When Big Data are Too Much: Effects of LiDAR Returns and Point Density on Estimation of Forest Biomass

Kunwar K. Singh, Gang Chen, John B. Vogler, and Ross K. Meentemeyer

***Abstract*—Analysis of light detection and ranging (LiDAR) data is becoming a mainstream approach to mapping forest biomass and carbon stocks across heterogeneous landscapes. However, large volumes of multireturn high point-density LiDAR data continue to pose challenges for large-area assessments. We are beginning to learn when and where point density can be reduced (or aggregated), but little is known regarding the degree to which multireturn data—at varying levels of point density—improve estimates of forest biomass. In this study, we examined the combined effects of LiDAR returns and data reduction on field-measured estimates of aboveground forest biomass in deciduous and mixed evergreen forests in an urbanized region of North Carolina, USA. We extracted structural metrics using first returns only, all returns, and rarely used laser pulse first returns from reduced point densities of LiDAR data. We statistically analyzed relationships between the field-measured biomass and LiDAR-derived variables for each return type and point-density combination. Overall, models using first return data performed only slightly better than models that utilized multiple returns. First-return models and multiple-return models at one percent point density resulted in 14% and 11% decrease in the amount of explained variation, respectively, compared to models with 100% point density. In addition, variance of modeled biomass across all point densities and return models was statistically similar to the field-measured biomass. Taken together, these results suggest that LiDAR first returns at reduced point density provide sufficient data for mapping urban forest biomass and may be an effective alternative to multireturn data.**

***Index Terms*—Aboveground biomass, data reduction, large-area assessments, light detection and ranging (LiDAR), multiple linear regression (MLR), point density and returns, urban forest.**

Manuscript received July 30, 2015; revised January 12, 2016; accepted January 26, 2016. Date of publication March 01, 2016; date of current version August 12, 2016. This work was supported in part by the National Science Foundation ULTRA-Ex program (BCS-0949170), in part by the Garden Club of America Zone VI Fellowship in Urban Forestry, and in part by the American Association of Geographers Dissertation Research Grants.

K. K. Singh is with the Center for Geospatial Analytics, North Carolina State University, Raleigh, NC 27695 USA, and also with the Northern Plant Ecology Laboratory, Department of Biology, University of Saskatchewan, Saskatoon, SK S7N 5E2, Canada (e-mail: kunwar.singh@usask.ca).

G. Chen is with the Laboratory for Remote Sensing and Environmental Change, University of North Carolina at Charlotte, Charlotte, NC 28233 USA (e-mail: gang.chen@uncc.edu).

J. B. Vogler is with the Center for Geospatial Analytics, North Carolina State University, Raleigh, NC 27695 USA (e-mail: jbvogler@ncsu.edu).

R. K. Meentemeyer is with the Center for Geospatial Analytics, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC 27695 USA (e-mail: rkmeente@ncsu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2016.2522960

I. Introduction

ACCURATE assessment of large-area aboveground forest biomass is essential for developing effective carbon management strategies and understanding the impact of climate change on carbon stocks at regional to global scales [1]. Light detection and ranging (LiDAR) remote sensing is being increasingly applied for such assessment with a high degree of accuracy [3], [4]. Compared to other sensors, LiDAR has the ability to capture three-dimensional (3-D) forest structure, making it a key data source for biomass mapping [15]. With recent advancements in sensor technology, the point density and number of returns per pulse generated by discrete-return LiDAR systems have grown exponentially. While LiDAR facilitates very accurate representations of forest structure from the stand to the individual tree level, data procurement costs remain high, and processing voluminous LiDAR data for large-area assessments of forest biomass remains a challenge.

To overcome these challenges, large-area biomass estimation is typically based on two LiDAR data reduction strategies. First, LiDAR data are often collected for a portion of the study area, and then derived biomass estimates are extrapolated to the entire site using sampling theories and/or relatively low-cost data sets, such as satellite optical imagery [2], [3]. In such cases, plot-level field-measured biomass is regressed against the statistics derived from LiDAR data [14]. Although plot-level biomass estimates are typically accurate, the bias in sample selection and the models used for plot-to-landscape generalization can introduce large errors [31]. Particularly, urban environments with high degrees of spatial heterogeneity and diversity are problematic. Second, researchers have evaluated reducing point density in LiDAR data to help lower procurement costs and computational overhead while retaining spatial variation in forest structure over large areas [12], [26]. However, identifying a cut-off point for data resampling remains a challenge as point density varies from tile to tile and requires additional thorough analysis.

Another rarely discussed data reduction strategy uses a smaller number of returns per emitted laser pulse. Discrete-return LiDAR sensors are capable of recording up to five returns per pulse [24]. Each return represents an individual point in the LiDAR data with an associated return number and unique id. The first and last returns are normally associated with the highest (e.g., canopy top) and lowest (e.g., ground surface) landscape features, respectively, while intermediate returns define the middle structure of vegetation. Ideally,

multitiered forest structures, like those typically found in urban forests, may have many intermediate returns as compared to an even-aged and well-managed forest stand [29]. Therefore, multiple return LiDAR data have become a natural choice in many studies for achieving high accuracy biomass estimates [5], [17], [22], [25], [29].

Theoretically, as few as one LiDAR point representing the tree height and a few points representing the shape and diameter of a tree crown, all from the first return should be sufficient to retain model performance in biomass estimation. This warrants questioning the advantages of multiple LiDAR return data, especially the intermediate returns, for estimating forest biomass at large extents. Studies [4], [10], [14] have found that metrics derived from the LiDAR first returns adequately estimate or improve biomass compared to estimates based on all returns. However, these studies were not only limited to leaf-on LiDAR data and natural forested landscapes of small extents but also rarely addressed the "first return of each laser pulse" (hereafter referred to as "first-in-pulse" return) for estimating urban forest biomass. First-in-pulse returns are those returns labeled as first return for a pulse but not included in the tile. This happens when LiDAR data are clipped into tiles to overcome the complexities of a single LiDAR file with the large data volume [19]. While reducing either LiDAR point density or the number of returns has been evaluated, we still lack a basic understanding about how the accuracy of biomass estimation is affected by jointly tuning these two LiDAR variables. Is it possible that their cumulative impact is greater than the impact of changing one variable only?

