carbon quantification and uncertainty estimation from remotely sensed data is timely for estimating current carbon stocks and parameterizing models to project carbon positive benefits of silvicultural prescriptions.

2.2. Field data

We used two sets of field measurements: one for fitting a height-based allometric equation, and the other for deriving LiDAR processing-associated errors that stem from tree height estimation as well as omission and commission errors. The Forest Inventory and Analysis (FIA) dataset (PNW-FIA Field Manual, 2015) was used for estimating parameters of the height-based allometric equation. Each FIA plot has four subplots (radius 7.32 m) with one in the center and three on the

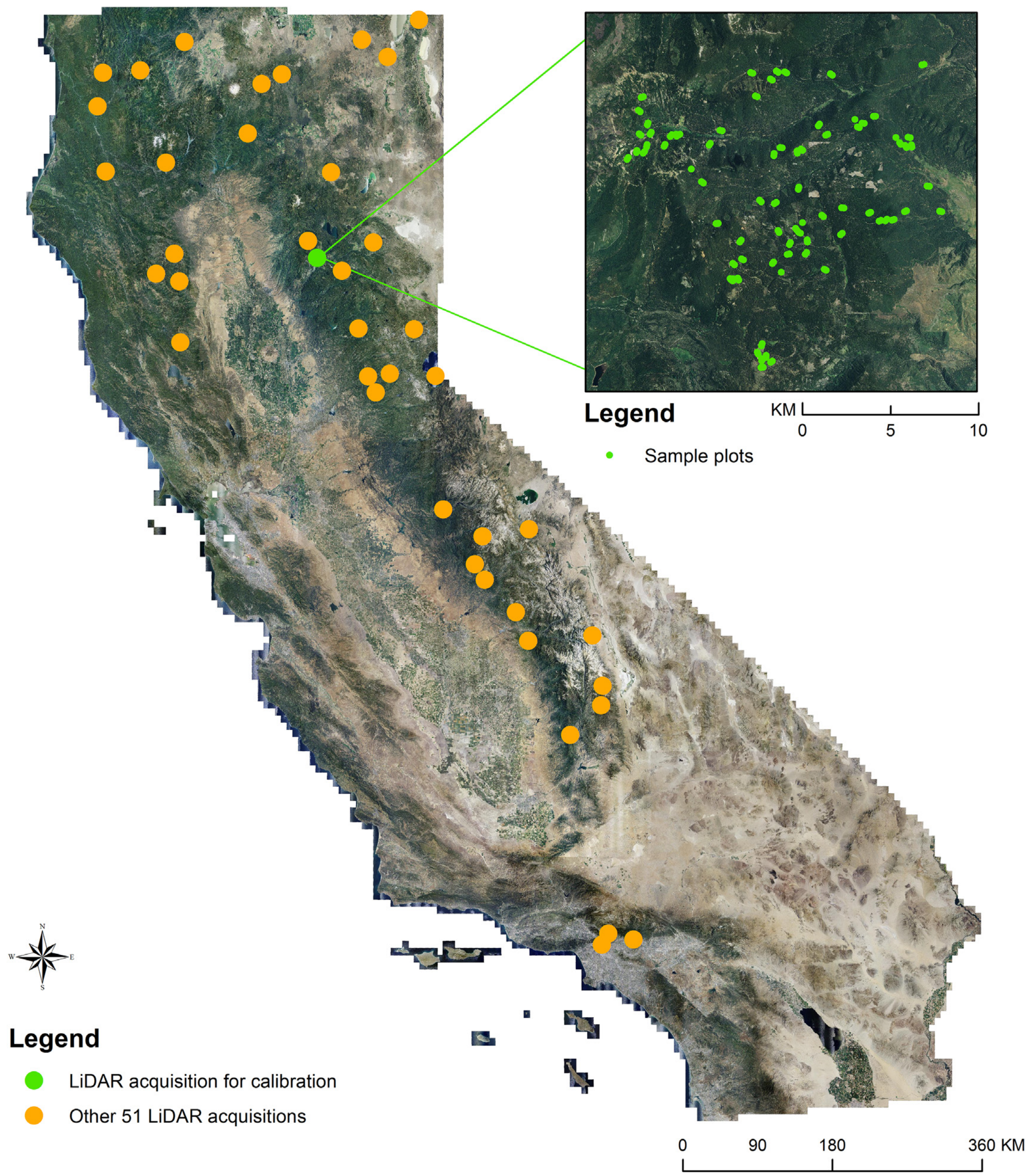

Fig. 1. LiDAR acquisition locations within California used in this analysis, with LiDAR calibration plots located in the Lassen National Forest in the inset.

Table 1
Summary statistics of tree sizes in the Lassen National Forest dataset.

Tree dimension	Minimum	Median	Mean	Maximum	Standard deviation
Height (m)	2.13	9.14	12.20	61.57	8.97
DBH (cm)	5.08	18.03	25.01	206.80	20.21

periphery. Species, diameter at breast height (DBH), and height for all trees with DBH ≥12.7 cm are measured. The FIA dataset was subsetted to California, resulting in a dataset with over 208,000 individual tree measurements.

In addition to the California-wide FIA dataset, we also used a dataset collected by the USDA Forest Service (PNW-FIA Field Manual, 2015) for the purposes of determining the errors associated with LiDAR height measurement, omission and commission. This field campaign was performed in the Lassen National Forest (Fig. 1) during the summers of 2013 and 2014. Lassen has a Mediterranean climate characterized by hot dry summer and cool wet winter. Douglas fir (*Pseudotsuga menziesii*), ponderosa pine (*Pinus ponderosa*), Jeffrey pine (*Pinus jeffreyi*), lodgepole pine (*Pinus contorta*), white fir (*Abies concolor*)) and red fir (*Abies magnifica*) are the dominant tree species in this region. Detailed tree information was collected for 8313 trees in 146 circular fixed-area sample plots of 900 m^2 (radius 16.93 m) that were randomly located in pairs 100 m from each other within various ecological strata. The plot locations were recorded using a differential GPS unit. Among other variables, tree species, height, DBH, and crown widths were collected for all trees with height ≥ 2 m. Descriptive statistics of DBH and height of the tallied Lassen trees are presented in Table 1.

2.3. Airborne LiDAR data

Fifty-two (52) LiDAR acquisitions were collected in the growing season (leaf-on) and compiled by the US Forest Service from 2005 to

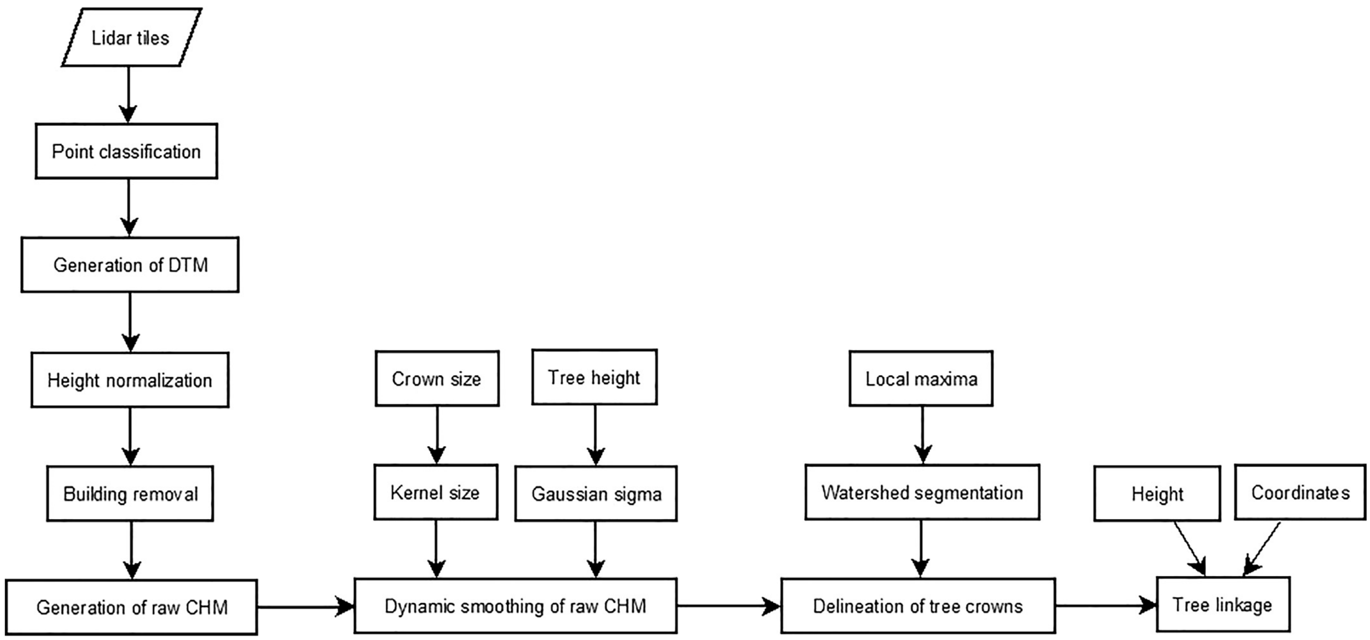

Fig. 2. Flowchart of the standardized preprocessing chain and individual tree detection.

