Remote Sensing of Environment 121 (2012) 108–117

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

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

Integration of airborne lidar and vegetation types derived from aerial photography for mapping aboveground live biomass

Qi Chen [a,*], Gaia Vaglio Laurin [b,c], John J. Battles [d], David Saah [e,f]

[a] Department of Geography, University of Hawai'i at Manoa, 422 Saunders Hall, 2424 Maile Way, Honolulu, HI 96822, USA
[b] Department of Computer, System and Production Engineering, University of Tor Vergata, Rome 00133, Italy
[c] CMCC - Centro Mediterraneo per i Cambiamenti Climatici, via Augusto Imperatore (Euro-Mediterranean Center for Climate Change), Lecce 73100, Italy
[d] Department of Environmental Science, Policy, and Management, 137 Mulford Hall, University of California at Berkeley, Berkeley, CA 94720, USA
[e] Spatial Informatics Group, LLC, 3248 Northampton Ct., Pleasanton, CA 94588, USA
[f] College of Arts and Sciences, Environmental Science, University of San Francisco, San Francisco, CA 94117, USA

ARTICLE INFO

Article history:
Received 3 August 2011
Received in revised form 29 December 2011
Accepted 25 January 2012
Available online xxxx

Keywords:
Biomass
Mixed-effects model
Airborne lidar
Aerial photos
Vegetation type

ABSTRACT

The relationship between lidar-derived metrics and biomass could vary across different vegetation types. However, in many studies, there are usually a limited number of field plots associated with each vegetation type, making it difficult to fit reliable statistical models for each vegetation type. To address this problem, this study used mixed-effects modeling to integrate airborne lidar data and vegetation types derived from aerial photographs for biomass mapping over a forest site in the Sierra Nevada mountain range in California, USA. It was found that the incorporation of vegetation types via mixed-effects models can improve biomass estimation from sparse samples. Compared to the use of lidar data alone in multiplicative models, the mixed-effects models could increase the R^2 from 0.77 to 0.83 with RMSE (root mean square error) reduced by 10% (from 80.8 to 72.2 Mg/ha) when the lidar metrics derived from all returns were used. It was also found that the SAF (Society of American Forest) cover types are as powerful as the NVC (National Vegetation Classification) alliance-level vegetation types in the mixed-effects modeling of biomass, implying that the future mapping of vegetation classes could focus on dominant species. This research can be extended to investigate the synergistic use of high spatial resolution satellite imagery, digital image classification, and airborne lidar data for more automatic mapping of vegetation types, biomass, and carbon.

1. Introduction

Vegetation biomass, the weight of plant materials that exist over an area, is a critical measure of ecosystem structure and productivity that informs a range of applications such as fire emission calculations (e.g., De Santis et al., 2010), wildlife habitat analysis (e.g., Morris et al., 2009), hydrological modeling (e.g., Ursino, 2007), and greenhouse gas accounting (e.g., De Jong et al., 2010). In particular, accurate estimates of biomass are needed in order to inform national policies and international treaties regarding forest management and carbon sequestration (Malmsheimer et al., 2011).

Lidar is a state-of-the-art remote sensing technology with a proven ability to map aboveground biomass (AGB). The accuracy and sensitivity of the metrics derived from optical and radar imagery (such as NDVI and backscatter coefficient) decline with increasing AGB (Waring et al., 1995). In contrast, vegetation height metrics derived from lidar have been found to be highly correlated to biomass even when the biomass density is very high (Gonzalez et al. 2010, Means et al., 1999). In the past, much research has been done to estimate AGB using airborne discrete-return lidar (e.g., Asner et al., 2009; Banskota et al., 2011; Lim et al., 2003), airborne profiling lidar (e.g., Nelson et al., 2009, 1988; Stahl et al., 2011), airborne waveform lidar (e.g., Dubayah et al. 2010; Lefsky et al., 1999; Ni-Meister et al., 2010), satellite lidar (e.g., Boudreau et al., 2008; Guo et al., 2010; Nelson et al., 2009), and ground-based lidar (e.g., Loudermilk et al., 2009; Ni-Meister et al., 2010). In these applications, statistical models were used to quantify the relationship between biomass measurements and vegetation structure metrics derived from lidar for a number of forest plots or stands. Their performance varies depending on the vegetation conditions, the density of field observations, and the approach used for statistical modeling.

Most of these existing studies have focused on the use of lidar-derived canopy structure metrics, such as height and canopy cover, for biomass estimation. However, studies of plant allometry suggested that biomass at the individual tree level is determined not only by canopy structure but also by factors such as trunk taper and wood density (Chave et al., 2006; Niklas, 1995), which are closely related to the floristic characteristics of the plants. As a result, biomass should be related to vegetation types. For example, Drake et al. (2003) examined the relationships between lidar metrics from an

* Corresponding author. Tel.: +1 808 956 3524; fax: +1 808 956 3512.
E-mail address: qichen@hawaii.edu (Q. Chen).

0034-4257/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2012.01.021

airborne waveform lidar LVIS (Laser Vegetation Imaging Sensor) and AGB for two study sites in Central America, one in a tropical moist forest in Panama and the other in a tropical wet forest in Costa Rica. They found that the relationships between lidar metrics and AGB differ between these two sites even after the models had adjusted for the fraction of crown area that was deciduous (FCAD) of canopy trees. They attributed the differences to the underlying allometric relationships between stem diameter and AGB in tropical forests. Næsset and Gobakken (2008) estimated the aboveground and belowground biomass for 1395 sample plots in young and mature coniferous forests located in ten different areas within the boreal forest zone of Norway. With one canopy height metric and one canopy density metric derived from airborne discrete-return lidar, they were able to estimate aboveground and belowground biomass with R^2 of 0.82 and 0.77, respectively. When variables including tree species composition were included, the R^2 increased to 0.88 and 0.85. In a recent study, Ni-Meister et al. (2010) found that the relationships between biomass and canopy structure are distinctly different for deciduous and conifer trees in temperate forests in New England, U.S. Their analysis was based on the canopy structure information measured in the field as well as those derived from LVIS and Echidna® validation instrument (EVI), a ground-based lidar system.

The dependence of biomass-canopy structure relationship on vegetation types is well-known (e.g., Nelson et al., 1988; Ni-Meister et al., 2010). One approach for incorporating vegetation type information into biomass estimation is to stratify the forest plots according to vegetation types, for each of which a separate statistical model is developed (e.g., MacLean and Krabill 1986; Nelson et al., 1988). However, such an approach has practical and theoretical limitations. First, in most previous studies, only a limited number (typically 20–60 in total) of field plots were available for biomass modeling due to issues such as accessibility and cost. The stratification of a study area will lead to even fewer number of field plots per vegetation type, making it difficult to fit reliable statistical models for each vegetation type. Another problem of such an approach is that it assumes that the field data contains an exhaustive list of all vegetation types which exist in a given area. This is hardly true for natural forests because the vegetation types collected through field measurements are typically only a sample of the all vegetation types which exist over that area.

Recent advances in mixed-effects modeling can circumvent the aforementioned problems. In a conventional statistical model, the regression coefficients (such as intercept and slopes) are treated as constants. However, in mixed-effects models, these coefficients could be modeled as random Gaussian variables with their specific values varying among vegetation types. This approach makes it feasible to estimate biomass even when the sample size per vegetation type is small. Mixed-effects models have recently been used to estimate canopy height from satellite lidar (GLAS) data (Chen, 2010) and tree diameter from airborne discrete-return lidar data (Salas et al., 2010). Chen (2010) used mixed-effects model to test the generalizability of height estimation from GLAS data within and across three study sites in the Pacific coast region (one conifer site and one woodland site in California and another conifer site in Washington). He found significant random effects between the conifer and woodland sites but not between the two conifer sites. Salas et al. (2010) compared four statistical models including ordinary least squares (OLS), generalized least squares with a non-null correlation structure (GLS), linear mixed-effects model (LME), and geographically weighted regression (GWR) for estimating diameter of individual trees using discrete-return lidar data. They found that LME was significantly better than the other three models. Despite the promising results obtained in these two lidar remote sensing studies, no studies, to our best knowledge, have been done to explore the use of mixed-effects model for biomass estimation using lidar data.