In this study, we evaluated the effects of the number of LiDAR returns and LiDAR point-density reduction on the estimation of aboveground biomass of highly fragmented and spatially heterogeneous urban forests of Mecklenburg County, North Carolina, USA. We extracted structural metrics using all returns, first returns, and first-in-pulse from original LiDAR data and reduced point densities (80%, 60%, 40%, 20%, 10%, 5%, and 1% samples) of the original LiDAR data. We used multiple linear regression (MLR) to establish statistical relationships between field-measured biomass and LiDAR-derived predictor variables (PVs) for each return and point-density combination. We compared the performance of biomass models developed for each return category across the range of reduced point densities. Finally, we compared the modeled biomass estimates with field-measured biomass.

II. Materials and Methods

A. Study Area

This study focuses on the urban forests of Mecklenburg County located in the center of the Charlotte Metropolitan Region of North Carolina (Fig. 1). Forested landscapes in the area are primarily oak-hickory-pine forests that have developed on former timber plantation sites and through natural regeneration on abandoned farmland. The southern Piedmont physiographic region is characterized by gently rolling topography. Continued urban growth and the inaccessibility of the high proportion of privately owned land in the region [20] make it even more meaningful to study the utility of LiDAR data for efficient assessment of urban forest biomass.

B. Field Sites and Plant Measurements

We collected field data at 70 forested sites along urban-rural gradients of the study area from 2010 to 2012 as part of the Charlotte urban long-term research areas exploratory (ULTRA-Ex) study. At each site, we established three to five randomly located, 11.5-m fixed-radius field plots by defining a center point and recording coordinates using a Garmin GPSMAP 62s device (± 12 feet accuracy). Within each 415 m^2 plot, we measured the diameter of all plants greater than 5 cm at breast height, and their geographic coordinates, merchantable height, species' name and type (deciduous vs. coniferous), and predominant land-cover type [26].

C. Field-Measured Biomass Estimation

We used [13] allometric equation (1) to estimate biomass at plot level using field-based *dbh* and species group parameters (Table I). First, we grouped field-observed tree species into hardwood and softwood species. Second, we applied species group parameters with field-based *dbh* to estimate individual tree biomass. Finally, the calculated biomass of each tree was aggregated at plot level followed by conversion to a tons per hectare unit

$$bm = Exp\left(\beta_0 + \beta_1 * \ln\left(dbh\right)\right) \tag{1}$$

where *bm* is total aboveground biomass (kg dry weight), *Exp* is the exponential function, *dbh* is the diameter at breast height (cm), ln is the natural log, and β_0 and β_1 are parameters for hardwood and softwood tree species groups.

D. LiDAR Data Processing and Extraction of Return-Based Metrics

We obtained leaf-off multiple return airborne LiDAR data from the GIS mapping and project services of Charlotte-Mecklenburg County government in the state plane coordinate system (NC FIPS 3200, NAD 1983, meters). A total of 1896 tiles cover the study area with each tile having dimensions 914.40 m $\times$ 914.40 m. Pictometry International Corp. (Rochester, New York, USA) acquired the LiDAR data over six missions between April 11 and 14, 2012, using an Optech ALTM Gemini 3100 LiDAR system. The sensor recorded four returns with an average point spacing of 1 m between any two neighboring points. We selected 79 LiDAR tiles (several of our 70 field plots required two tiles for complete coverage) for further processing to derive structural metrics.

We merged tiles where necessary to provide seamless coverage of each field plot and then clipped the LiDAR data to each plot using a 12-m radius buffer around plot centers. We applied a 1-m resolution digital elevation model derived from the last returns of LiDAR to remove topographic effects. We then reduced the original LiDAR data (100%) to 80%, 60%,