2013 over California forests (Fig. 1). Although they were acquired by different surveyors to differing specifications in terms of sensor type, pulse wavelengths, flight altitude, pulse density, point classification, and rasterized products, they met the consolidated requirements for point density of approximately 8 points/m^2. The mean point density of all acquisitions is 11.9 points/m^2, with 4.3 as the minimum and 51.2 as the maximum. Point density of the Lassen training set is about 9.3 points/m^2. To create a consistent dataset for analysis, we developed a standardized preprocessing chain. Starting from the raw point clouds, the processing chain consisted of point classification, generation of the canopy height model (CHM), and dynamic smoothing of the raw CHM (see Fig. 2). To avoid edge artifacts, the non-overlapping LiDAR point cloud tiles were retiled with a 40 m buffering zone around each of them. Raw LiDAR points were first classified into ground points and non-ground points using the progressive triangular irregular network (TIN) densification algorithm (Axelsson, 1999) performed on last returns only with a step size of 3 m and an intensified search for initial ground points. Next, a digital terrain model (DTM) was generated with the ground points using a TIN (Isenburg et al., 2006). Raw LiDAR points were height-normalized by subtracting the DTM from the orthometric heights at the corresponding locations. The normalized LiDAR points with above-ground heights were first filtered by 320 m, the tallest building in California, and later filtered by removing points below 2 m and points that are multi-returns. An algorithm that examines the planarity of the object surface was applied to detect buildings in point clouds, which were afterwards removed by changing their heights into zero. The filtered points were then added back to restore the LiDAR tile that had identical point number with the original tile, and further filtered by 115 m, the tallest tree in California. The pit-free CHM algorithm proposed by Khosravipour et al. (2014) was applied to generate the canopy height model. A number of temporary rasterized CHMs were created using the first returns with "-kill" set up to 1 m to limit interpolation distance and to minimize the risk of creating false canopies. The raw CHM was determined by taking the highest value of these temporary CHMs for each pixel of the size of 0.25 m.

In order to delineate intact tree crowns with compact boundaries, the raw CHM was smoothed by a fully dynamic Gaussian filter, the kernel size of which is a function of predicted crown size, and the standard deviation of the Gaussian distribution (sigma) is a function of the pixel's relative height in a defined neighborhood. Crown size was predicted for each pixel using a regression model that relates tree height to crown diameter fitted to the Lassen dataset. Kernel size was adjusted based on canopy size, so that larger crowns have a larger smoothing kernel, and smaller crowns use a smaller kernel. The dynamic kernel size assists in filling holes in crown area, as well as removing noises from outside of crowns. The smoothing parameter sigma was higher for dominant canopies, and lower for understory canopies. This assists in avoiding blurring between crown boundaries and between trees and non-tree objects (shrubs, herbs, rocks and soil) so that small trees at the edge of taller crowns can be identified. Rather than basing the smoothing parameter on a fixed canopy height, we used the relative height calculated within a 10 m window surrounding a given location to correctly represent the canopy layers of the crown. The smoothing parameter was linearly scaled between 0.2 (lowest relative canopy height) to 1.2 (highest relative canopy height).

3. Methods

3.1. Overview

This analysis propagated AGB uncertainty through two scales: 1) at the individual tree level and 2) at the plot and pixel level. To accomplish this, we first established a height-based allometric equation relating LiDAR-measured tree height to per-tree AGB. Second, we propagated errors in LiDAR-measured heights into per-tree AGB estimates using the height-based AGB model. Finally, we integrated estimates of omission and commission errors from the individual tree crown delineation to determine plot-level errors.

3.2. Height-based allometric equation

A height-based allometric equation that related per-tree heights to AGB in a species-agnostic fashion was developed using a generalized allometric equation ("GAE") approach calibrated using trees in the California-wide FIA dataset. To generate semi-synthetic AGB for each tree in the FIA dataset, published species-specific allometric equations compiled as part of the GlobAllomeTree database (GlobAllomeTree, 2013) were used to predict per-tree AGB from field-measured DBH. For each tree, 1000 random values of Gaussian noise with zero mean and a variance estimated using the published R^2 value for the allometric

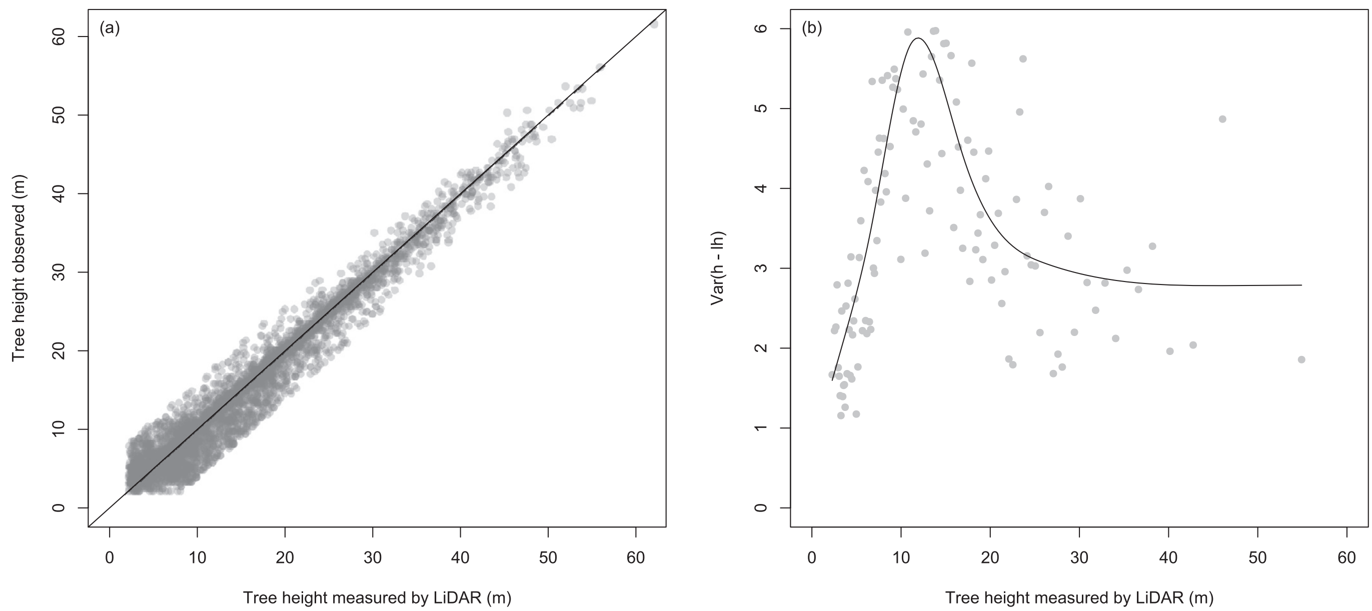

Fig. 3. Uncertainty in the LiDAR height measurements. a. LiDAR performance in measuring tree heights observed in the Lassen National Forest field campaign. b. The variance of the difference between field-measured and LiDAR-measured heights is modeled as a function of LiDAR-measured tree heights.

equation (Appendix A) were sampled and added to the predicted AGB. This essentially estimated a distribution of actual AGB for each tree in the FIA dataset. A reasonable point estimate of a tree's exact AGB, namely the semi-synthetic AGB mentioned above, is the mean of this distribution. Using ordinary least squares, a height-based allometric equation (Eq. (1)) of a common form of the published allometric equations was fitted to the dataset that consisted of FIA dataset and the semi-synthetic AGB.

$$ln(b_i) = \beta_0 + \beta_1\sqrt{h_i} + \varepsilon_i \tag{1}$$

where b_i is the semi-synthetic above-ground biomass at the i-th tree, β_0 and β_1 are coefficients to be estimated, h_i is field-measured tree height and ε_i is the model error that follows a normal distribution with mean zero and variance as σ^2, denoted by $\varepsilon_i \sim N(0, \sigma^2)$.

3.3. Individual tree delineation and LiDAR-derived tree height

LiDAR-derived individual tree heights were estimated based on results from an individual tree crown delineation analysis (Fig. 2). A local maxima-incorporated watershed segmentation algorithm (Gauch, 1999; Pitkänen et al., 2004) was applied to the low-pass filtered CHM to identify individual trees by delineating the crown boundaries. Segments were formed around local maxima, which were first converted into local minima. Tree height was extracted from the maximum height value of each segmented crown, and the coordinates of the maximum height were assumed to be the trunk or tree top location. Crown diameters were measured by taking the maximum diameter and its perpendicular of each segmented crown. Other tree-level attributes were calculated including distance to the nearest tree and the mean distance to all trees within a sample plot to account for the local stem density.

The LiDAR-detected trees were linked with the field-measured trees in the Lassen National Forest based on their similarity in three-dimensional space, taking both tree height and tree location into account (Xu et al., 2014). The threshold of 6 m was used as the maximum height difference to remove possible false tree linkages.

3.4. Tree-level AGB prediction and uncertainty decomposition

For LiDAR-detected trees in the Lassen National Forest, per-tree AGB were predicted using Eq. (2) by transforming the log scale AGB in Eq. (1) back into the original units. Additionally, a correction (Baskerville, 1972; Brown et al., 1989) was applied to ensure the unbiasedness of the regression mean in the original scale (Eq. (2)). The field-measured height was replaced by the LiDAR-measured tree height, denoted by lh_i to obtain the LiDAR-derived biomass, denoted by lb_i.

$$E(b_i) = \exp\left(f(h_i) + \frac{\hat{\sigma}^2}{2}\right) \tag{2}$$

where b_i and h_i are AGB and field-measured height at the i-th tree respectively, E indicates the regression mean, $\frac{\hat{\sigma}^2}{2}$ is the bias correction term and $\hat{\sigma}^2$ is the estimated variance of residuals in Eq. (1).