In this study, vegetation types derived from aerial photographs are used to stratify forest for biomass modeling. Aerial photography is a fundamental remote sensing data source that possesses fine spatial and temporal details for producing base maps and performing environmental analysis (Lillesand et al., 2008). It has been widely used for mapping vegetation types for decades (e.g., Avery, 1978; Colwell, 1946; Fensham and Fairfax, 2002; Morgan et al., 2010). The recent advances in digital imaging and analysis also make aerial photography a rapidly-evolving tool for environmental analysis and ecological management (Morgan et al., 2010). In the U.S., a number of national programs such as NHAP (National High Altitude Program), NAPP (National Aerial Photography Program), NAIP (National Agriculture Imagery Program), and NDOP (National Digital Orthophoto Program) have collected and delivered aerial photographs every 3–10 years that cover the conterminous states from the late 1980s. Besides their wide temporal and spatial coverage, the aerial photographs acquired through these programs are usually free or at low cost for public use, making them ideal for detailed vegetation type mapping (Davies et al., 2010; Higinbotham et al., 2004).

The main goal of this study is to investigate whether integrating airborne lidar data with traditional vegetation maps derived from aerial photographs can improve biomass estimation for forest landscape in California. We specifically explore the efficacy of mixed-effects modeling to integrate the two remotely sensed data sources. We also compare the performance of two common but different approaches to vegetation classification.

2. Study area and data

2.1. Study area

Our study area is located in the United States Forest Service Sagehen Creek Experimental Forest in California, which covers approximately 3925 ha and is on the eastern slope of the Sierra Nevada approximately 32 km north of Lake Tahoe (Fig. 1). Conifer species present include white fir (*Abies concolor*), red fir (*Abies magnifica*), mountain hemlock (*Tsuga mertensiana*), lodgepole pine (*Pinus contorta*), Jeffrey pine (*Pinus jeffreyi*), sugar pine (*Pinus lambertiana*), and western white pine (*Pinus monticola*) (Table 1). Non-forested areas include fens, wet and dry montane meadows and shrub fields. Elevation ranges from 1862 m to 2670 m with slopes averaging 18% but can reach 70% in parts of the watershed.

2.2. Field data collection

A systematic grid of geo-referenced 0.05 ha circular plots was installed with a random starting location (Fig. 2). The grid consists of three sampling densities, 500 m, 250 m, and 125 m spacing. The entire watershed was sampled by plots spaced on a 500 m interval. Areas not occupied by Jeffrey pine plantations were further sampled at 250 m spacing; 125 m spacing was used in 10 unique forest types to conduct high density sampling. A total of 523 plots were established in the field between 2004 and 2006. These field plots were located with a handheld Garmin eTrex recreational GPS with horizontal accuracy of 3 to 11 m, which are called RGPS plots hereinafter. Nine of the ten locations of 125 m plot spacing were revisited in 2006 and a Trimble® GeoXH™ handheld GPS with Zephyr Geodetic antenna was used to re-measure the center of 81individual plots. The average horizontal accuracy of the new GPS measurements is 0.1 m with the majority <0.2 m and, at the worse case, 1.5 m. These plots are DGPS plots hereinafter.

At each plot, all trees greater than 5 cm in diameter at breast height (DBH, breast height = 1.37 m) were measured with a nested sampling design. Canopy trees ($\geq$19.5 cm DBH) were tagged and measured in the whole plot; Understory trees ($\geq$5 cm DBH to $<$19.5 cm DBH) were measured in a randomly selected third of the plot. Tree measurements include species, DBH, tree height, and vigor. Vigor was defined into six different classes: 1) healthy trees with no visible defects, 2) healthy trees with minimal damage or

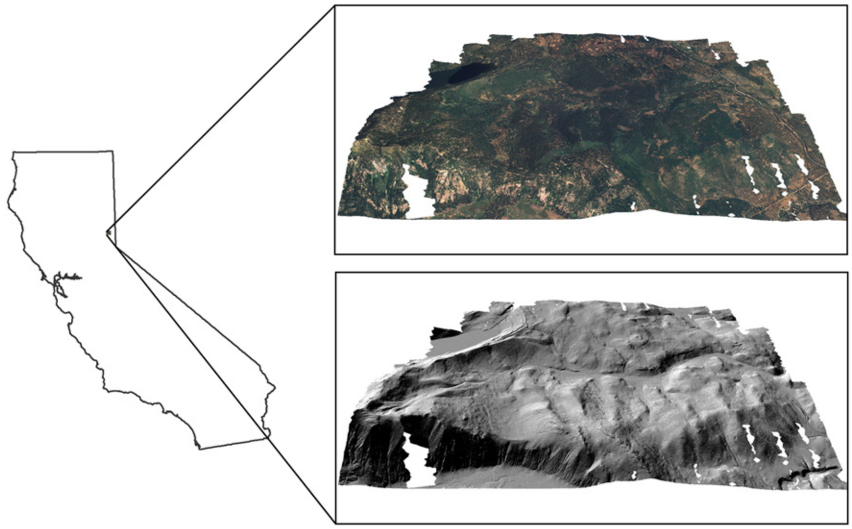

Fig. 1. Location of the study area. Top-right: Aerial photographs draped over the lidar DEM. Bottom-right: A hillshade of the lidar DEM.

defect (broken top/dead top, abnormal lean, etc.), 3) live trees that are near death or will be dead in the next five years, 4) recently dead trees with little decay and that retain their bark, branches and top, 5) trees that show some decay and have lost some bark, branches and may have a broken top, and 6) extensive decay and missing bark and most branches and have a broken top. The first three vigor classes are for live trees and the last three are for dead trees.

2.3. Lidar data

Lidar data were collected from September 14 to 17, 2005 for the study area using an Optech ALTM 2050 system on an airplane flying at an altitude of ~800 m and average velocity of 260 km per hour. The ALTM 2050 acquired up to three returns per pulse at a pulse frequency of 50 kHz, scan frequency of 38 Hz, and a maximum scan angle of 15°, creating a swath width of ~580 m. The point density is about 2–4 returns per square meter. Optech, Inc. rates the RMSE precision of individual point locations surveyed by the ALTM 2050 as $\pm$ 15 cm vertical and $\pm$50 cm horizontal.

2.4. Vegetation types from aerial photographs

USDA Forest Service (USFS) provided a vegetation type map, which was produced by visually interpreting 1 m NAIP (National Agricultural Imagery Program) Digital Orthophoto Quadrangles (scale 1:15,840, natural color) taken on September 16, 2005 and manually delineating the vegetation polygons. The vegetation polygons were initially typed using the CALVEG (Classification and Assessment with Landsat of Visible Ecological Groupings) classification system (USDA, 1981), which is a provisional system that meets the floristically based level of the U.S. National Vegetation Classification Standard (NVCS) hierarchy. The CALVEG system was designed to classify California's existing vegetation communities and the CALVEG types are also called "Dominant Types" in accordance with the USFS Existing Vegetation Classification and Mapping Technical Guide (Brohman and Bryant, 2005). The CALVEG types were crosswalked to other classification systems including SAF (Society of American Forester) (Eyre, 1980), CWHR (California Wildlife-Habitat Relationships) (Meyer and Laudenslayer, 1988), and U.S. NVC (National Vegetation Classification) alliance-level vegetation types (FGDC, 2008).

In this study, the two national-wide vegetation classification systems, SAF and NVC alliance-level vegetation types, were chosen for biomass estimation due to their broad applicability (Fig. 3). The NVC alliance-level vegetation types are based on NVCS, which establishes national procedures for field plot records and classification of existing vegetation types for the United States. These procedures provide a dynamic and practical way to publish new or revised descriptions of vegetation types while maintaining a current, authoritative list of types for multiple users to access and apply (Jennings et al., 2009). The early efforts of NVC started in 1994 and the first NVCS was adopted in 1997 by FDGC (Federal Geographic Data Committee). As early as of April 1997, a total of 1571 NVC types had been identified at the alliance-level (Grossman et al., 1998). Since then, the vegetation classification has been continuously evolving and updated (Jennings et al., 2009). In contrast to NVC that uses all vascular plant species present in a community to help define vegetation classes, the SAF types emphasize dominant species of a stand. In many cases, the SAF types are more broad-ranging over both structural and environmental gradients than are the

Table 1
Common tree species in this study area and their allometric equations for calculating biomass. BAT = total above ground biomass; BST = biomass of stem with bark; BSW = biomass of stem without bark; CIR = stem basal circumference; DBH = diameter at breast height; HT = tree height.