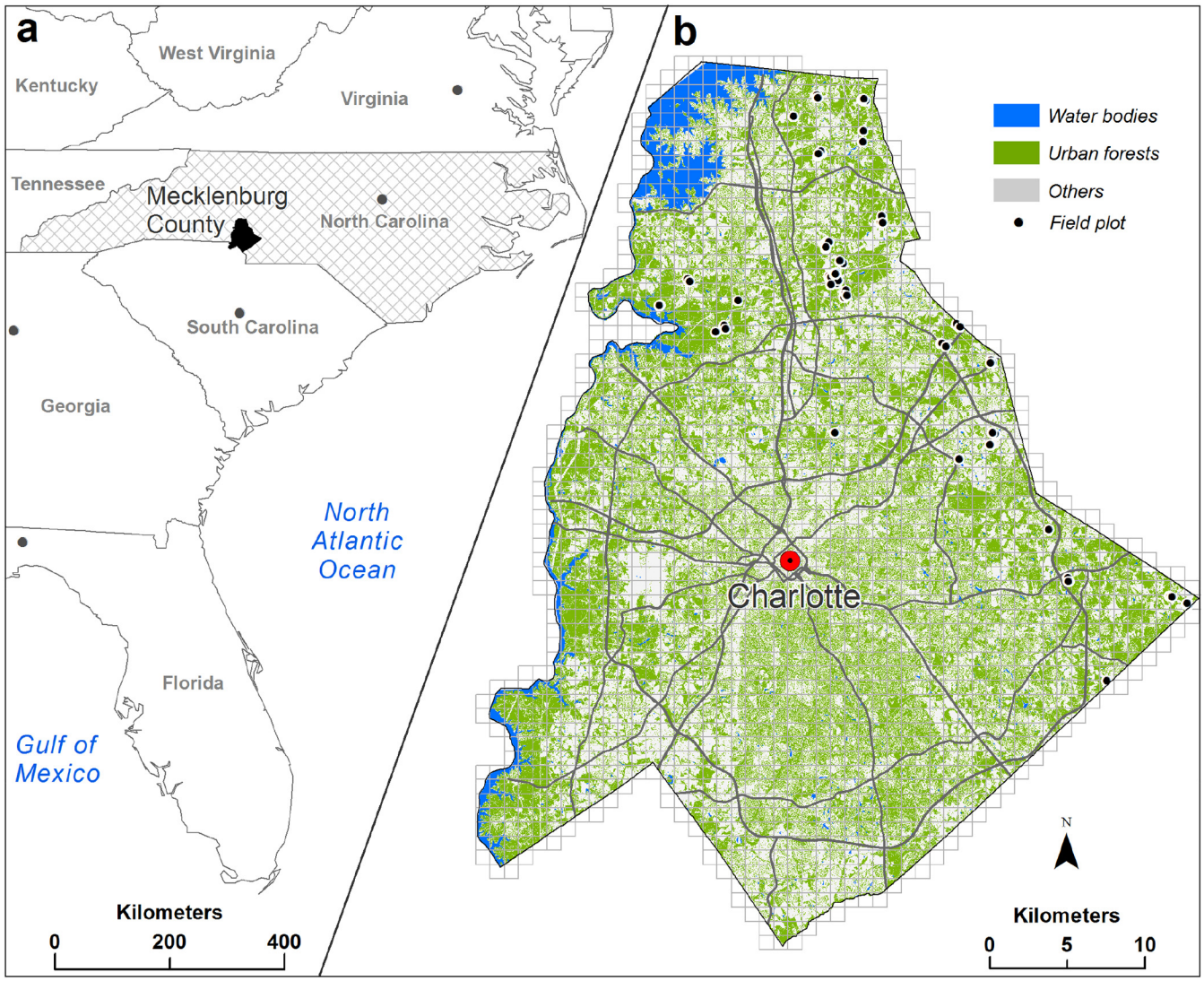


Fig. 1. Study area. (a) Mecklenburg County, North Carolina, USA. (b) Distribution of forest cover and field plots across the county with an overlay of LiDAR tiles.

TABLE I
PARAMETERS USED FOR ESTIMATING FIELD-MEASURED ABOVEGROUND BIOMASS FOR ALL HARDWOOD AND SOFTWOOD SPECIES

	Species group	Parameters		R^2
		β_0	β_1	
Hardwood	Soft maple/birch	–1.9123	2.3651	0.958
	Mixed hardwood	–2.4800	2.4835	0.980
	Hard maple/oak/hickory/beech	–2.0127	2.4342	0.988
Softwood	Cedar/larch	–2.0336	2.2592	0.981
	Pine	–2.5356	2.4349	0.987

40%, 20%, 10%, 5%, and 1% point densities using the "percentage of the total points" reduction algorithm [33] (Figs. 2 and 3). This produced eight LiDAR datasets for each return category, including 100% data and seven sets of reduced point densities (see [26], [27] for full description of data reduction method).

Next, we selected 37 metrics commonly identified in previous research [6]–[8], [10], [26] as the best predictors of biomass (Table II).We extracted these plot-level tree metrics from the eight LiDAR datasets using three LiDAR return combinations: all returns, first returns, and first-in-pulse. We used the height range of 1.5–35 m to exclude understory vegetation (<1.5 m) and objects taller than the trees (>35 m). The metrics extraction process produced 24 sets of plot-level tree metrics.

E. Data Analysis and Model Evaluation

We employed MLR to analyze the effects of LiDAR returns and point density on biomass estimation by establishing the linear statistical relationship between field-measured biomass and selected PVs across the datasets. First, we analyzed the median and mean of field-measured biomass to identify outliers (a median value lower than the mean suggests the presence of outliers). Second, we used scatterplot matrices to assess the collinearity of selected PVs. We also used logarithmic transformations to achieve linearity between field-based biomass and nonlinear forest structural parameters [7], [11], [15]. We then applied the variance inflation factor (VIF) to select noncollinear PVs to overcome issues of over fitting in the model [21]. To assess PV robustness, we used the "regsubsets" function from the LEAPS package in R [18]. This function develops a list

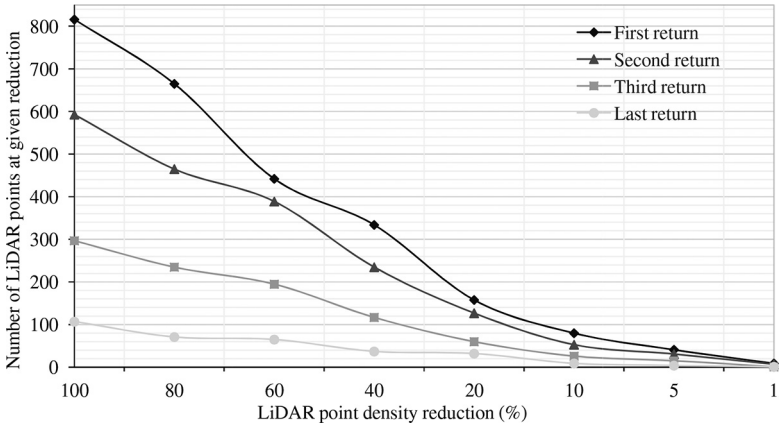


Fig. 2. An example of LiDAR data reduction using percentage algorithm and the distribution of points in each return across reduced point densities.

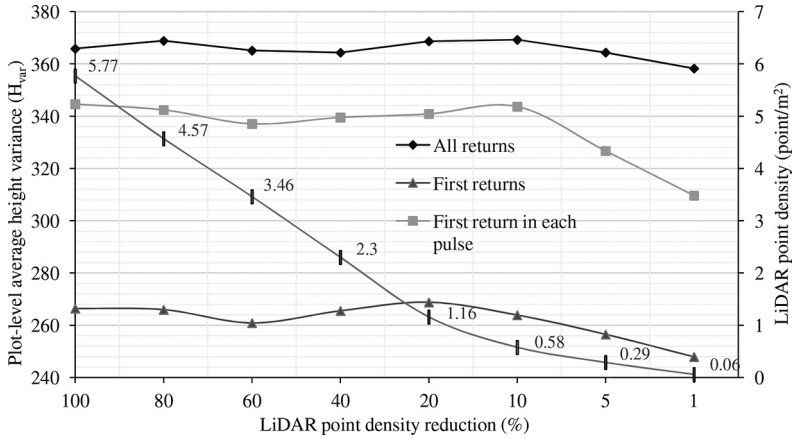


Fig. 3. Distribution of average height variance (H_{var}) derived from LiDAR point density reductions using: (1) all returns, (2) first returns, and (3) first-in-pulse returns. Greater height variance indicates data points that are more widely spread out around the mean and from each other.