To investigate uncertainty in per-tree AGB estimation, we needed AGB estimates in their original scale. We therefore exponentiated both sides of Eq. (1) to convert back from the log scale:

$$b_i = e^{\beta_0 + \beta_1 \times \sqrt[2]{h_i}} \times e^{\varepsilon_i} \tag{3}$$

We denoted the right side of Eq. (3) as $g(h, \beta, \varepsilon)$, a function of three parameters, each of which represented one error source of the estimated per-tree AGB. β is a vector consisting of β_0 and β_1. Since in practice we only observe lh, $\hat{\beta}$ and ε, instead of their true values, we used a first-order Taylor expansion around the observed values to express the AGB at h, β and ε in Eq. (4), as in Gertner et al. (1995) (see Appendix B):

$$b_i \cong g(lh_i, \hat{\beta}, \hat{\varepsilon}) + \frac{dg}{dh}(h_i - lh_i) + \frac{dg}{d\beta}(\beta - \hat{\beta}) + \frac{dg}{d\varepsilon}(\varepsilon - \hat{\varepsilon}) \tag{4}$$

where lh is the LiDAR-measured tree height and $\frac{dg}{dh}$, $\frac{dg}{d\beta}$ and $\frac{dg}{d\varepsilon}$ are partial derivatives of g function with respect to h, β and ε, respectively.

The variance of the LiDAR-derived biomass was formulated in Eq. (5) (see Appendix C), with two assumptions that the three error terms are independent from each other, and that the model parameters are unbiased.

$$var(b_i) \cong \left(\frac{dg}{dh}\right)^2 var(h_i - lh_i) + \left(\frac{dg}{d\beta}\right)^2 var(\hat{\beta}) + \left(\frac{dg}{d\varepsilon}\right)^2 var(\hat{\varepsilon}) \tag{5}$$

The first term of Eq. (5) is the error associated with LiDAR height measurements with an assumption that field-measured heights are errorless. $var(h - lh)$ was estimated using known loctions in which both field- and LiDAR-measured heights are observed. Differences between the field-measured and LiDAR-measured heights were binned into classes of the LiDAR-measured heights. Empirical variances (grey dots) in height differences were calculated for bins and modeled with mean height of each bin using natural cubic splines (Fig. 3b), which is commonly used for nonparametric interpolation (Liu, 1980). A cubic spline is also flexible enough to model the non-negative $var(h - lh)$ for trees that are taller than any observed trees in Lassen.

The second term of Eq. (5) is the error associated with model parameters, and $var(\hat{\beta})$ is the variance-covariance matrix of the estimated β. The third term is the error associated with model residuals, and $var(\hat{\varepsilon})$ is the model residual error. $\frac{dg}{dh}$, $\frac{dg}{d\beta}$ and $\frac{dg}{d\varepsilon}$ are essentially functions of h (see Appendix D). The variance of the LiDAR-derived tree biomass is ultimately a function of the LiDAR-measured tree height and can be decomposed into three separate error components associated with the height-based allometric equation.

3.5. Plot-level AGB and uncertainty decomposition

Conventionally, plot-level biomass (Mg/ha) was obtained by aggregating per-tree biomass in a sample plot and multiplying the result by an expansion factor. Since the intrinsic challenges of the ITD approach lie in underestimating number of trees (omission errors) and splitting intact tree crowns (commission errors), we suggest a new method for estimating plot-level biomass density formulated in Eq. (6) that takes omission and commission errors as a correction to the LiDAR direct detection of AGB.

$$B_j = \left(\sum_{i=1}^{n} b_{i,j} - \sum_{c=1}^{m} b_{c,j} + \sum_{o=1}^{l} b_{o,j} \right) \times \frac{10000}{A_j} \tag{6}$$

where B_j is the LiDAR-derived AGB density at plot j, and A_j is the area of plot j, $b_{i,\ j}$, $b_{c,\ j}$ and $b_{o,\ j}$ are AGB of the LiDAR-detected trees, commission errors and omission errors,and n, m and l are the number of the LiDAR-detected trees, commission errors and omission errors.

Variance of the LiDAR-derived plot-level AGB is therefore formulated in Eq. (7) as the sum of the variance in the LiDAR-detected AGB and the variance in remote sensing-associated omission and commission errors assuming that within a sample plot the error for each tree is independent of other trees, and that commission error, omission error and the error in the LiDAR-detected AGB are independent from each other.

$$var(B_j) = var(B_{i,j}) + var(B_{o,j} - B_{c,j}) \tag{7}$$

where $B_{i,\ j}$, $B_{o,\ j}$ and $B_{c,\ j}$ are AGB density of LiDAR-detected trees, omission errors and commission errors respectively at the j-th plot.

The first term of Eq. (7) can be written into $\sum_{i=1}^{n} var(b_{i,j}) \times \frac{10000}{A_j}$, the summation of variance of each LiDAR-detected tree. $var(b_{i,\ j})$ was solved by inserting each LiDAR-measured height into Eq. (5). The second term is the variance in plot-level omission and commission AGB, and it was derived by modeling omission and commission AGB as a whole. The empirical AGB associated with omission and comission errors was calculated for each sample plot between the field data and the LiDAR tree detection results. Since the omission-commission errors are closely related to forest horizontal and vertical structures, to scale the errors to a landscape we modeled them with LiDAR height and density metrics extracted at the plot level from delineated tree crowns. For each sample plot, we calculated a set of LiDAR metrics based on four tree-level attributes: tree height, crown area, distance to the nearest tree and mean distance to all trees. Statistics such as mean, percentiles with an interval of 10%, standard deviation, skewness and kurtosis were generated for each of the attributes. Canopy cover was calculated at a set of height thresholds at 5, 10, 15 and 20 m. These LiDAR metrics, together with the number of tree crowns and the total AGB of tree crowns, were used as candidates for predictor variables in a linear model to determine the AGB associated with omission-comission errors. A stepwise AIC method was used for variable selection. Variability in the omission-commission AGB was estimated in a leave-one-out cross validation as the root mean squared error of the omission-commission AGB.

3.6. Application of the methods to the California LiDAR dataset

We implemented the whole framework of biomass estimation and uncertainty analysis on a set of 52 LiDAR acquisitions over California covering > 18,780 km^2. Relationships we obtained for the Lassen National Forest between forest structures and errors from the LiDAR-derived height, omission and commission were assumed to represent forests in the whole state.

4. Results

4.1. Overview of results

Error propagation was performed at the tree level based on the height-based allometric equation. The AGB uncertainty associated with the allometric model is assessed in Section 4.2 using confidence intervals. The accuracy of the LiDAR-derived height is reported in Section 4.3. After field-measured tree height was replaced by the LiDAR-derived height, we investigated tree-level uncertainty components associated with the height-based allometric equation and the LiDAR height measurements respectively in Section 4.4. In Section 4.5, uncertainty components are assessed at the plot level by aggregating tree-level uncertainty and integrating the uncertainty associated with tree detection algorithms. Finally, AGB uncertainty is analyzed at the landscape scale in Section 4.6.

4.2. Height-based allometric equation

The generalized allometric equation approach yielded a height-based allometric equation with estimated parameters specified in Eq. (8). Fig. E1 in Appendix E shows point estimation and interval estimation of AGB as a function of tree height. The width of the confidence interval increased with tree height, which suggests that uncertainty in the AGB can be expressed as a function of tree height. Fig. E2 presents predicted AGB using the height-based allometric equation versus the semi-synthetic AGB. The RMSE of predicted AGB is 4771.22 kg (243.09% for the relative RMSE). The height-based model underestimated per-tree AGB by 834.74 kg (42.53% for the relative bias).

$$ln(b_i) = 0.9329 + 1.1846\sqrt{h_i} \tag{8}$$

4.3. LiDAR performance in height measurement

LiDAR-estimated tree height was found to have a high correlation with the field-measured tree height ($R^2 = 0.98$) (Fig. 3a), with an RMSE of 2.05 m (14.40% for the relative RMSE) and a bias of −0.23 m (−1.61% for the relative bias). LiDAR systematically overestimated tree height by 1.61% of the mean tree height.

4.4. LiDAR-derived tree-level biomass and components of its uncertainty

LiDAR-derived biomass at the tree level had its uncertainty associated with the generalized allometric equation in terms of three error sources: 1) model predictors: the LiDAR-measured tree heights, 2) allometric model residuals, and 3) allometric model parameters. With standard deviation in the y-axis, Fig. 4a shows the uncertainty in the LiDAR-derived AGB as a function of tree height (see Eq. (5)), and the cumulative uncertainties by the three components. The variance estimator corresponded well with the empirical variances (grey dots), except that it underestimated variance for trees of height < 16 m. The empirical variances $var(b - lb)$ were calculated by first dividing trees into bins according to the LiDAR-derived height and then calculating