Species	Abbr.	Common name	Equation	Units (biomass, DBH or CIR, height)	Source
Abies concolor	ABCO	White fir	$\ln(BST) = 3.011904 + 2.7727 \times \ln(DBH)$	g, cm, –	Halpern and Means, 2004
Abies magnifica	ABMA	Red fir	$\ln(BST) = 3.020046 + 2.7590 \times \ln(DBH)$	g, cm, –	Halpern and Means, 2004
Juniperus occidentalis	JUOC	Sierra juniper	$\ln(BSW) = -8.5802 + 2.6389 \times \ln(CIR)$	kg, cm, –	Means et al., 1994
Pinus contorta	PICO	Lodgepole pine	$\ln(BST) = -9.10508 + 2.3363 \times \ln(DBH)$	mg, cm, –	Means et al., 1994
Pinus jeffreyi	PIJE	Jeffrey pine	$\ln(BST) = 1.817891 + 2.952 \times \ln(DBH)$	g, cm, –	Halpern and Means, 2004
Pinus lambertiana	PILA	Suger pine	$\ln(BST) = 3.229148 + 2.6863 \times \ln(DBH)$	g, cm, –	Halpern and Means, 2004
Pinus monticola	PIMO	Western white pine	$BAT = 20{,}800 + 0.1544 \times (DBH^2 \times HT)$	g, cm, cm	Halpern and Means, 2004
Populus tremuloides	POTR	Quaking aspen	$\ln(BAT) = -2.6224 + 2.4827 \times \ln(DBH)$	kg, cm –	Jenkins et al., 2004
Tsuga mertensiana	TSME	Mountain hemlock	$\ln(BAT) = -10.1688 + 2.5915 \times \ln(DBH)$	mg, cm, –	Jenkins et al., 2004

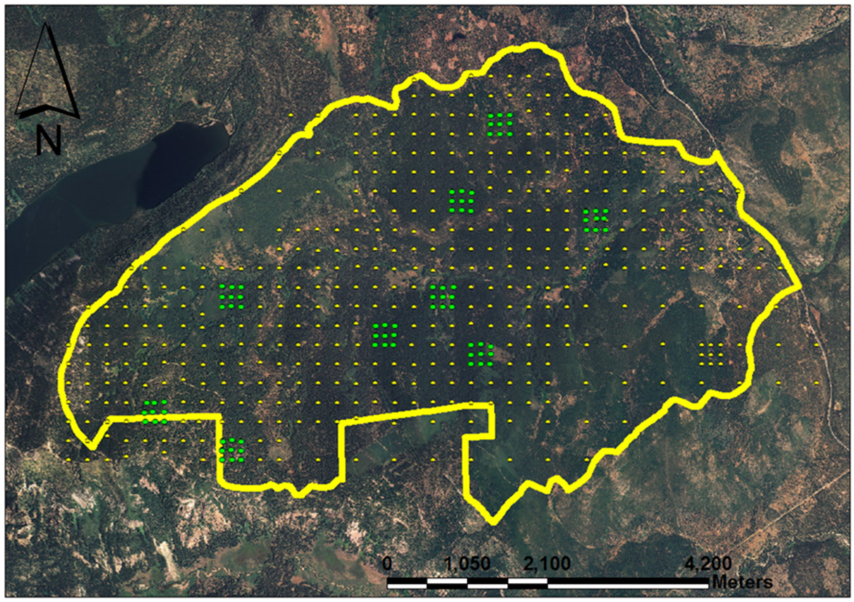

Fig. 2. Field plots of vegetation measurements. The smaller dots indicate the plots located with a recreational GPS. The larger dots indicate the plots located with both a recreational GPS and a differential GPS. The thick line is the boundary of the vegetation type map.

alliances recognized in NVC (Grossman et al., 1998), so in total a much smaller number of SAF types (86 forest types) have been identified for the whole United States.

3. Methods

3.1. Biomass calculation at the plot-level

Biomass can be most accurately calculated using species-specific allometric equations. A comprehensive review of the literature was conducted to search species-specific allometric equations and, during the selection process, preference was given to equations meeting all or most of the following criteria: 1) being derived from a high number (~40–100) of sample trees, 2) from DBH ranges similar to those in our dataset, 3) from geographical sites most similar to our study location, and 4) including all or the most relevant biomass components of a tree. The final equations we selected are from Halpern and Means (2004), Jenkins et al. (2004), and Means et al. (1994) (see Table 1). To derive the biomass at the plot level, we summed the biomass of live trees with DBH > 5 cm (the total biomass of understory trees was multiplied by three given that only a random third of each plot was measured for them) and converted the biomass total to density based on the area of each plot. We only consider live trees because the dead trees usually have few or no leaves and thus generate much fewer laser returns.

3.2. Lidar data processing

The first step of lidar data processing is to filter the raw lidar points and separate them into ground and non-ground returns (Chen et al., 2007). Then, the ground returns identified were interpolated to generate a Digital Elevation Model (DEM) of 1 m cell size. The canopy height of individual points was calculated as the difference between their original Z values and the corresponding DEM cell elevations. Based on the canopy height, the following statistics were calculated for all points within a given field plot: mean (h_u), standard deviation (h_{std}), skewness (h_{skn}), and kurtosis (h_{kurt}); proportion of lidar points within different height bins (0 to 5 m, 5 to 10 m, …, 45 to 50 m, and >50 m, denoted as p_{0to5}, p_{5to10}, …, p_{45to50}, and $p_{>50}$, respectively); percentile heights (5, 10, …, 100 percentile, denoted as h_5, h_{10}, …, h_{100}, respectively; note that 100 percentile height corresponds to maximum height); and quadratic mean height (h_{qm}) (see Table 2). The quadratic mean height was calculated as Lefsky et al. (1999). Two sets of lidar metrics were generated: one is based on all lidar returns and the other is based on first returns since some studies have found that first returns may have better performance in predicting vegetation attributes (e.g., Kim et al., 2009). All of the above lidar data processing was conducted using the Tiffs (Toolbox for Lidar Data Filtering and Forest Studies) software (Chen, 2007).

3.3. Statistical analysis

The mixed-effects model used to predict plot-level biomass from lidar metrics and vegetation types is as follows:

$$\begin{aligned} &\mathbf{Y} = \mathbf{Xb} + \mathbf{Zb} + \mathbf{e} \\ &\mathbf{b} \sim \mathrm{N}(\mathbf{0}, \mathbf{G}) \\ &\mathbf{e} \sim \mathrm{N}(\mathbf{0}, \mathbf{R}) \\ &\mathrm{cov}(\mathbf{b}, \mathbf{e}) = 0 \end{aligned} \tag{1}$$

where $\mathbf{Y}$ is a vector of biomass for n field plots, $\mathbf{X}$ is the $n \times p$ design matrix for the p fixed effects, $\mathbf{Z}$ is the $n \times q$ design matrix for q random effects, $\boldsymbol{\beta}$ is a $p \times 1$ vector for the fixed effects, $\mathbf{b}$ is $q \times 1$ vector for the q random effects, and $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector for the error random effects. Note that 1) the random effect vector $\mathbf{b}$ has Gaussian (Normal) distributions with zero means and variance–covariance matrix $\mathbf{G}$, which is called the G covariance structure; 2) the error vector $\boldsymbol{\varepsilon}$ could be correlated with variance–covariance matrix $\mathbf{R}$, which is modeled with variograms in this study; and 3) the random effects $\mathbf{b}$ and $\boldsymbol{\varepsilon}$ are independent. Given that there usually exist power–law relationships between biomass and other vegetation attributes such as DBH or height (Zianis and Mencuccini, 2004), biomass and all lidar metrics were log-transformed so that the developed models are linear at the log-scale.