TABLE II
PVs DERIVED FROM ALL LiDAR RETURN TYPES AND POINT-DENSITY REDUCTIONS COMBINATIONS AND USED TO DEVELOP REGRESSION MODELS FOR ESTIMATING PLOT-LEVEL BIOMASS

Variable	Description
TRC	Total return count
H_{min}	Height minimum
H_{max}	Height maximum
H_{mean}	Height mean
H_{mode}	Height mode
H_{SD}	Height standard deviation
H_{var}	Height variance
H_{cv}	Height coefficient of variation
H_{ske}	Height skewness
H_{kur}	Height kurtosis
H_{AAD}	Height average absolute deviation
H_{MAD}	Height median absolute deviation
HMADme	Median of the absolute deviations from the overall median
HMADmo	Median of the absolute deviations from the overall mode
H_{IQR}	Height interquartile range
$HP_{(1–99)}$	Height percentile: 1st, 5th, 10th, 20th, 25th, 30th, 40th, 50th (median), 60th, 70th, 75th, 80th, 90th, 95th, and 99th
CRR	Canopy relief ratio
PFRame	Percentage first returns above mean
PFRamo	Percentage first returns above mode
FRaMe	First returns above mean
FRaMo	First returns above mode
ARaMe	All returns above mean
ARaMo	All returns above mode

of models based on multiple scoring criteria [R^2, *adjusted* R^2, Mallows' Cp, and Bayesian information criterion (BIC)] and helps select the best regression model [30]. We also applied Bootstrap measures of relative importance for each predictor in the selected model to determine the relative contribution of PVs in the model. We selected the best subsets of PVs within each model using the lowest Akaike information criterion (AIC) value. Finally, we compared the performance of models across reduced point densities within each LiDAR return category, and between LiDAR return categories using an *adjusted* R^2, and root-mean-squared-error (RMSE) based on a 10-fold cross validation (10-fold CV) analysis. We applied the F-test to determine if the variance of observed biomass was significantly different from the predicted biomass.

III. RESULTS

A. All Returns Biomass Estimates

While data reduction revealed a significant variation in average height variance (H_{var}) across the three LiDAR return categories and with decreasing point densities, we found biomass models showed similar predictive power (Figs. 3 and 4, and

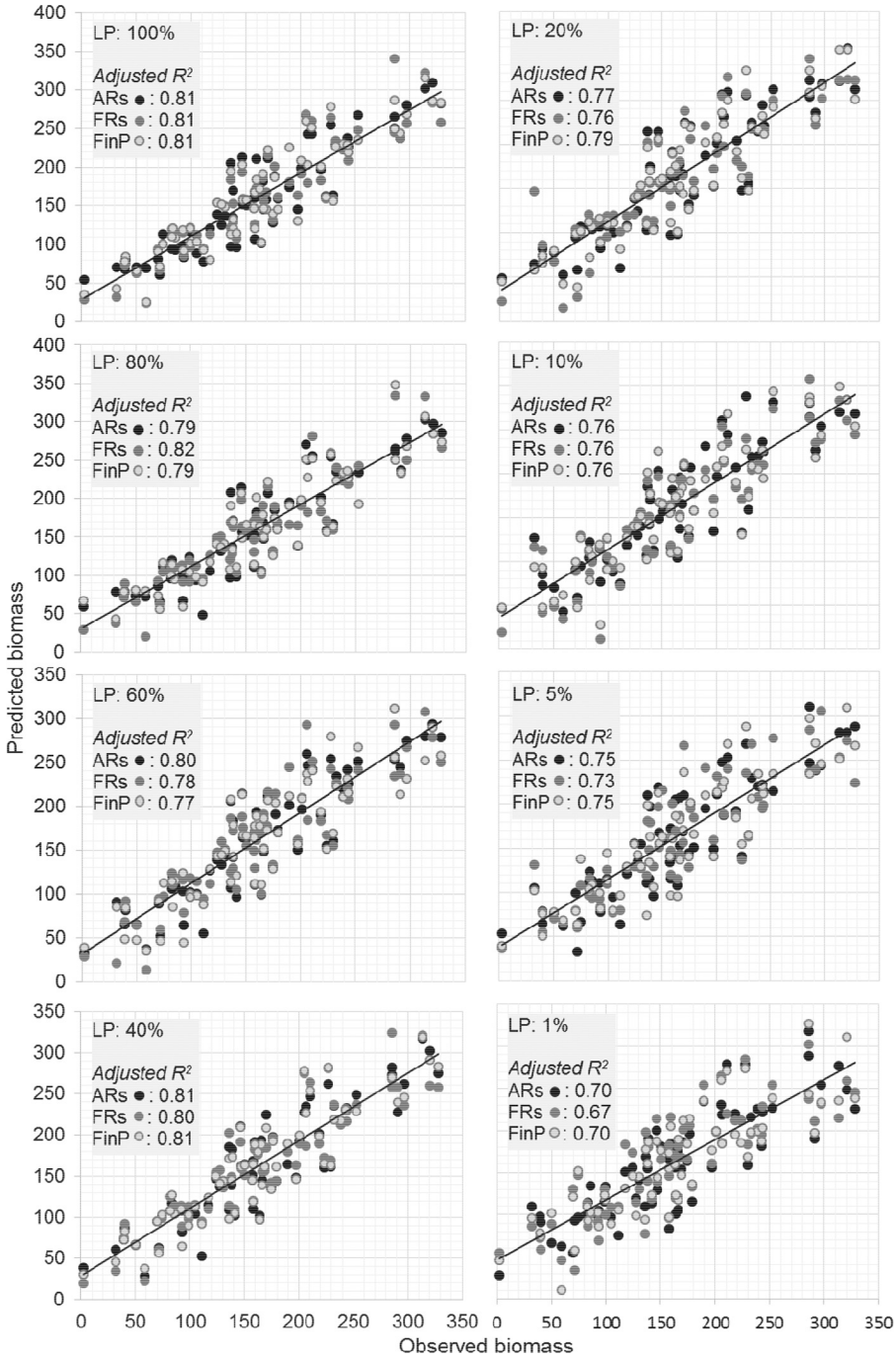


Fig. 4. Predicted versus observed aboveground biomass for each combination of LiDAR return type-and point density reduction (LP – LiDAR point density (%), ARs – all returns, FRs – first returns, and FinP – first-in-pulse returns).