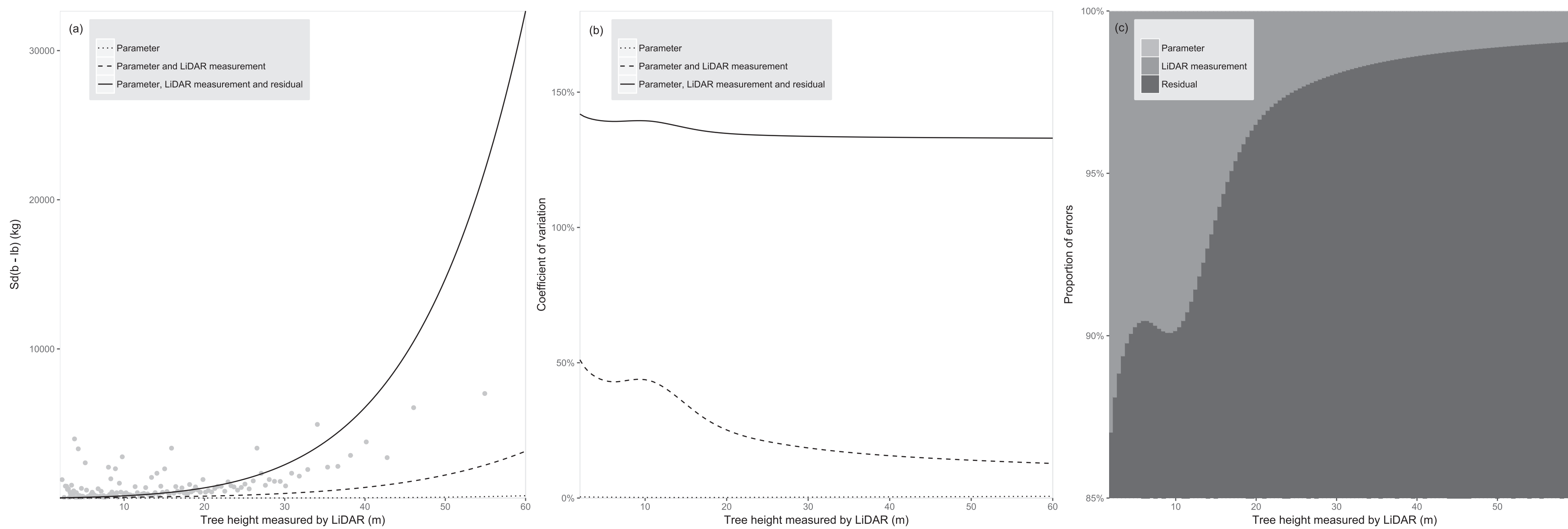

Fig. 4. Biomass uncertainty at the tree level. a. Uncertainty in the LiDAR-derived individual tree biomass as a function of the LiDAR-measured tree height. b. Coefficient of variation does not vary largely with tree height. c. Decomposition of uncertainty into three error sources shows that the largest contribution is from residuals, the second from LiDAR height measurements and a negligible contribution from parameters.

the variance of the difference between the semi-synthetic AGB and the AGB estimated by the height-based model for each bin. As a standardized measure of dispersion of a probability distribution, the coefficient of variation was utilized to assess the spread of the distribution of the LiDAR-derived AGB. The coefficient of variation is defined as a ratio of the standard deviation of AGB to the mean of AGB. Fig. 4b presents the cumulative coefficient of variation of the LiDAR-derived AGB at the tree level. The coefficient of variation does not vary largely with tree height, with a mean at 135% and a range between 133% and 142%.

When per-tree uncertainty was decomposed into these three components, we found that the errors associated with model residuals contributed the largest proportion (from 87% to 99%) of the total uncertainty and increased along with an increase in tree height (Fig. 4c). Errors associated with Lidar height measurement contributed a smaller proportion that varied from 13% to 1% of the total uncertainty when tree height increased. Errors associated with model parameters were negligible, yielding < 1% of the total uncertainty and did not change with tree height.

4.5. Plot-level biomass and components of its uncertainty

For the Lassen dataset, 5101 trees were correctly detected and measured by LiDAR out of a total of 8313 trees that were measured in the field, resulting in a detection rate of 61.36% and an omission rate of 38.64%. Since the ITD algorithm occasionally splits a single crown into multiple segments, those segments that fail to be linked with any field-measured trees are commission errors and accounted for 24.61% of the field-measured trees. Fig. 5a shows that on average, LiDAR underestimated 38.64% of the number of stems measured in the field, which corresponds to the omission rate reported earlier. The per-plot number of stems mapped by LiDAR has a RMSE of 31.61 (55.52%) and a bias of 22 (38.64%).

The omission-commission model was used to correct the LiDAR direct detection of AGB and to derive the AGB uncertainty associated with omission-commission errors. Three LiDAR height metrics (the maximum height, 20% and 50% height percentiles), three LiDAR density metrics (40% and 90% canopy area percentiles, canopy cover of all trees taller than 20 m), number of tree crowns and total AGB of tree crowns were eventually selected as predictor variables in the omission-commission model. Fig. 5b shows the omission-commission model fit, with model residuals approximating a normal distribution. The AGB uncertainty associated with omission-commission errors was the RMSE after a leave-one-out cross validation and was found to be 74.24 Mg/ha which indicated the underlying assumption that airborne laser scanning has a consistent performance in detecting and measuring trees in this area. Fig. 5c shows the LiDAR-derived AGB vs. field-based AGB with and without the omission-commission correction. Compared with the LiDAR directly detected AGB (without correction), the LiDAR derived AGB with correction decreased RMSE from 49.52% to 47.95%, but increased bias (underestimation) from 25.26% to 28.15%.

Once all the errors considered at the plot level were incorporated, we found the total uncertainty in AGB increased along with an increase in AGB (Fig. 6a). The mean coefficient of variation (Fig. 6b) across the LiDAR-derived AGB was 214% with the standard deviation of 8.69%. Sites with a lower AGB magnitude had a proportionally larger error than sites with larger AGB. When AGB reached 200 Mg/ha, the coefficient of variation stabilized around 50%.

Across the range of plot-level AGB found in the Lassen dataset, the two errors associated with remote sensing (omission-commission errors and the LiDAR height measurement errors) accounted for, on average, 48.68% and 0.87% of the total uncertainty in the LiDAR-derived AGB respectively, with a range from 0.83% to 98.19%, and a range from 0.13% to 1.25% respectively. The remaining 50.45% of the uncertainty was due to the errors from the tree-level allometric equation (Fig. 6c). As plot-level AGB increases, the proportion of errors due to the allometric equation increases while the proportion contributed by remote

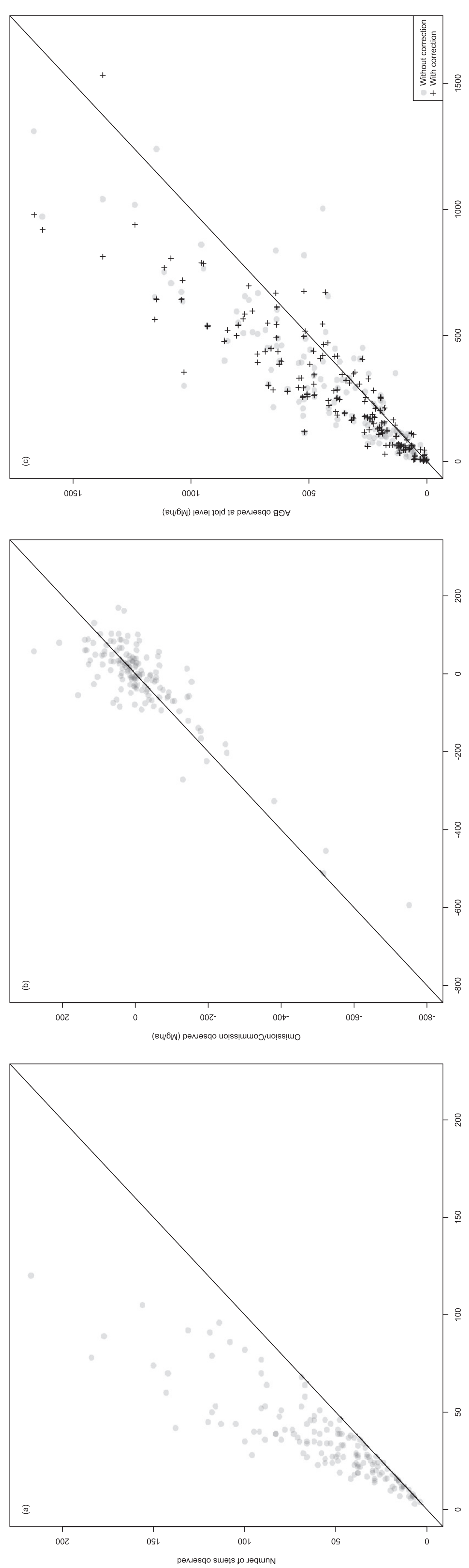

Fig. 5. Omission and commission errors. a. LiDAR omitted one-third of trees observed in the Lassen National Forest. b. Modeled AGB associated with omission-commission errors correspond well to the observed values. c. LiDAR-derived AGB at the plot level with correction slightly outperformed the version without correction.