We used stepwise regression to select the statistically significant lidar metrics for predicting biomass. Since both the response and predictor variables are at the log-scale, the developed models are *multiplicative* at the original scale. The multiplicative models served as the benchmark and starting point for developing mixed-effects models; in other words, we added and tested random effects only for the lidar metrics selected in the multiplicative models. Modeling variance structure is

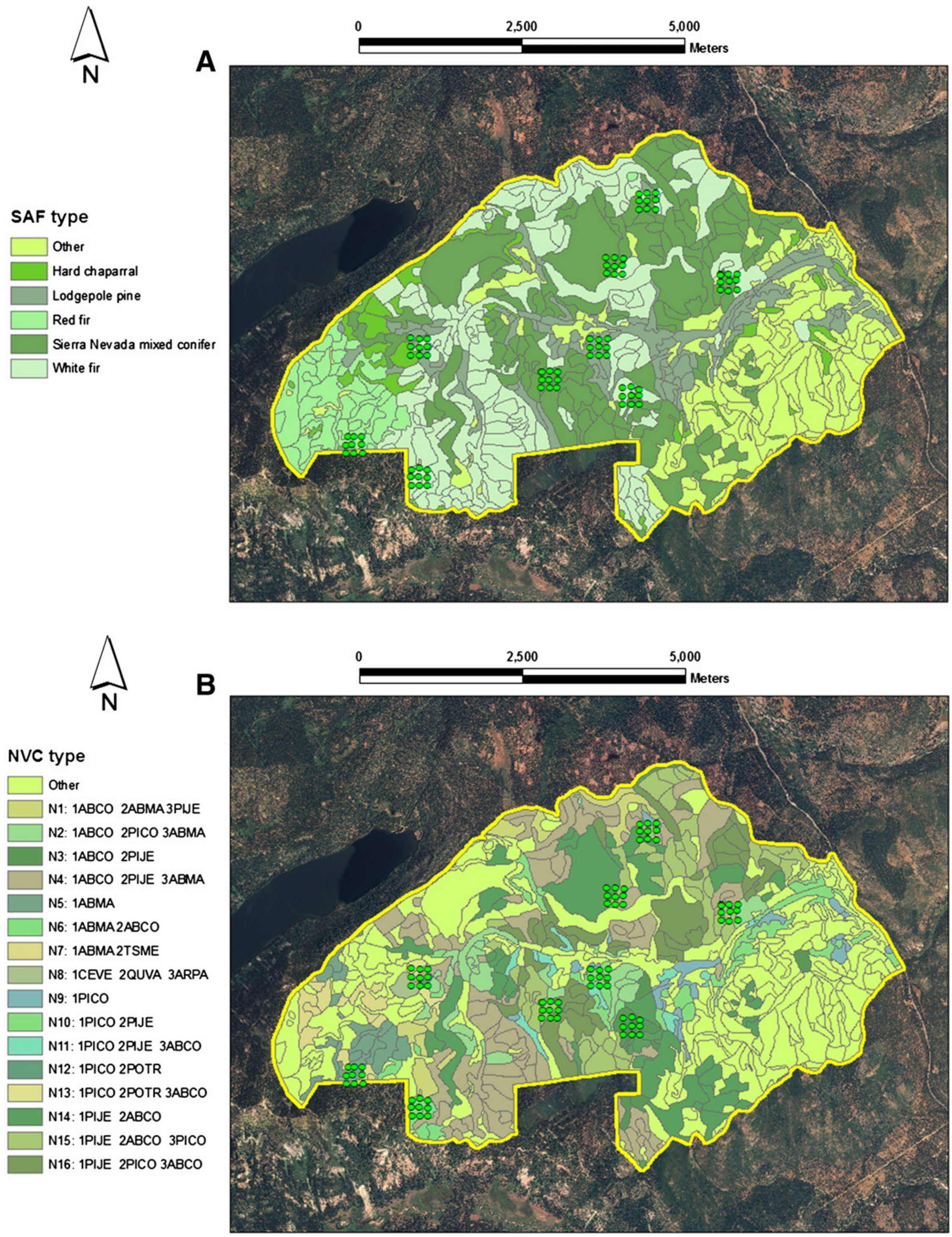

Fig. 3. Vegetation type maps of the study area. (a) SAF type, (b) NVC alliance-level type.

probably the most powerful and critical feature of mixed-effects models, which allows correlation among observations. To find the most parsimonious yet effective **G** covariance structure, we initially fit a model with all predictor variables having random effects and their covariance matrix being unstructured (UN), then fit models with a Variance Components (CV) covariance structure, which means that the individual random effects are independent and the off-diagonal elements of the covariance matrix are zeros. If the estimate of any random effects is statistically insignificant from zero across all different vegetation types, the random effect was dropped from the model. A total of four different types of variogram models (exponential, spherical, Gaussian, and Matern) were tested to model the spatial dependence of the residuals and calculate the variance–covariance matrix **R**. AIC (Akaike Information Criteria) was used to help select the best models, which usually have the lowest AIC. However, if the AIC values of two models have a difference less than 2, such models are considered indistinguishable (Burnham and Anderson, 2002). Once the best models had been selected, leave-one-out cross-validation was used to calculate the model coefficient of

Table 2
Lidar metrics for predicting forest attributes.

Lidar metrics	Description
h_u, h_{std}, h_{skn}, h_{kurt}	Mean, standard deviation, skewness, kurtosis of height of lidar points
p_{0to5}, p_{5to10}, …, p_{45to50}, $p_{>50}$	Proportion of lidar points within height bins (0 to 5 m, 5 to 10 m, …, 45 to 50 m, and > 50 m)
h_5, h_{10}, …, h_{100},	Percentile height of lidar points
h_{qm}	Quadratic mean height of lidar points

determination (R^2) and RMSE so that a straightforward comparison can be made between the results from this study and those from others. We used SAS 9.1.3 (SAS Institute Inc.) to fit mixed-effects models.

Among the 81 DPGS plots, one has questionable GPS accuracy and another three plots are outside of the vegetation type map so they are excluded from our analysis. When stepwise regression was used to estimate biomass for the remaining 77 DGPS plots, it was found that four plots have large residuals (>3 standard deviations). After a careful examination of the tree characteristics of the four plots and inspection of their corresponding point clouds, it was found that there was obvious mismatch of tree information (e.g., tree density, size) between lidar point clouds and field data for three plots. It is suspected that there might be large errors of plot coordinates or vegetation measurements in the field data of these three plots, so they were excluded from our analysis as well. However, the remaining plot was kept since no distinct mismatch can be identified, resulting in a total of 74 DGPS plots in our ground truth data. Table 3 shows the cross-tabulation of the 74 DPGS plots in the NVC alliance-level and SAF vegetation type classification systems. We developed mixed-effects models based on two vegetation types (SAF vs. NVC alliance-level) and two sets of lidar metrics (derived from all returns vs. first returns), which lead to a total of four sets of mixed-effects models for the DGPS plots.

4. Results

When the lidar metrics from all returns were used for the 74 DGPS plots, the two-way stepwise regression (with an enter probability of 0.05 and leave probability of 0.1) selected two lidar metrics h_{qm} and p_{35to40} in the multiplicative model:

$$\ln(\mathrm{AGB}) = 1.571\ln\left(h_{qm}\right) + 0.055\ln(p_{35to40}) + 2.066 \quad (2)$$

where AGB is the aboveground live tree biomass in Mg/ha and h_{qm} is the quadratic mean height in meters, and p_{35to40} is the proportion of lidar points between 35 and 40 m. Starting with the two lidar metrics selected in Eq. (2) and using the SAF vegetation types, we follow the procedure described in Section 3.3 to develop and test mixed-effects models (see DGPS.A.SAF.M1-7 in Table 4 for the models developed). When both h_{qm} and p_{35to40} are modeled as random effects, it was found that the model with the variance components (VC) covariance structure of random effects (model DGPS.A.SAF.M2) produced much smaller AIC compared to the one with the unconstructed (UN) covariance matrix (model DGPS.A.SAF.M1, Table 4), indicating that model DGPS.A.SAF.M2 should be preferred. An examination of model DGPS.A.SAF.M2 revealed that (1) the estimates of the random effects of intercept and the metric p_{35to40} are zeros and (2) the fixed-effects p_{35to40} is not statistically significant. So, the lidar metric p_{35to40} was removed and no random effect for intercept was modeled, resulting in model DGPS.A.SAF.M3. This further reduced the AIC to 23.4 compared to the AIC of 26.3 from model DGPS.A.SAF.M2. Starting with model DGPS.A.SAF.M3, four different variogram models (exponential, spherical, Gaussian, and Matern) were used to model the variance-covariance matrix **R** (models DGPS.A.SAF.M4-7). It was found that these models have higher AICs (models DGPS.A.SAF.M5-7) or very small (=0.2) AIC differences (model DGPS.A.SAF.M4) compared to model DGPS.A.SAF.M3. This indicates that, after incorporating the fixed and random effects in model DPGS.A.SAF.M3, the residuals of AGB have no significant spatial autocorrelation at the scale of current minimal plot spacing (125 m) or larger. As a result, model DGPS.A.SAF.M3 was chosen as the final mixed-effects model in this case of using SAF vegetation type and the lidar metrics from all returns for the 74 DGPS plots.