Table III). H_{var} in the all returns models was, on average, 103.6 and 30.0 m^2 higher compared to the models based on first returns and first-in-pulse categories, respectively. A high H_{var} value in the all returns category indicated the presence of middle return points and their wide distribution around the mean and from each other as compared to the first return and first-in-pulse categories (Fig. 3). Biomass models below the 20% point density produced slightly lower *adjusted* R^2 in each LiDAR return category with slightly higher variation in the median biomass (168.0 t/ha at 10%, 150.9 t/ha at 5%, and 168.2 t/ha at 1% point densities) compared to the field-measured biomass (157.9 t/ha) (Fig. 5). Similarly, the RMSE values were approximately 41 t/ha at 1% point density in each LiDAR return category (Table IV).

B. First Return Biomass Estimates

Biomass models based on first return data produced similar estimates across all the reduced LiDAR point densities with the greatest difference in variance being 13.3% when comparing the 100% and 1% density models. We observed a moderate change in predictive power between the 10% and 1%

TABLE III
MLR MODELS FOR PREDICTING ABOVEGROUND BIOMASS OF URBAN FORESTS

LiDAR (%)	Return type	Urban forest model with coefficients	R^2	R^2 adjusted	RMSE (t/ha)	Xval† (t/ha)
	ARs‡	$y = 44.39 + 0.31_{Hvar} - 3.72_{HMADMo} + 3.46_{HP05}$	0.8203	0.8114	31.99	34.02
100	*FRs*	$y = -10.49 + 0.27_{Hvar} - 4.05_{HMADMo} + 2.11_{HP05} + 1.58_{PFRame}$	0.8191	0.8070	32.11	34.95
	FinP	$y = -6.34 + 0.32_{Hvar} - 5.39_{HMADMo} + 2.07_{HP05} + 1.73_{PFRame}$	0.8195	0.8074	32.06	33.87
	ARs‡	$y = 49.26 + 0.31_{Hvar} - 3.65_{HMADMo} + 3.28_{HP05}$	0.8086	0.7992	33.02	35.34
80	*FRs*	$y = -10.56 + 0.29_{Hvar} - 4.20_{HMADMo} + 2.10_{HP05} + 1.52_{PFRame}$	0.8272	0.8157	31.38	34.17
	FinP	$y = 63.90 + 0.32_{Hvar} - 5.29_{HMADMo} + 3.17_{HP05}$	0.8038	0.7941	33.43	35.85
	ARs‡	$y = 5.40 + 0.28_{Hvar} - 2.58_{HMADMo}{}^{**} + 3.12_{HP05} + 0.62_{PFRame}{}^{\times}$	0.8076	0.7948	33.10	36.93
60	*FRs*	$y = -9.34 + 0.28_{Hvar} - 4.41_{HMADMo} + 2.14_{HP05} + 1.56_{PFRame}{}^{**}$	0.7970	0.7835	34.00	37.03
	FinP	$y = 156.75 + 0.29_{Hvar} - 411.18_{Hcv} + 0.84_{PFRame}{}^{\times}$	0.7819	0.7712	35.24	37.96
	ARs‡	$y = 11.13 + 0.28_{Hvar} - 3.36_{HMADMo} + 2.67_{HP05} + 0.75_{PFRame}{}^{*}$	0.8210	0.8090	31.94	34.06
40	*FRs*	$y = -14.56 + 0.27_{Hvar} - 3.99_{HMADMo} + 2.02_{HP05} + 1.70_{PFRame}{}^{**}$	0.8108	0.7982	32.83	35.19
	FinP	$y = 6.03 + 0.29_{Hvar} - 4.21_{HMADMo} + 2.11_{HP05} + 1.08_{PFRame}{}^{**}$	0.8233	0.8115	31.72	33.89
	ARs‡	$y = 26.16 + 0.28_{Hvar} - 3.44_{HMADMo} + 2.70_{HP05} + 0.51_{PFRame}{}^{\times}$	0.7872	0.7730	34.82	37.58
20	*FRs*	$y = -16.39 + 0.27_{Hvar} - 3.46_{HMADMo}{}^{**} + 2.15_{HP05} + 1.53_{PFRame}{}^{**}$	0.7708	0.7555	36.13	39.79
	FinP	$y = 18.87 + 0.28_{Hvar} - 3.71_{HMADMo} + 2.08_{HP05} + 0.86_{PFRame}{}^{*}$	0.8019	0.7887	33.59	35.84
	ARs‡	$y = 19.57 + 0.27_{Hvar} - 3.64_{HMADMo}{}^{*} + 2.84_{HP05} + 0.62_{PFRame}{}^{\times}$	0.7747	0.7597	35.82	38.87
10	*FRs*	$y = 118.64 + 0.19_{Hvar} + 1.83_{HMADMo}{}^{*} - 332.11_{Hcv} + 1.28_{Hmode}{}^{**}$	0.7752	0.7602	35.78	41.49
	FinP	$y = 163.76 + 0.24_{Hvar} - 328.82_{Hcv} + 0.60_{Hmode}{}^{*}$	0.7678	0.7564	36.37	39.26
	ARs‡	$y = 34.12 + 0.28_{Hvar} - 3.47_{HMADMo}{}^{*} + 2.30_{HP05} + 0.56_{PFRame}{}^{\times}$	0.7624	0.7466	36.79	39.35
5	*FRs*	$y = -4.40 + 0.21_{Hvar} + 2.35_{HP05} + 13.10_{Hkur}$	0.7388	0.7260	38.57	40.59
	FinP	$y = -68.49 + 0.25_{Hvar} - 1.77_{HMADMo}{}^{\times} + 1.75_{HP05} + 219.76_{CRR}{}^{*}$	0.7642	0.7484	36.65	40.01
	ARs‡	$y = -119.15 + 1.27_{TRC}{}^{*} + 2.56_{Hmin} + 0.17_{Hvar} + 259.90_{CRR}$	0.7164	0.6975	40.19	43.63
1	*FRs*	$y = 23.45 + 2.71_{Hmin} + 0.17_{Hvar} - 102.48_{CRR} + 30.68_{Hkur}$	0.6941	0.6737	41.75	45.86
	FinP	$y = -10.83 + 2.55_{Hmin} + 0.16_{Hvar} + 26.57_{Hkur}$	0.7093	0.6950	40.51	43.52

ARs, all returns; FRs, first returns; and FinP, first-in-pulse returns.
% Percentage of original LiDAR data.
‡Biomass model [26].
†10-fold cross validation.
Level of significance: 0.001 "**" 0.01 "*" 0.05 "×" 0.1.