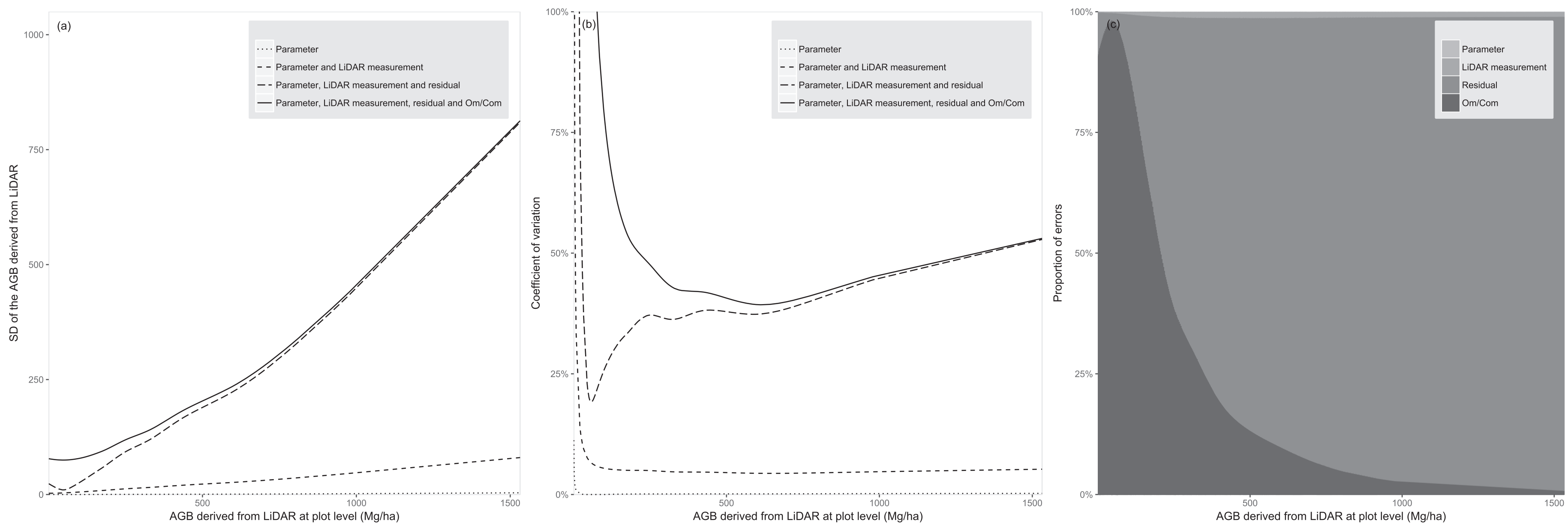

Fig. 6. Biomass uncertainty at the plot level. a. Uncertainty in the LiDAR-derived AGB at the plot level increases with AGB. b. Coefficient of variation is higher at lower-biomass forests, but it decreases to about 50% when AGB reaches 200 Mg/ha. Decomposition of uncertainty into four error sources shows that the largest contribution is from model residuals, the second from omission-commission errors, the third from LiDAR height measurements and a negligible contribution from model parameters.

sensing errors decreases. The allometric equation errors and the omission-commission errors are approximately equal when the AGB reaches 200 Mg/ha. As AGB increases beyond approximately 200 Mg/ha, the allometric equation errors became the dominant source of total uncertainty. When AGB reaches 500 Mg/ha, the uncertainty proportion due to remote sensing errors dropped to < 15%.

4.6. Application of the methods to California LiDAR datasets

We implemented the uncertainty analysis framework on a set of 52 LiDAR acquisitions over California covering over 18,780 km^2. Errors in the LiDAR-derived heights, omission-commission errors modeled using the LiDAR-field dataset collected at the Lassen National forest were assumed to represent the forests in the state as a whole.

At the tree level, the CA-wide analysis yielded a set of individual tree crown maps with the following attributes: 1) maximum crown height, 2) x,y position of maximum crown height, 3) crown diameters (maximum width and perpendicular), 4) distance to the nearest tree, 5) estimated aboveground biomass in kg, and 6) error in AGB estimate in kg. In total, 789,983,802 individual trees were identified, with a mean AGB of 494.88 kg (standard deviation of 1189.16 kg). The mean coefficient of variation is 138% with a standard deviation of 0.02%.

After taking into consideration the additional information on omission-commission errors, tree-level AGB (Fig. 7) and their uncertainties were aggregated into a 30 m ground sampling distance (GSD). At the pixel level, we found the LiDAR footprints yielded a mean AGB of 276.33 Mg/ha (standard deviation of 336.39 Mg/ha). We summarized mean, standard deviation and coefficient of variation for these footprints (Appendix F), stratified by California Wildlife Habitat Relationship (WHR) types that are existing vegetation types important to wildlife (Mayer and Laudenslayer, 1988). By WHR type, "Sierran Mixed Conifer" forests were found to have the highest AGB and lowest error (351 Mg/ha, CV 43%) and "Barren" the lowest AGB and highest error (48 Mg/ha, CV 171%).

The individual tree crown polygons, 30 m AGB maps, and corresponding canopy height models are archived at the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) (Xu et al., 2017, https://doi.org/10.3334/ORNLDAAC/1537).

5. Discussion

Our analysis focused on the question: which error sources contribute the most to the total AGB uncertainty at both tree and plot levels when performing ALS-based mapping of AGB, errors associated with remote sensing processing or errors associated with allometric models? After we considered three error sources at the tree level, we found the residual variance of the allometric equation was responsible for the largest proportion of per-tree uncertainty, whereas variance in LiDAR height measurement had a smaller contribution. This suggests that tree height alone did not explain sufficient variability in AGB, especially for tall trees. Additional predictors such as crown variables and species may assist in strengthening the predictability of the height-based allometric equation (Duncanson et al., 2015). Even after we integrated the remote sensing-associated omission and commission errors with the tree-level error sources at the plot level, we found that the allometric equation continued to dominate AGB uncertainty. Errors related to the remote sensing analysis were secondary to the allometric errors and became less important as plot-level AGB increased.

Previous ALS-based biomass studies were under the framework of area-based approaches, in which LiDAR metrics were related to the field-based AGB estimates via a regression model at the plot level (Montesano et al., 2014; Chen et al., 2015). This study is among the first cases of AGB uncertainty analysis based on the individual tree detection method that starts with a generalized allometric equation to assess per-tree AGB and their uncertainties, which were subsequently scaled up into a plot and even pixel level while integrating the remote sensing-associated omission and commission errors. This new framework allowed us to propagate two error sources from remote sensing performance into AGB uncertainty: 1) LiDAR uncertainty in measuring tree height was analyzed in terms of its contribution to the total uncertainty at both tree and plot level and 2) LiDAR uncertainty in detecting tree stems was taken into consideration in estimating AGB at the plot and pixel level. These two error sources were found to have a lower contribution to AGB uncertainty than error sources from the allometric equation. This conclusion suggests that it may be more worthwhile to devote future research efforts into reducing uncertainties in allometric equations than into remote sensing techniques, and that AGB mapping approaches that did not fully consider allometric uncertainties may have published overly optimistic accuracies.

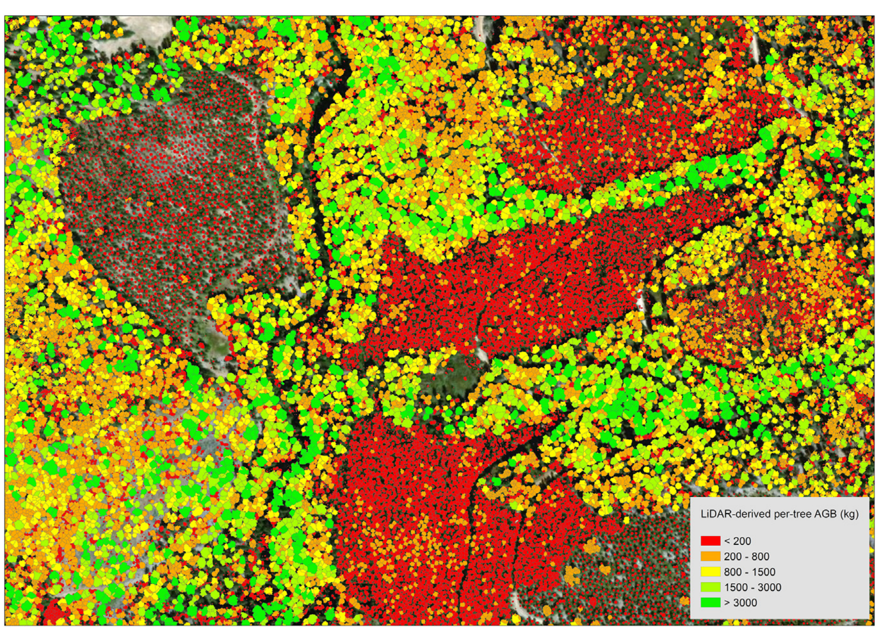

Fig. 7. Per-tree aboveground biomass (kg) derived from LiDAR-measured tree heights (Basemap source: Esri, DigitalGlobe, GeoEye, i-cubed, USDA FSA, USGS, AEX, Getmapping, Aerogrid, IGN, IGP, swisstopo, and the GIS User Community).

A larger uncertainty than previously reported at the tree level was found when field-based AGB mapping was replaced by airborne laser scanning. Based on the generalized allometric equation, we found that with a mean coefficient of variation for individual trees in the Lassen National Forest at 135%, the uncertainty in the per-tree AGB was larger than AGB estimate itself. Chave et al. (2004) used the same measure to assess biomass uncertainty for tropical forests based on field measurements and the allometric equations that related height, DBH and wood density to AGB. They found that the biomass uncertainty at the tree level is 48% of the estimated AGB for trees of diameter 10 cm or greater, and 78% of the estimated AGB for trees of diameter < 10 cm. Chen et al. (2015) reported a coefficient of variation of 50% for trees in tropical forests as well. When airborne LiDAR was introduced to replace the field survey of per-tree biomass, we found an uncertainty at least twice as large as reported by Chave et al. (2004) and Chen et al. (2015). By comparing the three studies for the contribution of residual variance to the total uncertainty, we found that the error associated with residuals of the allometric equation in our study accounted for a larger proportion (87–99%) than Chave et al. (2004) and Chen et al. (2015) (58–66%). This suggests that the height-based allometric equation failed to explain about 30% of variance in AGB, compared with allometric equations that used DBH and species. However, these parameters are currently inaccessible through airborne laser scanning. Without breakthroughs in retrieving either DBH or species from airborne laser scanning, it may be challenging to reduce uncertainties in LiDAR-derived per-tree AGB. Another reason for larger uncertainties found in our study is related to the semi-synthetic AGB we used to fit the height-based allometric equation. Unlike other remote sensing studies that utilized estimated AGB from allometric equations for reference, we utilized simulated AGB that allowed us to analyze the uncertainty with respect to actual AGB and to have a more comprehensive insight into per-tree uncertainties. Finally, our assumption of independent error components in the error propagation may have increased our variance estimates.