Similarly, we developed models for the cases of using 1) SAF vegetation types and lidar metrics from first returns (see models DPGS.F.SAF.M1-7 in Table 4), 2) NVC alliance-level vegetation types and lidar metrics from all returns (see models DGPS.A.NVC.M1-7 in Table 5), and 3) NVC alliance-level vegetation types and lidar metrics

Table 3
Cross-tabulation of 74 DGPS plots in two vegetation classification systems: NVC alliance-level type (rows N1–N16) and SAF type (HRC-Hard Chaparral, LPN-Lodgepole Pine, RFR-Red Fir, SMC-Sierra Nevada Mixed Conifer, WFR-White Fir). The name of each NVC alliance-level type lists 1–3 dominant tree species in that type. See Table 1 for the abbreviated species names in each NVC type.

NVC type	SAF type					
	HRC	LPN	RFR	SMC	WFR	Total
N1: 1ABCO 2ABMA 3PIJE	0	0	0	0	1	1
N2: 1ABCO 2PICO 3ABMA	0	0	0	0	4	4
N3: 1ABCO 2PIJE	0	0	0	0	5	5
N4: 1ABCO 2PIJE 3ABMA	0	0	0	0	13	13
N5: 1ABMA	0	0	4	0	0	4
N6: 1ABMA 2ABCO	0	0	0	0	7	7
N7: 1ABMA 2TSME	0	0	1	0	0	1
N8: 1CEVE 2QUVA 3ARPA	2	0	0	0	0	2
N9: 1PICO	0	3	0	0	0	3
N10: 1PICO 2PIJE	0	7	0	0	0	7
N11: 1PICO 2PIJE 3ABCO	0	1	0	0	0	1
N12: 1PICO 2POTR	0	1	0	0	0	1
N13: 1PICO 2POTR 3ABCO	0	3	0	0	0	3
N14: 1PIJE 2ABCO	0	0	0	8	0	8
N15: 1PIJE 2ABCO 3PICO	0	0	0	10	0	10
N16: 1PIJE 2PICO1 3ABCO	0	0	0	4	0	4
Total	2	15	5	22	30	74

Table 4
Different mixed-effects models of biomass estimation based on differential GPS plots (denoted as DGPS in the model no.), lidar metrics derived from all or first returns (A or F in the model no.), and SAF forest cover type. UN means that the G covariance matrix is unconstructed; VC means that the Variance Components matrix is used as the G covariance structure. The best model in each set is bolded.

Model no.	Fixed effects	Random effects	G cov. structure*	Variogram model	AIC
DGPS.A.SAF.M1	Intercept, h_{qm}, p_{35to40}	Intercept, h_{qm}, p_{35to40}	UN	None	49.2
DGPS.A.SAF.M2	Intercept, h_{qm}, p_{35to40}	Intercept, h_{qm}, p_{35to40}	VC	None	26.3
DGPS.A.SAF.M3	**Intercept, h_{qm}**	**h_{qm}**	**VC**	**None**	**23.4**
DGPS.A.SAF.M4	Intercept, h_{qm}	h_{qm}	VC	Exponential	23.2
DGPS.A.SAF.M5	Intercept, h_{qm}	h_{qm}	VC	Spherical	23.8
DGPS.A.SAF.M6	Intercept, h_{qm}	h_{qm}	VC	Gaussian	23.6
DGPS.A.SAF.M7	Intercept, h_{qm},	h_{qm}	VC	Matern	25.2
DGPS.F.SAF.M1	Intercept, h_{qm}, h_{40}, p_{35to40}	Intercept, h_{qm}, h_{40}, p_{35to40}	UN	None	50.7
DGPS.F.SAF.M2	Intercept, h_{qm}, h_{40}, p_{35to40}	Intercept, h_{qm}, h_{40}, p_{35to40}	VC	None	28.9
DGPS.F.SAF.M3	**Intercept, h_{qm}**	**h_{qm}**	**VC**	**None**	**26.0**
DGPS.F.SAF.M4	Intercept, h_{qm}	h_{qm}	VC	Exponential	26.0
DGPS.F.SAF.M5	Intercept, h_{qm}	h_{qm}	VC	Spherical	25.9
DGPS.F.SAF.M6	Intercept, h_{qm}	h_{qm}	VC	Gaussian	25.5
DGPS.F.SAF.M7	Intercept, h_{qm},	h_{qm}	VC	Matern	27.5

Table 5
Different mixed-effects models of biomass estimation based on differential GPS plots (DGPS in the model no.), lidar metrics derived from all or first returns (A or F in the model no.), and NVC alliance-level vegetation type. UN means that the G covariance matrix is unconstructed; VC means that the Variance Components matrix is used as the G covariance structure. The best model in each set is bolded.

Model no.	Fixed effects	Random effects	G cov. structure*	Variogram model	AIC
DGPS.A.NVC.M1	Intercept, h_{qm}, p_{35to40}	Intercept, h_{qm}, p_{35to40}	UN	None	104.0
DGPS.A.NVC.M2	Intercept, h_{qm}, p_{35to40}	Intercept, h_{qm}, p_{35to40}	VC	None	31.8
DGPS.A.NVC.**M3**	**Intercept, h_{qm}**	**h_{qm}**	**VC**	**None**	**25.8**
DGPS.A.NVC.M4	Intercept, h_{qm}	h_{qm}	VC	Exponential	25.2
DGPS.A.NVC.M5	Intercept, h_{qm}	h_{qm}	VC	Spherical	25.9
DGPS.A.NVC.M6	Intercept, h_{qm}	h_{qm}	VC	Gaussian	25.5
DGPS.A.NVC.M7	Intercept, h_{qm},	h_{qm}	VC	Matern	27.7
DGPS.F.NVC.M1	Intercept, h_{qm}, h_{40}, p_{35to40}	Intercept, h_{qm}, h_{40}, p_{35to40}	UN	None	161.4
DGPS.F.NVC.M2	Intercept, h_{qm}, h_{40}, p_{35to40}	Intercept, h_{qm}, h_{40}, p_{35to40}	VC	None	33.7
DGPS.F.NVC.M3	**Intercept, h_{qm}**	**h_{qm}**	**VC**	**None**	**28.8**
DGPS.F.NVC.M4	Intercept, h_{qm}	h_{qm}	VC	Exponential	28.3
DGPS.F.NVC.M5	Intercept, h_{qm}	h_{qm}	VC	Spherical	28.6
DGPS.F.NVC.M6	Intercept, h_{qm}	h_{qm}	VC	Gaussian	28.5
DGPS.F.NVC.M7	Intercept, h_{qm},	h_{qm}	VC	Matern	30.5

from first returns (see models DGPS.F.NVC.M1-7 in Table 5). With the same rationale as above, we selected the best models for these three cases, which are model DPGS.F.SAF.M3, DGPS.A.NVC.M3, and DGPS.F.NVC.M3, respectively. Note that the multiplicative models based on first returns include an additional lidar metric, h_{40}. However, in the corresponding mixed-effects models, this metric is not statistically significant any more. As a result, all best mixed-effects models have the same model structure in terms of the fixed effects, random effects, **G** covariance structure, and **R** covariance matrix (i.e., variogram models).

Table 6 summarizes the fitting statistics of multiplicative and the best mixed-effects models developed based on the 74 DGPS plots. The examination of fitting statistics of the mixed-effects models indicates that all mixed-effects models, based on either NVC alliance-level or SAF vegetation types, outperformed the corresponding multiplicative models. For example, when all returns were used, the R^2 increased from 0.77 to 0.83 for NVC alliance-level types and to 0.82 for SAF types. The RMSE decreased by about 10% for all returns and by about 5% for first returns. Among all models based on the DGPS plots, the mixed-effects model based on all returns and NVC alliance-level vegetation type (model DGPS.A.NVC.M3) has the highest R^2 (0.83) and the lowest RMSE (72.2 Mg/ha). However, since the model based on all returns and SAF vegetation type (model DGPS.A.SAF.M3) has almost the same R^2 (0.82) and RMSE (72.8 Mg/ha) as model DGPS.A.NVC.M3 while using a smaller number of vegetation classes (5 SAF classes instead of 16 NVC classes), it was considered as the best model from the aspects of both model parsimony and fitting statistics. Fig. 4 shows the biomass map of the study area based on model DGPS.A.SAF.M3.

Table 6
Model fitting statistics calculated with leave-one-out cross validation for multiplicative and mixed-effects models.