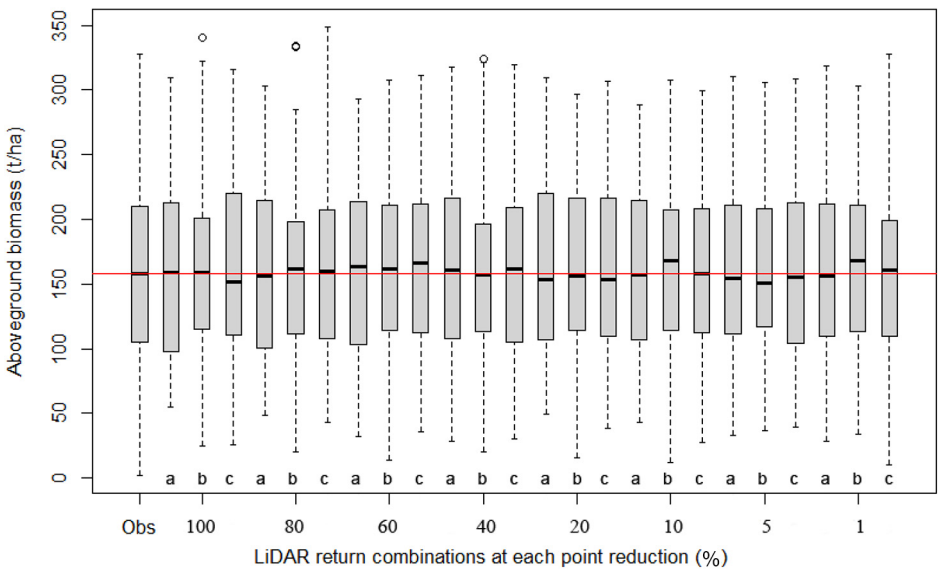


Fig. 5. Predicted biomass (tons/ha) categorized by each combination of LiDAR return type and point-density reduction. Red horizontal line indicates the median of observed biomass (Obs, field-measured biomass; a, all returns, b, first returns, and c, first-in-pulse returns). The bottom and top of box in each biomass model are the first and third quartiles. The division inside each box represents the median. The ends of the whiskers represent the minimum and maximum of all of the data with the suspected outliers in each biomass model.

reduced point-density models (Fig. 4 and Table III). RMSE values of the biomass models exhibited a similar trend with a maximum difference of 3.7 t/ha between 100% and 10%, and 5.9 t/ha between 10% and 1% point densities (Table IV). The largest difference in the 10-fold CV was 10.9% when comparing models based on 100% and 1% point densities. While we observed the presence of similar PVs in models (except at 10%, 5%, and 1% point densities), H_{var} was common to all models (Table III) and contributed up to 65% in total biomass variance (*adjusted* R^2). The *F*-test showed that

TABLE IV
CHANGE IN ADJUSTED R^2 AND RMSE FOR EACH COMBINATION OF LIDAR RETURN TYPE AND POINT-DENSITY REDUCTION

LiDAR reductions	*Adjusted R^2*			RMSE (t/ha)		
	ARs	*FRs*	*FinP*	*ARs*	*FRs*	*FinP*
100	0.81	0.81	0.81	31.99	32.11	32.06
80	0.80	0.82	0.80	33.02	31.38	33.43
60	0.79	0.78	0.77	33.10	34.00	35.24
40	0.81	0.80	0.81	31.94	32.83	31.72
20	0.77	0.76	0.79	34.82	36.13	33.59
10	0.76	0.76	0.76	35.82	35.78	36.37
5	0.75	0.73	0.75	36.79	38.57	36.65
1	0.70	0.67	0.70	40.19	41.75	40.51

ARs, all returns; FRs, first returns; FinP, first-in-pulse returns.

predicted biomass estimates were similar to field measurements, including at 10%, 5%, and 1% point densities.

C. First-in-Pulse Biomass Estimates

We found biomass estimates and general performance of first-in-pulse models to be consistently similar to the results based on first returns models across all reduced point densities. Overall, biomass estimates (*adjusted* R^2) and RMSE differences between models based on first returns and first-in-pulse categories varied by 2% and 1 t/ha, respectively. We observed trivial differences in biomass variance among the three LiDAR returns categories and across all reduced LiDAR point densities with one exception. At 1% LiDAR point density, we found a 3% difference in *adjusted* R^2 between the first returns and other two LiDAR return categories (Fig. 4; Table IV). We observed the presence of similar PVs and their relative contributions to biomass estimates across all the reduced point densities, including the presence of the H_{var} variable.

IV. DISCUSSION

LiDAR data have proven valuable in the assessment of biomass and other forest biophysical parameters. Typically, LiDAR data with higher point densities (both horizontally and vertically) help produce a higher quality 3-D representation of the earth's surface and may prove useful for multiple purposes, including subcanopy forest studies, e.g., detection and mapping evergreen understory invasive Chinese privet (*Ligustrum sinense*) [27]. However, given that the higher data procurement costs and data volume pose major challenges to large-area assessments, this study provides a framework for using reduced volume of data for estimating aboveground biomass of highly fragmented and spatially heterogeneous urban forests. Our urban environment study suggests that using first returns or first-in-pulse of LiDAR data, normalized by topography, produces biomass estimates similar to estimates derived using all returns of LiDAR data. These findings validate our assumption that LiDAR first return data are sufficient for biomass assessment. Since our study is conducted in temperate forests and uses leaf-off season LiDAR data, we note the effects of return number could be different for forests with thick evergreen understory vegetation or where LiDAR data are acquired during the leaf-on season. For example, [26] suggested considering forest types model to address greater diversity of forest types for reasonable biomass estimates at regional scales. Across all data reductions from 100% to 1%, we observed biomass estimates from both the first returns and first-in-pulse that are similar to estimates based on all returns LiDAR data. This also suggests that direct use of first returns or first-in-pulse data may be a more efficient and effective alternative to point-density reduction approach in estimating biomass.