Smaller and even negligible errors were found to be associated with LiDAR-derived heights and model parameters. We benefited from calibrating the height-based allometric equation using the massive FIA dataset, as the variance of the estimated model parameters decreases as sample size increases. The use of the whole California-wide FIA dataset also generalizes the applicability of the allometric model to the whole state, although this biomass model was not validated on real tree AGB. The error associated with LiDAR height measurements, calculated from the Lassen dataset, contributed on average only 5% of the total uncertainty, and remained constant for trees taller than 20 m (Fig. 3b). The Lassen dataset was more than sufficient to represent the other California forests (the other 51 LiDAR acquisitions) in terms of the errors associated with LiDAR-derived tree height. Previous studies have demonstrated that tree height could be accurately measured by airborne LiDAR across a diverse set of forest ecosystems (Maltamo et al., 2004; Jakubowski et al., 2013). Furthermore, Vauhkonen et al. (2008) showed that accuracy of LiDAR-measured tree dimensions is largely invariant to changes in point density when the original point cloud was thinned from 40 to 3 points/m^2.

Model extrapolation is an important issue to consider when applying an allometric equation to trees of interest because it introduces extrapolation error in the estimated biomass. Chave et al. (2004) used a pan-tropical equation for trees having extrapolation issue (DBH > 148 cm), and decreased standard deviation of the estimated biomass by 3% of the mean biomass. Extrapolation error arises from unrepresentative samples and occurs when the training dataset does not cover the range of the testing dataset. The California FIA dataset used in this study to fit the height-based allometric equation has > 208,000 trees with DBHs ranging from 12 cm to 537 cm, and heights ranging from 1 m to 98 m. This model is general enough to cover most trees in the Lassen National Forest except for younger trees of DBH < 12 cm. Chave et al. (2004) used a separate allometric equation for trees of diameter < 10 cm because of the discrepancy in allometry of young trees from mature trees. We applied the height-based allometric equation to younger trees identified in our analysis, but observed a larger empirical uncertainty than eventually modeled since they exceeded range of the FIA training data. The extrapolation errors were demonstrated in Fig. 4(a) for trees with height < 16 m in the Lassen dataset, where the per-tree variance estimator underestimated the empirical variance. One possible solution to the extrapolation issue is to fit a separate allometric equation for younger trees, and to derive a separate variance estimator that works better for younger trees.

At the plot and pixel level, we went a step beyond individual tree errors and included omission and commission errors in the plot-level biomass estimation. Watershed segmentation based on a CHM mainly captures information of dominant canopies, but often results in a large number of omission errors for forests with significant understory vegetation, and commission errors originating from split crowns in dominant canopy layers. Although reducing omission and commission errors from the perspective of remote sensing algorithm development is currently an active research topic, we found that these errors could be reasonably modeled with area-based LiDAR metrics that reflect vertical canopy layers of the Lassen National Forest. However, the assumption on which we extended the omission-commission model to the whole California is worth discussing because the Lassen does not fully represent the diversified ecosystems in California in terms of species composition, as well as vertical and horizontal forest structure. Differences in species composition and forest structure can have a large impact on omission and commission errors. For instance, deciduous forests, which are underrepresented in our training set, are more prone to omission and commission errors because individual canopies are more difficult to be identified by LiDAR (Duncanson et al., 2014), resulting in larger commission errors than coniferous forests. Failure to cover deciduous forests intensifies the risk of less confident extrapolation of the omission-commission model into deciduous forests, leading to inaccurate estimation of plot-level AGB and its uncertainties. In addition, the omission-commission model did not extend well to sparsely populated plots since neither had we similar plots to train our model on.

We found that the uncertainty in plot-level AGB estimates was 214% of the mean AGB in the Lassen National Forest, two or four times larger than those found in previous remote sensing-based biomass studies. Montesano et al. (2014) obtained 50–100% error for forest biomass levels of < 80 Mg/ha while using airborne laser scanning. Chen et al. (2015) reported a pixel-level uncertainty at 49% of the mean AGB for tropical forests with trees of diameter no < 10 cm. We attribute the larger uncertainty in our analysis mainly to a wider range of tree sizes characterized by the Lassen National Forest, as well as the underrepresentation of the Lassen dataset for low biomass forest sites. If we exclude these low biomass plots (< 74.24 Mg/ha, the variability in omission-commission errors), the mean coefficient of variation was lowered into 47%, very similar to what was reported by Chen et al. (2015).

6. Conclusions

By scaling up individual tree AGB and its uncertainty to the landscape, we found 1) at the tree level, the contribution from the residual variance of the allometric equation to total AGB uncertainty was the largest, 2) per-tree AGB uncertainty increased with tree height, 3) the allometric equation continued to dominate the total AGB uncertainty at the plot level, 4) plot-level AGB uncertainty increased with plot AGB, 5) at the plot level, remote sensing-associated uncertainty was larger at lower-biomass forests, and decreased as AGB increased, 6) at the landscape scale, Sierran mixed conifer forests were found to have the highest AGB but the lowest uncertainty.

Acknowledgements

The study is funded by NASA Carbon Monitoring System (proposal number: 14-CMS14-0048) under the title of "Reducing uncertainties in estimating California's forest carbon stocks". We thank James Balamuta, Jiancong Zhu and Zhen Zhuo at the University of Illinois at Urbana-Champaign for their work at the earlier stage of the project, and Kirk Evans and Carol Clark from USDA Forest Service for sharing knowledge and key datasets.

In particular, we greatly appreciate the contributions of the late Dr. Brian Wing to this research. One of his career goals that he was very passionate about was to help improve land managers' understanding of how LiDAR could be used to address forest resource inventory and monitoring needs. We view this research as an extension of his vision. He is greatly missed.

Appendix A

Estimation of residual error of published equations using R^2 and FIA data. For a linear regression equation $w = \widehat{w} + \epsilon$, we can obtain estimates of the residual error variance $\sigma_\epsilon{}^2 = Var(\epsilon)$ if we have the equation R^2 as well as $Var(\widehat{w})$. Since we do not have $Var(\widehat{w})$, we instead estimate $Var(\widehat{w})$ from the FIA data. We assume that the distribution of fitted AGB in the FIA data is the same as the distribution of fitted AGB for the published equation, and we predict AGB of FIA trees using the published equation.

$$R^2 = 1 - \frac{Var(\epsilon)}{Var(w)}$$
$$R^2 = 1 - \frac{Var(\epsilon)}{Var(\widehat{w}) + Var(\epsilon)}$$
$$R^2 - 1 = -\frac{Var(\epsilon)}{Var(\widehat{w}) + Var(\epsilon)}$$
$$(1 - R^2)(Var(\widehat{w}) + Var(\epsilon)) = Var(\epsilon)$$
$$(1 - R^2)Var(\widehat{w}) = R^2 Var(\epsilon)$$
$$Var(\epsilon) = \frac{1 - R^2}{R^2} Var(\widehat{w})$$

Appendix B

First-order Taylor Expansion.

$$ln(b_i) = \beta_0 + \beta_1\sqrt{h_i} + \epsilon_i$$
$$ln(b_i) = f(h_i, \boldsymbol{\beta}) + \epsilon_i$$
$$b_i = exp(f(h_i, \boldsymbol{\beta})) \times exp(\epsilon_i)$$
$$b_i = g(h_i, \boldsymbol{\beta}, \epsilon_i)$$
$$b_i \approx g(\mathrm{lh}_i, \widehat{\boldsymbol{\beta}}, \widehat{\epsilon}_i) + \frac{dg}{dh} \times (h_i - \mathrm{lh}_i) + \frac{dg}{d\boldsymbol{\beta}} \times (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \frac{dg}{d\epsilon} \times (\epsilon_i - \widehat{\epsilon}_i)$$

Appendix C

Variance decomposition from the Taylor Expansion in Appendix B.

$$Var(b) \approx \left(\frac{dg}{dh}\right)^2 \times Var(\mathrm{h} - \mathrm{lh}) + \left(\frac{dg}{d\boldsymbol{\beta}}\right)^2 \times Var(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \left(\frac{dg}{d\epsilon}\right)^2 \times Var(\epsilon - \widehat{\epsilon})$$
$$\approx \left(\frac{dg}{dh}\right)^2 \times Var(\mathrm{h} - \mathrm{lh}) + \left(\frac{dg}{d\boldsymbol{\beta}}\right)^2 \times Var(\widehat{\boldsymbol{\beta}}) + \left(\frac{dg}{d\epsilon}\right)^2 \times Var(\widehat{\epsilon})$$

h differs for each tree, hence we calculate $Var(h - \mathrm{lh})$ rather than $Var(lh)$. However, β and ϵ are not specific to a single tree.

Appendix D

Partial derivatives and calculations for quantities used in Appendix C.

$$\left.\frac{dg}{dh}\right|_{h=lh, \boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \epsilon=\widehat{\epsilon}} = \frac{\widehat{\beta}_1}{2\sqrt{lh}} exp(\widehat{\beta}_0 + \widehat{\beta}_1\sqrt{lh})$$
$$\left.\frac{dg}{d\beta_0}\right|_{h=lh, \boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \epsilon=\widehat{\epsilon}} = exp(\widehat{\beta}_0 + \widehat{\beta}_1\sqrt{lh})$$
$$\left.\frac{dg}{d\beta_1}\right|_{h=lh, \boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \epsilon=\widehat{\epsilon}} = \sqrt{lh} \times exp(\widehat{\beta}_0 + \widehat{\beta}_1\sqrt{lh})$$
$$\left.\frac{dg}{d\epsilon}\right|_{h=lh, \boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \epsilon=\widehat{\epsilon}} = exp(\widehat{\beta}_0 + \widehat{\beta}_1\sqrt{lh})$$
$$\nabla_\beta{}^2 g \times Cov(\beta) = \nabla^T{}_\beta g \times \Sigma_\beta \times \nabla_\beta g$$

where

$$\nabla_{\beta} g = \left(\frac{dg}{d\beta_0} \bigg|_{h=lh, \boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \varepsilon=\widehat{\varepsilon}}, \frac{dg}{d\beta_1} \bigg|_{h=lh, \boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}, \varepsilon=\widehat{\varepsilon}} \right)^T$$

and

Σ_β is the covariance matrix of the estimated parameters from the GAE.