	Multiplicative model		Mixed effects model			
			NVC alliance-level type		SAF type	
	R^2	RMSE (Mg/ha)	R^2	RMSE (Mg/ha)	R^2	RMSE (Mg/ha)
DGPS plots (n = 74)						
All returns	0.77	80.8	0.83	72.2	0.82	72.8
First returns	0.77	80.2	0.81	74.5	0.81	75.1
RGPS plots (n = 74)						
All returns	0.66	98.7	0.70	94.0	0.72	92.5
First returns	0.67	97.4	0.70	95.2	0.68	98.2

5. Discussion

5.1. Comparison with previous studies

Our results indicate that the mixed-effects models have better performance than the corresponding fixed-effects models. This finding is consistent with previous studies that used other remotely-sensed data: Meng et al. (2007) used NDVI derived from Landsat ETM + imagery and forest inventory data to develop a linear fixed-effects model and linear mixed-effects models to estimate merchantable biomass for the state of Georgia. They found that the linear mixed-effects model with random effects in both intercept and slope best fits the data and achieved a R^2 of 0.57 while the fixed-effects model produced a R^2 of 0.31 only.

Some previous studies found that the integration of lidar data and optical or radar imagery does not necessarily produce better results in biomass modeling. For example, Hyde et al. (2006) found that the addition of Quickbird and SAR/InSAR structure metrics (such as NDVI and backscatter intensity) to LVIS (Laser Vegetation Imaging Sensor) resulted in no improvement for estimating biomass across 120 one-hectare circular plots in the Sierra Nevada of California. This was explained by the fact that the structure metrics from lidar, radar, and Quickbird are redundant (Hyde et al., 2006). Using different inputs (categorical vegetation types instead of continuous structure metrics such as NDVI) and statistical approaches (mixed-effects instead of fixed-effects models), we found that it is possible to improve biomass estimation by integrating lidar and optical remote sensing data. The difference between this study and Hyde et al. (2006) might be attributed to our different modeling strategy (i.e., considering the biomass dependence on vegetation types) and our use of mixed-effects models and vegetation types, but more research is needed to further investigate this issue.

The two multiplicative models have the same R^2 (0.77) and similar RMSE (80.8 Mg/ha for all returns; 80.2 Mg/ha for first returns). These fitting statistics are comparable to those from Gonzalez et al. (2010), which used the lidar data collected by the same lidar system (Optech ALTM 2050). They used field measurements of 39 plots collected by a modified FIA (Forestry Inventory and Analysis) design to develop stepwise regression model to estimate aboveground live tree biomass in North Yuba in the Tahoe National Forest in California, a site only ~50 km away from our study area. Their final model included lidar metrics such as quadratic mean height and five percentile heights (p10, p20, p30, p40, and p50), with R^2 of 0.80, slighter higher than our fixed-effects model. However, their model RMSE is 123 Mg/ha, much larger than 80.8 Mg/ha of our fixed-effects model (Eq. 2). The causes of the large RMSE difference between this study and Gonzalez et al. (2010) are multifaceted: besides using fixed- instead mixed-effects models, their study uses field plots consisting of four subplots of 17.95 m radius while we use single plots of 12.62 m radius: the larger field plot introduces more variability of canopy structure, making it more difficult to characterize using a single set of metrics; another reason for the larger RMSE in Gonzalez et al. (2010) is that they incorporated uncertainty in their field biomass using Monte Carlo simulations.

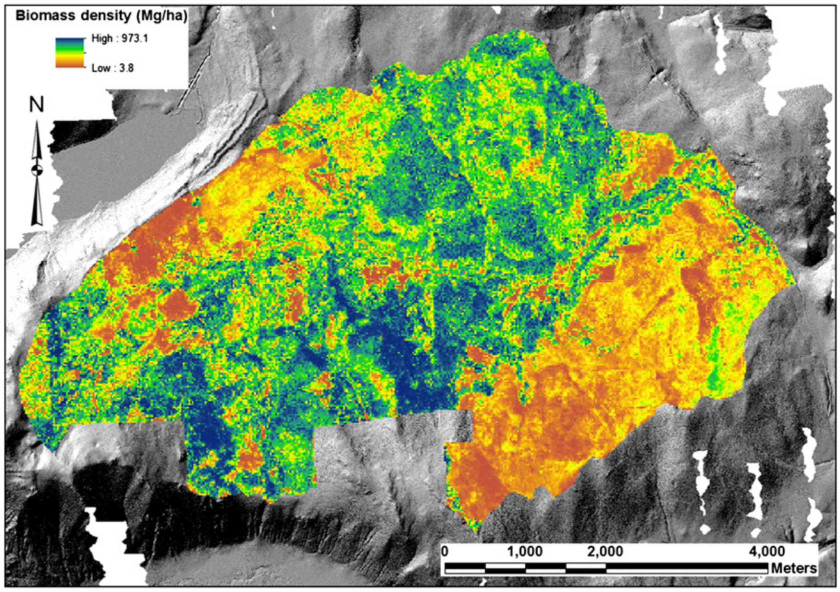

Fig. 4. Biomass map of the study area (based on model DGPS.A.SAF.M3).

5.2. Advantages of mixed-effects models

When vegetation types or other information are available to partition a study area into different strata, an alternative approach is to fit a statistical model for each stratum. Compared to the approach of fitting individual stratum-specific regression models, mixed-effects models have the advantage of using fewer parameters while possibly achieving comparable or even better performance. For instance, Meng et al. (2007) compared a mixed-effects model with the approach of fitting an individual regression model within each region (IRR). They divided the state of Georgia in USA into five eco-regions, each including 10 to 67 counties. The mixed-effects models were developed at the county-level, so the minimal sample size of each stratum (eco-region) is 10. They found that the mixed-effects model obtained slightly better performance than the IRR approach even though the mixed-effects model used much fewer parameters.

In our study, the need to use mixed-effects models instead of fitting individual vegetation type specific statistical models is obvious because we have very limited numbers of plots per vegetation type (see Table 3). For example, among the 16 NVC alliance-level vegetation types, 11 types have 5 or less field plots associated with each. Using the SAF classification system leads to fewer types and thus higher average number of plots per vegetation type. However, there are only 2 plots for the Hard Chaparral (HRC) and 5 plots for the Red Fir (RFR) types. Fitting statistical models for such small samples is clearly questionable from the statistical standpoint (Green, 1991).

Mixed-effects models deal with the biomass dependence on vegetation types from a different perspective: the coefficients of the biomass models for different vegetation types could be assumed to vary as random Gaussian variables. This assumption puts a constraint on the variability of model coefficients and prevents unreliable estimates of model coefficients from being produced even when the sample size is small. Take model DGPS.A.SAF.M3 as an example (see Table 4), which is essentially a random slope model with the coefficient of h_{qm} (at the log scale) varying among different vegetation types:

$$\ln AGB_{ij} = 1.6971\ln\left(h_{qm,ij}\right) + b_i{}^{*}\ln\left(h_{qm,ij}\right) + 1.3860 \quad (3)$$

where AGB_{ij} is the aboveground live tree biomass for plot j of vegetation type i; $h_{qm,ij}$ is the quadratic mean height of all lidar points for the plot j of vegetation type i. b_i is the random coefficient estimated with the empirical best linear unbiased predictions (EBLUPs) for vegetation type i and it represents the estimated deviation from the mean slope (i.e., 1.6971). Fig. 5 shows the estimated biomass models for different vegetation types. These regression lines could be much different from the ones derived from vegetation type specific regression models. For instance, a regular "least squares" regression model for the HRC vegetation type will create a line that passes through the two HRC plots; such a line will be highly sensitive to the small sample size problem and thus will have less generalization ability for prediction. Instead, in the mixed-effects models, the regression line of a given vegetation type is a combination of a) the coefficients of the fixed effects (1.3860 for intercept, and 1.6971 for slope for this example as shown in Eq. 3), and b) the estimated coefficients of the random effects (b_i in Eq. 3). The mean regression line (determined by the coefficients of the fixed effects) can be thought as an initial estimate for the regression model of a specific vegetation type, much like the *prior* estimate in Bayesian statistics. This is the essential reason why mixed-effects models could be less susceptible to the small sample size issue. Additionally, if more samples are available for a given vegetation type, mixed-effects modeling will take advantage of the available sample data and the estimated model will be closer to the one derived from the regular least square regression, exemplified by the models for the WFR, SMC, and LPN vegetation types shown in Fig. 5. This explains why mixed-effects models are effective in modeling biomass when vegetation types of a wide range of sample sizes exist.