Estimates derived from first return and first-in-pulse models were similar to findings of [5]. Dalponte *et al.* observed a slight increase of 2% (*adjusted* R^2 of all return model—0.77 and first return model—0.75) in the overall model accuracy for estimating individual stem volume using variables from all returns over first returns model. Similarly, [4], [10], and [14] observed that the metrics derived from first returns were not significantly different than using all returns in predicting biomass. The comparable performance is perhaps due to the leaf-off season LiDAR data as open canopy forest allows good infiltration of first returns to the forest floor and permits sufficient representation of forest vertical structure [28]. However, the increasingly similar performance of first return and first-in-pulse models to all return models across the range of reduced point densities invalidates the seasonal factor. Moreover, we observed the presence of the same PVs in biomass models of all three return categories. Therefore, it may be possible to use first returns instead of all returns for estimating biomass for large areas without a significant decrease in overall accuracy. The first-in-pulse may offer similar prospects where LiDAR data are available in tile format covering large areas of interest. In addition, these approaches could be particularly useful in analyzing LiDAR data collected for hydrological studies and containing only the first and last returns.

We found virtually identical predictive power for the first return models across the reduced point densities. This finding validates our hypothesis that given a minimal number of LiDAR points suitable for extracting structural metrics, there is no significant compromise in overall model performance. For example, we found overall accuracies, on average, only 1.5% lower for the first return models below 80% point densities compared to using all returns (Table IV). As suggested by [9] and [16], a higher point density does not necessarily produce data or outputs with greater information content. We observed that models developed using the first return and first-in-pulse accompanied by data reduction maintained overall model performance as compared to all return models. This suggests that first return data combined with point-density reduction is sufficient for large-area biomass assessments while maintaining the predictive power in regression models. However, point-density reduction approaches generally require time-intensive data analysis to identify an ideal cutoff point for data resampling. We observed missing values in many metrics derived at 5% and 1% reductions in both the first returns and first-in-pulse categories. This suggests that variations in point density within plots compared to the overall study area limit the utility of point-density reduction algorithms. This warrants further research to determine the requisite minimum number of LiDAR points for an individual tree or plot that maintains the structural integrity of the metrics suitable for accurate biomass estimation.

Evaluation of the MLR equations revealed that the H_{var} variable contributed substantially to overall performance of the first returns and first-in-pulse models. This corroborates the findings of [26] and [32] suggesting that H_{var} is an effective metric for classifying vertical forest structural configurations. In addition, we observed the median of the absolute deviations from the overall mode (HMADmo), height 5th percentile (HP_{05}), and the percentage of first returns above mean (PFRame) in all models except those developed from point densities below 20%. Among these, HMADmo is a robust measure of height variability, HP_{05} represents the lowermost height in the plot for each return category, and PFRame is the percentage of first returns above the mean height within the data sample. However, it is worth noting that the consideration of all LiDAR-derived PVs for model development is useful as some PVs might reflect finer scale forest structures and contribute to improved biomass estimation, such as height minimum and lower percentiles. The recurrence of these four PVs suggests that: 1) field-measured biomass is primarily related to LiDAR points representing tree height and top-of-canopy configuration, and 2) first returns provide the greatest contribution to aboveground biomass estimation regardless of the presence of multiple returns in LiDAR data. Overall, our study demonstrates that, LiDAR first return data generate aboveground biomass estimates that are comparable to estimates obtained using first-in-pulse and all returns.

V. Conclusion

Repeated large-area assessments of forest biomass using LiDAR data are required to understand the dynamic relationships between changing climate, carbon stocks, and land-use change in urbanizing areas. Across the spectrum of models developed in this research, we observed variance in predicted biomass estimates similar to field-measured biomass. Our findings suggest that first return or first-in-pulse LiDAR data alone, compared to all returns, are sufficient for estimating biomass. The H_{var} distribution shows that the use of first returns LiDAR data normalized by elevation is an effective alternative to any other LiDAR returns combination, including point-density reduction, for estimating urban forest biomass without compromising biomass estimates accuracy. This also applies to the utilization of existing LiDAR data (first and last returns) originally acquired for hydrological studies. Additionally, using first return data combined with reduced point density can minimize data procurement costs and overcome computational overhead without compromising biomass estimation accuracy. Higher *adjusted* R^2 and lower RMSE values of biomass models above 20% point density suggest that the first returns of LiDAR data with an average point spacing of 0.70–1.5 m (approximately 1.35 points/m^2) may offer cost-effective data procurement and processing for urban forest landscape management. We found H_{var} to be the most robust plot-level structural LiDAR metric for predicting biomass, corroborating findings in [26], [27], and [32]. Overall, this study suggests that multiple return, high point-density LiDAR data may not significantly improve urban forest biomass estimations over large areas despite the greater procurement costs and computational overhead. We conclude that using: 1) first returns LiDAR data for biomass estimation is an effective and accurate alternative to using first-in-pulse data, multiple return data, and/or applying data reductions, and 2) first return data combined with reduced point density can overcome computational challenges in large-area applications.

Acknowledgment

The authors would like to thank Dr. R. Bianchetti, Michigan State University, for valuable comments and feedback on this paper. They would also like to thank Charlotte-Mecklenburg County Government Office, City of Charlotte, for providing LiDAR data, and fellow researchers at North Carolina State University for their valuable comments and feedback on this paper.

References

[1] S. G. Zolkos, S. J. Goetz, and R. Dubayah, "A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing," *Remote Sens. Environ.*, vol. 128, pp. 289–298, 2013.

[2] G. Chen and G. J. Hay, "An airborne LiDAR sampling strategy to model forest canopy height from quickbird imagery and GEOBIA," *Remote Sens. Environ*., vol. 115, pp. 1532–1542, 2011.

[3] G. Chen, M. A. Wulder, J. C. White, T. Hilker, and N. C. Coops, "LiDAR calibration and validation for geometric-optical modeling with Landsat imagery," *Remote Sens. Environ.*, vol. 124, pp. 384–393, 2012.