Appendix E

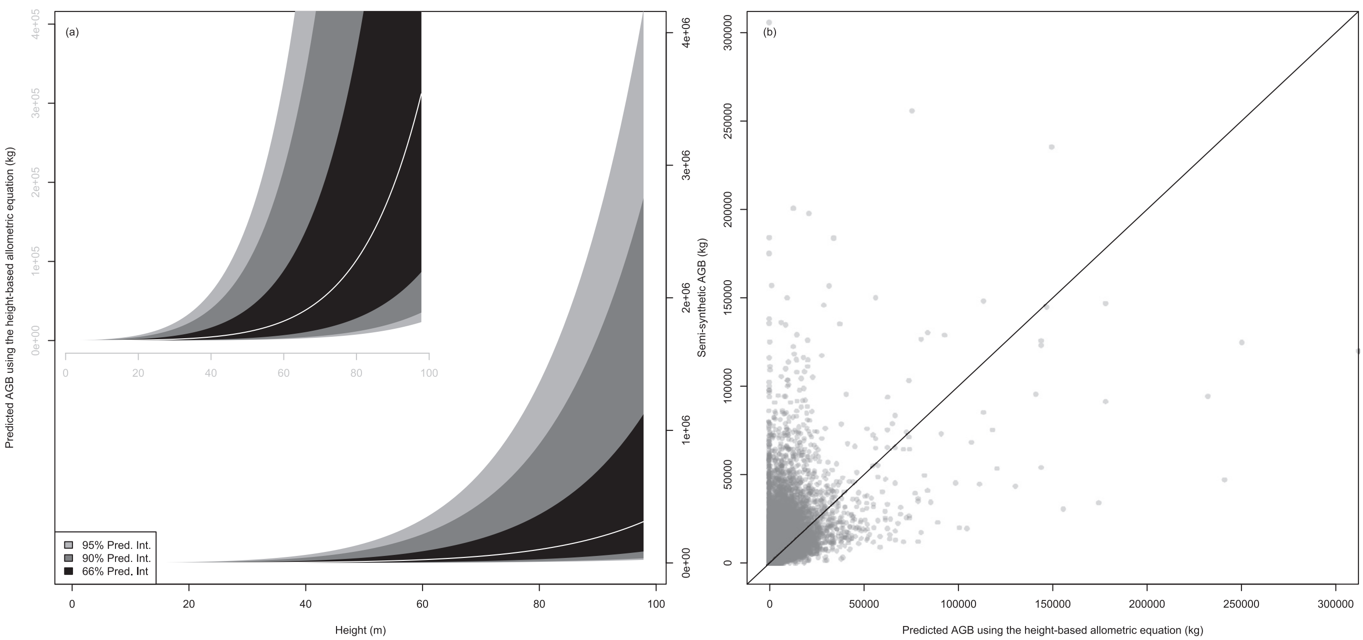

Fig. E. The height-based allometric equation. a. Confidence interval of the height-based allometric equation calibrated using Californian FIA dataset, with the inset zooming in on a smaller magnitude. b. The height-based allometric equation underestimates the semi-synthetic AGB.

Appendix F

Table F1
Summary of AGB (Mg/ha) and uncertainty according to the California WHR types.

ID	WHRTYPE	WHRNAME	Mean AGB	Mean sd	Mean CV (%)
1	ASP	Aspen	154.83	98.55	64
2	BOP	Blue oak-foothill pine	121.57	90.76	75
3	BOW	Blue oak woodland	110.76	85.47	77
4	COW	Coastal oak woodland	138.00	95.52	69
5	CPC	Closed-cone pine-cypress	142.06	92.79	65
6	DFR	Douglas-fir	275.36	125.15	45
7	DRI	Desert riparian	227.34	118.21	52
8	EPN	Eastside pine	225.71	118.55	53
9	EUC	Eucalyptus	226.29	115.92	51
10	JPN	Jeffrey pine	213.57	112.80	53
11	JST	Joshua tree	208.71	110.48	53
12	JUN	Juniper	205.68	111.04	54
13	KMC	Klamath mixed conifer	205.63	112.04	54
14	LPN	Lodgepole pine	224.33	115.90	52
15	MHC	Montane hardwood-conifer	235.14	116.67	50
16	MHW	Montane hardwood	181.70	100.55	55
17	MRI	Montane riparian	200.17	110.02	55
18	PJN	Pinyon-juniper	212.13	112.01	53
19	POS	Palm oasis	231.96	118.66	51
20	PPN	Ponderosa pine	239.64	122.55	51
21	RDW	Redwood	250.17	126.82	51

(*continued on next page*)

Table F1 (*continued*)

ID	WHRTYPE	WHRNAME	Mean AGB	Mean sd	Mean CV (%)
22	RFR	Red fir	278.25	130.58	47
23	SCN	Subalpine conifer	254.63	124.17	49
24	SMC	Sierran mixed conifer	350.88	150.27	43
25	VOW	Valley oak woodland	214.20	116.40	54
26	VRI	Valley foothill riparian	223.28	118.68	53
27	WFR	White fir	255.88	123.57	48
28	ADS	Alpine-dwarf shrub	162.96	103.78	64
29	ASC	Alkali desert scrub	160.66	103.15	64
30	BBR	Bitterbrush	161.25	105.15	65
31	CRC	Chamise-redshank chaparral	146.19	101.06	69
32	CSC	Coastal scrub	151.61	101.67	67
33	DSC	Desert scrub	137.39	98.85	72
34	DSS	Desert succulent shrub	140.01	99.99	71
35	DSW	Desert wash	131.95	96.45	73
36	LSG	Low sage	119.33	94.72	79
37	MCH	Mixed chaparral	94.61	89.48	95
38	MCP	Montane chaparral	85.12	87.53	103
39	SGB	Sagebrush	102.81	91.22	89
40	AGS	Annual grassland	91.46	91.31	100
41	FEW	Freshwater emergent wetland	120.52	95.37	79
42	PAS	Pasture	116.41	95.42	82
43	PGS	Perennial grassland	95.32	91.64	96
44	SEW	Saline emergent wetland	107.39	93.28	87
45	WTM	Wet meadow	84.97	90.68	107
46	EST	Estuarine	87.98	89.40	102
47	LAC	Lacustrine	67.72	90.17	133
48	MAR	Marine	93.39	89.56	96
49	RIV	Riverine	79.11	85.35	108
50	CRP	Cropland	87.64	86.98	99
51	DGR	Dryland grain crops	84.69	85.11	100
52	DOR	Deciduous orchard	82.47	89.20	108
53	EOR	Evergreen orchard	79.96	87.56	110
54	IGR	Irrigated grain crops	76.27	85.11	112
55	IRF	Irrigated row and field crops	73.30	85.25	116
56	IRH	Irrigated hayfield	71.08	84.21	118
57	OVN	Orchard and vineyard	76.45	88.93	116
58	RIC	Rice	70.68	85.61	121
59	URB	Urban	71.53	86.81	121
60	VIN	Vineyard	67.71	82.79	122
61	BAR	Barren	48.30	82.65	171

References

Axelsson, P., 1999. Processing of laser scanner data-algorithms and applications. ISPRS J. Photogramm. Remote Sens. 54, 138–147.

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

Brandtberg, T., 1999. Automatic Individual Tree-Based Analysis of High Spatial Resolution Remotely Sensed Data (Doctoral Thesis). Acta Universitatis Agriculturae Sueciae, Silvestria 118. Swedish University of Agricultural Sciences, Centre for Image Analysis, Uppsala, Sweden.

Brandtberg, T., 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne LiDAR. ISPRS J. Photogramm. Remote Sens. 61, 325–340.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., Solberg, S., 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. Remote Sens. Environ. 114 (4), 911–924.

Breidenbach, J., Antón-Fernandez, C., Petersson, H., McRoberts, R.E., Astrup, R., 2014. Quantifying the model-related variability of biomass stock and change estimates in the Norwegian national forest inventory. For. Sci. 60 (1), 25–33.

Brown, S., 1997. Estimating Biomass and Biomass Change of Tropical Forests: A Primer (FAO Forestry Paper-134). FAO, United Nations, Rome.

Brown, S., Lugo, A.E., 1982. The storage and production of organic matter in tropical forests and their role in the global carbon cycle. Biotropica 14 (3), 161–187.

Brown, S., Gillespie, A.J.R., Lugo, A.E., 1989. Biomass estimation methods for tropical forests with applications to forest inventory data. For. Sci. 35, 881–902.

CAL FIRE, 2017. http://calfire.ca.gov/Grants/downloads/UrbanForestry/2017-2018/CAL%20FIRE%2017-2018%20GGRF%20Forest%20Health%20Program%20Procedural%20Guide%20FINAL%2012_19_17.pdf (assessed 2018.03.12).

California Assembly Bill 32, 2006. California Government Code 38500.1. (Sacramento, California, State of California).

CARB, 2018. https://www.arb.ca.gov/cc/capandtrade/auctionproceeds/calfire_fh_draftqm_17-18.pdf (assessed 2018.03.12).