5.3. NVC alliance-level versus SAF vegetation types

One of the interesting results from this study is the lack of differences between the mixed-effects models developed from the two vegetation types (NVC alliance-level vs. SAF). The differences are less than 0.01 for R^2 values and less than 0.6 Mg/ha for RMSE. As introduced in Section 2.4, NVC alliance-level types define a vegetation class based on all vascular plants present while SAF types are defined by the dominant species. Thus, NVC alliance-level types represent a finer scale of vegetation classification. However, from the perspective of biomass estimation, the more coarsely scaled SAF types perform nearly as well because the dominant species account for the vast majority of AGB. These results suggest that we can focus classification and mapping schemes on the dominant species for mixed-effects modeling of biomass.

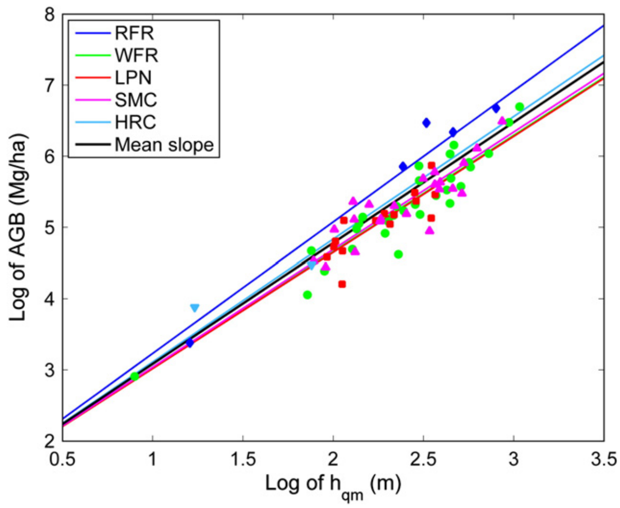

Fig. 5. Mixed-effects model of biomass estimation kmodel DGPS.A.SAF.M3). RFR-Red Fir, WFR-White Fir, LPN-Lodgepole Pine, SMC-Sierra Nevada Mixed Conifer, HRC-Hard Chaparral.

5.4. Lidar metrics from all returns vs. first returns

A few studies have also used lidar metrics derived from first returns only for predicting biomass. For example, Hall et al. (2005) found that a canopy cover metric derived from first returns was able to predict the foliage biomass and total aboveground biomass in a Ponderosa pine forest in Colorado with R^2 of 0.79 and 0.74, respectively. Kim et al. (2009) found that the model using first returns improved the R^2 by 0.1 for predicting the total aboveground biomass in a mixed coniferous forest in Arizona compared to the one using all returns.

Our results indicate that using first returns reduced RMSE by 0.6 Mg/ha compared to using all returns, which are in line with the findings from Kim et al. (2009). However, the improvement is too small to be considered statistically significant. It is interesting that, among the mixed-effects models, the ones based on all returns outperformed the ones based on first returns. The R^2 increased from 0.81 to 0.83 for NVC alliance-level type and, correspondingly, the RMSE decreased by 3%. The reasons for this contrasting pattern and the discrepancy between this study and previous studies are unclear, but it might be related to specific forest conditions and the way with which the specific lidar system generates individual returns (Wagner et al., 2007).

5.5. Biomass estimation using field data located with a recreational GPS

The accuracy of remote sensing based biomass maps is influenced not only by the specific earth observation data and the statistical approaches used, but also by the accuracy of the calibration data used for deriving models and estimates. One of the common problems in field data is the low geo-location accuracy caused by either the use of low-cost GPS or the existence of dense forests. Dominy and Duncan (2001) reported the difficulty of quality satellite reception beneath a dense forest canopy, with the degree of spatial error seriously affecting fine-scale vegetation mapping. Miura and Jones (2010) used a Garmin eTrex GPS (average ± 5.5 m horizontal error) to locate the centers of 25-m radius circular plots for field measurements and related to airborne lidar data. They had to manually shift the plots to achieve a better registration between lidar data and field measurements. However, few studies have evaluated the impacts of GPS accuracy on biomass estimation using lidar data.

The availability of both differential GPS coordinates and recreational GPS coordinates for the 74 plots in our study site made it possible to directly assess the impacts of plot coordinate accuracy on biomass estimation. Table 6 reported the fitting statistics of the multiplicative and mixed-effects models based on RGPS plot coordinates. The use of recreational instead of differential GPS in our study site resulted in a decrease of R^2 by 0.10–0.13 and an increase of RMSE by about 21–31%. This degradation in performance due to GPS accuracy will likely vary depending on the site-specific conditions (e.g., canopy structure, spatial heterogeneity, and topography). Nevertheless, our results emphasize the value of differential GPS to locate field plots for vegetation measurements.

6. Conclusions

Lidar is a state-of-the-art technology for mapping biomass, which relies on the fundamental relationship between biomass and canopy structure metrics such as height. Motivated by the biomass dependence on vegetation types, this study uses an innovative method, mixed-effects models, to integrate airborne lidar and vegetation types derived from aerial photographs to map biomass over the Sagehen Creek Experimental Forest in the Sierra Nevada of California. It was found that mixed-effects models can effectively deal with the small samples associated with each vegetation type and can improve biomass estimation compared to the use of lidar data alone in multiplicative models.

The vegetation of our study site was classified based on two different systems: SAF and NVC alliance-level classes. We found that, despite its emphasis on dominant species, the SAF cover types are as powerful as the NVC alliance-level vegetation types in the mixed-effects modeling of biomass. This result suggests that vegetation classification for carbon assessment could focus on dominant species given the strong relationship between forest stand biomass and dominant species. The vegetation types of this study were visually interpreted from aerial photographs. For many places, especially those in developing countries, updated aerial photographs are not always available. Due to the increasing accessibility of high spatial resolution satellite imagery such as Worldview-2, further research should be done in the future to investigate the use of high spatial resolution satellite imagery, digital image classification, and airborne lidar data for biomass and carbon mapping.

Acknowledgments

Gaia Vaglio Laurin acknowledges the ERC Africa GHG Grant for providing support to this research. We appreciate the constructive comments from the two anonymous reviewers.

References

Asner, G. P., Hughes, R. F., Varga, T. A., Knapp, D. E., & Kennedy-Bowdoin, T. (2009). Environmental and biotic controls over aboveground biomass throughout a tropical rain forest. *Ecosystems*, *12*, 261–278.

Avery, T. E. (1978). *Forester's guide to aerial photo interpretation*. Agriculture Handbook 308. : U.S. Department of Agriculture, Forest Service.

Banskota, A., Wynne, R. H., Johnson, P., & Emessiene, B. (2011). Synergistic use of very high-frequency radar and discrete-return lidar for estimating biomass in temperate hardwood and mixed forests. *Annals of Forest Science*, *68*(2), 347–356.

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., & Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sensing of Environment*, *112*, 3876–3890.

Existing vegetation classification and mapping technical guide. Brohman, R., & Bryant, L. (Eds.). (2005). *Gen. Tech. Rep. WO–67*. Washington DC: U.S. Department of Agriculture Forest Service, Ecosystem Management Coordination Staff.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd edition). New York: Springer-Verlag Press.

Chave, J., Muller-Landau, H. C., Baker, T. R., Easdale, T. A., Ter Steege, H., & Webb, C. O. (2006). Regional and phylogenetic variation of wood density across 2456 neotropical tree species. *Ecological Applications*, *16*, 2356–2367.

Chen, Q. (2007). Airborne lidar data processing and information extraction. *Photogrammetric Engineering and Remote Sensing*, *73*(2), 109–112.

Chen, Q. (2010). Retrieving canopy height of forests and woodlands over mountainous areas in the Pacific coast region using satellite laser altimetry. *Remote Sensing of Environment*, *114*, 1610–1627.

Chen, Q., Gong, P., Baldocchi, D. D., & Xie, G. (2007). Filtering airborne laser scanning data with morphological methods. *Photogrammetric Engineering and Remote Sensing*, *73* (2), 175–185.

Colwell, R. N. (1946). The estimation of ground conditions from aerial photographic interpretation of vegetation types. *Photogmmmetric Engineering*, *12*(2), 151–161.

Davies, K. W., Petersen, S. L., Johnson, D. D., Davis, D. B., Madsen, M. D., & Zvirzdin, D. L. (2010). Estimating juniper cover from National Agriculture Imagery Program (NAIP) imagery and evaluating relationships between potential cover and environmental variables. *Rangeland Ecology & Management*, *63*(6), 630–637.

De Jong, B., Anaya, C., Masera, O., Olguin, M., Paz, F., Etchevers, J., Martinez, R. D., Guerrero, G., & Balbontin, C. (2010). Greenhouse gas emissions between 1993 and 2002 from land-use change and forestry in Mexico. *Forest Ecology and Management*, *260*(10), 1689–1701.

De Santis, A., Asner, G. P., Vaughan, P. J., & Knapp, D. E. (2010). Mapping burn severity and burning efficiency in California using simulation models and Landsat imagery. *Remote Sensing of Environment*, *114*(7), 1535–1545.

Dominy, N. J., & Duncan, B. (2001). GPS and GIS methods in an African rain forest: Applications to tropical ecology and conservation. *Conservation Ecology*, *5*(2), 537–549.

Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B., & Hofton, M. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: Factors affecting the generality of relationships. *Global Ecology & Biogeography*, *12*, 147–159.

Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., & Chazdon, R. L. (2010). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research*, *115*, G00E09, doi:10.1029/2009JG000933.

Eyre, F. H. (1980). Forest cover types of the United States and Canada. Washington, D.C: Society of American Foresters (SAF) 148 p.

Fensham, R. J., & Fairfax, R. J. (2002). Aerial photography for assessing vegetation change: A review of applications and the relevance of findings for Australian vegetation history. *Australian Journal of Botany*, *50*(4), 415–429.

FGDC (Federal Geographic Data Committee) (2008). *National Vegetation Classification Standard*, Version 2 FGDC-STD-005-2008 (version 2). *Vegetation Subcommittee, Federal Geographic Data Committee, FGDC Secretariat*. Reston, Virginia, USA: U.S. Geological Survey.

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 Sensing of Environment*, *114*(7), 1561–1575.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, *26*(3), 499–510.

Grossman, D. H., Faber-Langendoen, D., Weakley, A. S., Anderson, M., Bourgeron, P., Crawford, R., Goodin, K., Landaal, S., Metzler, K., Patterson, K. D., Pyne, M., Reid, M., & Sneddon, L. (1998). International classification of ecological communities: Terrestrial vegetation of the United States. Volume I. *The National Vegetation Classification System: development, status, and applications*. Arlington, Virginia, USA: The Nature Conservancy.

Guo, Z., Hong, C., & Sun, G. (2010). Estimating forest aboveground biomass using HJ-1 Satellite CCD and ICESat GLAS waveform data. *Science in China Series D: Earth Sciences*, *53*(Supplement:1), 16–25.

Hall, S. A., Burke, I. C., Box, D. O., Kaufmann, M. R., & Stoker, J. M. (2005). Estimating stand structure using discrete-return lidar: An example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management*, *208*, 189–209.

Halpern, C., & Means, J. (2004). *Pacific Northwest Plant Biomass Component Equation Library*. Long-Term Ecological Research. Corvallis, OR: Forest Science Data Bank.

Higinbotham, C. B., Alber, M., & Chalmers, A. G. (2004). Analysis of tidal marsh vegetation patterns in two Georgia estuaries using aerial photography and GIS. *Estuaries*, *27*(4), 670–683.

Hyde, P., Dubayah, R., Walker, W., Blair, B., HOfton, M., & Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, *102*, 63–73.

Jenkins, J. C., Chojnacky, D. C., Heath, L. S., & Birdsey, R. A. (2004). Comprehensive database of diameter-based biomass regressions for North American tree species. *General technical report NE-319*. Newtown Square, PA: U.S. Department of Agriculture, Forest Service.

Jennings, M. D., Faber-Langendoen, D., Loucks, O. L., Peet, R. K., & Roberts, D. (2009). Standards for associations and alliances of the U.S. National Vegetation Classification. *Ecological Monographs*, *79*, 173–199.

Kim, Y., Yang, Z. Q., Cohen, W. B., Pflugmacher, D., Lauver, C. L., & Vankat, J. L. (2009). Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data. *Remote Sensing of Environment*, *113*, 2499–2510.

Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G., & Shugart, H. H. (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of Environment*, *67*, 83–98.

Lillesand, T., Kiefer, R., & Chipman, J. (2008). *Remote sensing and image interpretation* (6th edition). NY: John Wiley & Sons.

Lim, K., Treitz, P., Baldwin, K., Morrison, I., & Green, J. (2003). Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing*, *29*, 658–678.

Loudermilk, E. L., Hiers, J. K., O'Brien, J. J., Mitchell, R. J., Singhania, A., Fernandez, J. C., Cropper, W. P., & Slatton, K. C. (2009). Ground-based LIDAR: A novel approach to quantify fine-scale fuelbed characteristics. *International Journal of Wildland Fire*, *18*, 676–685.

MacLean, G. A., & Krabill, W. B. (1986). Gross-merchantable timber volume estimation using an airborne LiDAR system. *Canadian Journal of Remote Sensing*, *12*, 7–18.

Malmsheimer, R. W., Bowyer, J. L., Fried, J. S., Gee, E., Izlar, R. L., Miner, R. A., Munn, I. A., Oneil, E., & Stewart, W. C. (2011). Managing forests because carbon matters: Integrating energy, products, and land management policy. *Journal of Forestry*, *109*, S7–S48.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., Harmon, M. E., & McKee, W. A. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon. *Remote Sensing of Environment*, *67*, 298–308.

Means, J. E., Hansen, H. A., Koerper, G. J., Alaback, P. B., & Klopsch, M. W. (1994). Software for computing plant biomass — BIOPAK users guide. *Gen. Tech. Rep. PNW-GTR-340*. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Meng, Q., Cieszewski, C. J., Madden, M., & Borders, B. (2007). A linear mixed effects model of biomass and volume of trees using Landsat ETM+ images. *Forest Ecology and Management*, *244*(1–3), 93–101.

Meyer, K. E., & Laudenslayer, W. F. (1988). *A guide to wildlife habitats of California*. Sacramento: California Department of Fish and Game.

Miura, N., & Jones, S. D. (2010). Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. *Remote Sensing of Environment*, *114*, 1069–1076.

Morgan, J. L., Gergel, S. E., & Coops, N. C. (2010). Aerial photography: A rapidly evolving tool for ecological management. *BioScience*, *60*(1), 47–59.

Morris, D. L., Western, D., & Maitumo, D. (2009). Pastoralist's livestock and settlements influence game bird diversity and abundance in a savanna ecosystem of southern Kenya. *African Journal of Ecology*, *47*(1), 48–55.

Næsset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, *112*, 3079–3090.

Nelson, R., Boudreau, J., Gregoire, T. G., Margolis, H., Næsset, E., Gobakken, T., & Stahl, G. (2009). Estimating Quebec provincial forest resources using ICESat/GLAS. *Canadian Journal of Forest Research*, *39*, 862–881.

Nelson, R., Krabill, W., & Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment*, *24*, 247–267.

Niklas, K. J. (1995). Size-dependent allometry of tree height, diameter and trunk-taper. *Annals of Botany*, *75*, 217–227.

Ni-Meister, W., Lee, S. Y., Strahler, A. H., Woodcock, C. E., Schaaf, C., Yao, T. A., Ranson, K. J., Sun, G. Q., & Blair, J. B. (2010). Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *Journal of Geophysical Research-Biogeosciences*, 115, doi:10.1029/2009JG000936 Article No.: G00E11.

Salas, C., Ene, L., Gregoire, T. G., Næsset, E., & Gobakken, T. (2010). Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote Sensing of Environment*, *114*, 1277–1285.

Stahl, G., Holm, S., Gregoire, T. G., Gobakken, T., Næsset, E., & Nelson, R. (2011). Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Canadian Journal of Forest Research*, *41*(1), 96–107.

Ursino, N. (2007). Modeling banded vegetation patterns in semiarid regions: Interdependence between biomass growth rate and relevant hydrological processes. *Water Resources Research*, *43*(4), W04412.

USDA Forest Service (1981). CALVEG: A classification of California vegetation. San Francisco CA: Pacific Southwest Region, Regional Ecology Group 168 pp.

Wagner, W., Roncat, A., Melzer, T., & Ullrich, A. (2007). Waveform analysis techniques in airborne laser scanning. *IAPRS*, *39*(Part 3/W52), 413–418.

Waring, R. H., Way, J. B., Hunt, E. R., Morrissey, L., Ranson, K. J., Weishampel, J. F., Oren, R., & Franklin, S. E. (1995). Imaging radar for ecosystem studies. *Bioscience*, *45*, 715–723.

Zianis, D., & Mencuccini, M. (2004). On simplifying allometric analyses of forest biomass. *Forest Ecology and Management*, *187*(2–3), 311–332.