Chave, J., Riéra, B., Dubois, M.A., 2001. Estimation of biomass in a neotropical forest of French Guiana: spatial and temporal variability. J. Trop. Ecol. 17, 79–96.

Chave, J., Condit, R., Lao, S., Caspersen, J.P., Foster, R.B., Hubbell, S.P., 2003. Spatial and temporal variation in biomass of a tropical forest: results from a large census plot in Panama. J. Ecol. 91, 240–252.

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., Perez, R., 2004. Error propagation and scaling for tropical forest biomass estimates. Philos. Trans. R. Soc. Lond. B 359, 409–420.

Chave, J., Andalo, C., Brown, S., Cairns, M., Chambers, J.C., Eamus, D., Fölster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J., Nelson, B.W., Ogawa, H., Puig, H., Riéra, B., Yamakura, T., 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. Oecologia 145 (1), 87–99.

Chen, Q., Laurin, G.V., Valentini, R., 2015. Uncertainty of remotely sensed aboveground biomass over an African tropical forest: propagating errors from trees to plots to pixels. Remote Sens. Environ. 160, 134–143.

Clark, D.A., 2002. Are tropical forests an important carbon sink? Reanalysis of the long-term plot data. Ecol. Appl. 12, 3–7.

Clark, D.B., Kellner, J.R., 2012. Tropical forest biomass estimation and the fallacy of misplaced concreteness. J. Veg. Sci. 23, 1191–1196.

Duncanson, L.I., Cook, B.D., Hurtt, G.C., Dubayah, R.O., 2014. An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems. Remote Sens. Environ. 154, 378–386.

Duncanson, L., Rourke, O., Dubayah, R., 2015. Small sample sizes yield biased allometric equations in temperate forests. Sci. Rep. 5, 17,153.

Gauch, J.M., 1999. Image segmentation and analysis via multiscale gradient watershed hierarchies. IEEE Trans. Image Process. 8 (1), 69–79.

Gertner, G., Cao, X., Zhu, H., 1995. A quality assessment of a Weibull based growth projection system. For. Ecol. Manag. 71 (3), 235–250.

GlobAllomeTree, 2013. http://www.globallometree.org/about/ (accessed 2017.08.10).

Gonzalez, P., Asner, G.P., Battles, J.J., Lefsky, M.A., Waring, K.M., Palace, M., 2010. Forest carbon densities and uncertainties from LiDAR, QuickBird, and field measurements in California. Remote Sens. Environ. 114 (7), 1561–1575.

Gregoire, T.G., 1998. Design-based and model-based inference in survey sampling: appreciating the difference. Can. J. For. Res. 28 (10), 1429–1447.

Gregoire, T.G., Ståhl, G., Næsset, E., Gobakken, T., Nelson, R., Holm, S., 2011. Model-assisted estimation of biomass in a LiDAR sample survey in Hedmark County, Norway. Can. J. For. Res. 41, 83–95.

Hou, Z., Xu, Q., Vauhkonen, J., Maltamo, M., Tokola, T., 2016. Species-specific combination and calibration between area-based and tree-based diameter distributions using airborne laser scanning. Can. J. For. Res. 46, 753–765.

Hou, Z., Xu, Q., McRoberts, R.E., Greenberg, J.A., Liu, J., Heiskanen, J., Pitkänen, S., Packalen, P., 2017. Effects of temporally external auxiliary data on model-based inference. Remote Sens. Environ. 198, 150–159.

Hou, Z., McRoberts, R.E., Ståhl, G., Packalen, P., Greenberg, J., Xu, Q., 2018. How much can natural resource inventory benefit from a finer resolution auxiliary data? Remote Sens. Environ. 209, 31–40.

Hyyppä, J., Inkinen, M., 1999. Detecting and estimating attributes for single trees using laser scanner. Photogramm. J. Finland. 16 (2), 27–42.

Isenburg, M., Liu, Y., Shewchuk, J., Snoeyink, J., Thirion, T., 2006. Generating raster DEM from mass points via TIN streaming. In: Geographic Information Science: 4th International Conference, GIScience 2006. Münster, Germany, September, pp. 186–198.

Jakubowski, M.K., Guo, Q., Kelly, M., 2013. Tradeoffs between lidar pulse density and forest measurement accuracy. Remote Sens. Environ. 130 (15), 245–253.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2003. National-scale biomass estimators for United States tree species. For. Sci. 49 (1), 12–35.

Kaartinen, H., Hyyppä, J., 2008. Commission II “Tree Extraction”. (Final Report of ISPRS Project EuroSDR. European Spatial Data Research, Official Publication No. 53).

Kangas, A., 1994. Classical and model based estimators for forest inventory. Silva Fenn. 28 (1), 3–14.

Khosravipour, A., Skidmore, A.K., Isenburg, M., Wang, T.J., Hussin, Y.A., 2014. Generating pit-free canopy height models from airborne LiDAR. Photogramm. Eng. Remote. Sens. 80 (9), 863–872.

Korpela, I., Ørka, H.O., Maltamo, M., Tokola, T., Hyyppä, J., 2010. Tree species classification using airborne LiDAR – effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. Silva Fenn. 44 (2), 319–339.

Lambert, M.C., Ung, C.H., Raulier, F., 2005. Canadian national tree aboveground biomass equations. Can. J. For. Res. 35 (8), 1996–2018.

Liu, C.J., 1980. Log volume estimation with spline approximation. For. Sci. 26 (3), 361–369.

Malhi, Y., Baker, T.R., Phillips, O.L., Almeida, S., Alvarez, E., Arroyo, L., Chave, J., Czimczik, C.I., Fiore, A., Higuchi, N., et al., 2004. The above-ground coarse wood productivity of 104 neotropical forest plots. Glob. Chang. Biol. 10, 563–591.

Maltamo, M., Mustonen, K., Hyyppä, J., Pitkänen, J., Yu, X., 2004. The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. Can. J. For. Res. 34 (9), 1791–1801.

Marvin, D.C., Asner, G.P., 2016. Spatially explicit analysis of field inventories for national forest carbon monitoring. Carbon Balance Manag. 11 (1), 1–12.

Mayer, K.E., Laudenslayer, W.F., 1988. A Guide to Wildlife Habitats of California. 1988 State of California, Resources Agency, Department of Fish and Game, Sacramento, CA (166 pp).

McRoberts, R.E., Westfall, J.A., 2014. Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. For. Sci. 60 (1), 34–42.

Melson, S.L., Harmon, M.E., Fried, J.S., Domingo, J.B., 2011. Estimates of live-tree carbon stores in the Pacific Northwest are sensitive to model selection. Carbon Balance Manag. 6 (2), 16.

Montesano, P.M., Nelson, R.F., Dubayah, R.O., Sun, G., Cook, B.D., Ranson, K.J., Næsset, E., Kharuk, V., 2014. The uncertainty of biomass estimates from LiDAR and SAR across a boreal forest structure gradient. Remote Sens. Environ. 154, 398–407.

Næsset, E., 2004. Practical large-scale forest stand inventory using a small airborne scanning laser. Scand. J. For. Res. 19 (2), 164–179.

NASA Carbon Monitoring System, 2010. CMS scoping study workshop final report. https://carbon.nasa.gov/pdfs/2010%20CMS_Scoping%20Study%20Workshop%20Report.pdf (accessed 2017.08.10).

Olson, J.S., Watts, J.A., Allison, L.J., 1983. Carbon in live vegetation of major world ecosystems. In: Environmental Sciences Division Publication No. 1997. Oak Ridge National Laboratory, Oak Ridge.

Pitkänen, J., Maltamo, M., Hyyppä, J., Yu, X., 2004. Adaptive methods for individual tree detection on airborne laser based canopy height model. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 36 (8/W2), 187–191.

PNW-FIA Field Manual, 2015. https://www.fs.fed.us/pnw/rma/fia-topics/documentation/field-manuals/documents/Annual/2015_PFSL_FIA_Field_Manual.pdf (accessed 2017.08.10).

Popescu, S.C., Wynne, R.H., 2004. Seeing the trees in the forest: using LiDAR and multispectral data fusion with local filtering and variable window size for estimating tree height. Photogramm. Eng. Remote. Sens. 70 (5), 589–604.

Salas, C., Ene, L., Gregoire, T.G., Nasset, E., Gobakken, T., 2010. Modelling tree diameter from airborne laser scanning derived variables: a comparison of spatial statistical models. Remote Sens. Environ. 114 (6), 1277–1285.

Stocker, T.F., Qin, D., Plattner, G.-K., Alexander, L.V., Allen, S.K., Bindoff, N.L., Bréon, F.-M., Church, J.A., Cubasch, U., Emori, S., et al., 2013. Technical summary. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Doschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 33–115.

Vauhkonen, J., Tokola, T., Maltamo, M., Packalén, P., 2008. Effects of pulse density on predicting characteristics of individual trees of Scandinavian commercial species using alpha shape metrics based on airborne laser scanning data. Can. J. Remote. Sens. 34 (Suppl. 2), S441–S459.

Vauhkonen, J., Korpela, I., Maltamo, M., Tokola, T., 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. Remote Sens. Environ. 114 (6), 1263–1276.

Xu, Q., Hou, Z., Maltamo, M., Tokola, T., 2014. Calibration of area based diameter distribution with individual tree based diameter estimates using airborne laser scanning. ISPRS J. Photogramm. Remote Sens. 93, 65–75.

Xu, Q., Man, A., Frederickson, M., Hou, Z., Pitkänen, J., Wing, B., Ramirez, C., Li, B., Greenberg, J.A., 2017. CMS: LiDAR-derived Aboveground Biomass and Related Uncertainties in California, USA. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1537.