[4] Q. Chen, G. V. Laurin, J. J. Battles, and D. Saah, "Integration of airborne LiDAR and vegetation types derived from aerial photography for mapping aboveground live biomass," *Remote Sens. Environ.*, vol. 121, pp. 108–117, 2012.

[5] M. Dalponte, N. C. Coops, L. Bruzzone, and D. Gianelle, "Analysis on the use of multiple returns LiDAR data for the estimation of tree stems volume," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 4, pp. 310–318, Dec. 2009.

[6] R. O. Dubayah *et al.*, "Estimation of tropical forest height and biomass dynamics using LiDAR remote sensing at La Selva, Costa Rica," *J. Geophys. Res.*, vol. 115, p. G00E09, 2010.

[7] G. W. Frazer, S. Magnussen, M. A. Wulder, and K. O. Niemann, "Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass," *Remote Sens. Environ*., vol. 115, pp. 636–649, 2011.

[8] M. García, D. Riaño, E. Chuvieco, and F. M. Danson, "Estimating biomass carbon stocks for a mediterranean forest in central spain using LiDAR height and intensity data," *Remote Sens. Environ.*, vol. 114, pp. 816–830, 2010.

[9] N. R. Goodwin, N. C. Coops, and D. S. Culvenor, "Assessment of forest structure with airborne LiDAR and the effects of platform altitude," *Remote Sens. Environ.*, vol. 103, pp. 140–152, 2006.

[10] S. A. Hall, I. C. Burke, D. O. Box, M. R. Kaufmann, and J. M. Stoker, "Estimating stand structure using discrete-return LiDar: An example from low density, fire prone ponderosa pine forests," *For. Ecol. Manage.*, vol. 208, pp. 189–209, 2005.

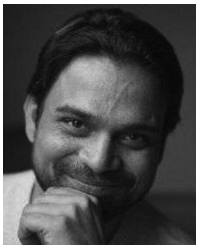
[11] A. T. Hudak *et al.*, "Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return LiDAR and multispectral satellite data," *Can. J. Remote Sens.*, vol. 32, pp. 126–138, 2006.

[12] M. K. Jakubowski, Q. H. Guo, and M. Kelly, "Tradeoffs between LiDAR pulse density and forest measurement accuracy," *Remote Sens. Environ.*, vol. 130, pp. 245–253, 2013.

[13] J. C. Jenkins, D. C. Chojnacky, L. S. Heath, and R. A. Birdsey, "National-scale biomass estimators for united states tree species," *For. Sci.*, vol. 49, pp. 12–35, 2003.

[14] Y. Kim, Z. Q. Yang, W. B. Cohen, D. Pflugmacher, C. L. Lauver, and J. L. Vankat, "Distinguishing between live and dead standing tree biomass on the north rim of grand canyon national park, USA using small-footprint LiDAR data," *Remote Sens. Environ.*, vol. 113, pp. 2499–2510, 2009.

[15] M. A. Lefsky, W. B. Cohen, D. J. Harding, G. G. Parker, S. A. Acker, and S. T. Gower, "LiDAR remote sensing of above-ground biomass in three biomes," *Global Ecol. Biogeogr.*, vol. 11, pp. 393–399, 2002.
[16] K. Lim, C. Hopkinson, and P. Treitz, "Examining the effects of sampling point densities on laser canopy height and density metrics," *For. Chron.*, vol. 84, pp. 876–885, 2008.
[17] K. S. Lim and P. M. Treitz, "Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators," *Scand. J. For. Res.*, vol. 19, 558–570, 2004.
[18] T. Lumley, (2013). Package 'leaps' [Online]. Available: http://cran.r-project.org/web/packages/leaps/leaps.pdf
[19] R. J. McGaughey *FUSION/LDV: Software for LIDAR Data Analysis and Visualization*. Seattle, WA, USA: Dept. Agric., Forest Service, Pacific Northwest Research Station, 2014.
[20] R. K. Meentemeyer, W. W. Tang, M. A. Dorning, J. B. Vogler, N. J. Cunniffe, and D. A. Shoemaker, "FUTURES: Multilevel simulations of emerging urban-rural landscape structure using a stochastic patch-growing algorithm," *Ann. Assoc. Amer. Geogr.*, vol. 103, pp. 785–807, 2013.
[21] R. M. O'Brien, "A caution regarding rules of thumb for variance inflation factors," *Qual. Quant.*, vol. 41, pp. 673–690, 2007.
[22] S. C. Popescu, R. H. Wynne, and R. F. Nelson, "Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass," *Can. J. Remote Sens.*, vol. 29, pp. 564–577, 2003.
[23] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013.
[24] M. Renslow, P. Greenfield, and T. Guay, "Evaluation of multi-return LIDAR for forestry applications," US Dept. Agric. For. Serv.—Remote Sens. Appl. Center, Salt Lake City, UT, USA, Tech. Rep. RSAC-2060/4810-LSP-0001-RPT1, 2000, p. 12.
[25] J. B. Riegel, E. Bernhardt, and J. Swenson, "Estimating above-ground carbon biomass in a newly restored coastal plain wetland using remote sensing," *Plos One*, vol. 8, pp. 1–10, 2013.
[26] K. K. Singh, G. Chen, J. B. McCarter, and R. K. Meentemeyer, "Effects of LiDAR point density and landscape context on estimates of urban forest biomass," *ISPRS J. Photogramm. Remote Sens.*, vol. 101, pp. 310–322, 2015.
[27] K. K. Singh, A. J. Davis, and R. K. Meentemeyer, "Detecting understory plant invasion in urban forests using LiDAR," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 38, pp. 267–279, 2015.
[28] V. Thomas, P. Treitz, J. H. McCaughey, and I. Morrison, "Mapping stand-level forest biophysical variables for a mixedwood boreal forest using LiDAR: An examination of scanning density," *Can. J. For. Res.-Revue Canadienne De Recherche Forestiere*, vol. 36, pp. 34–47, 2006.
[29] J. A. N. van Aardt, R. H. Wynne, and R. G. Oderwald, "Forest volume and biomass estimation using small-footprint LiDAR-distributional parameters on a per-segment basis," *For. Sci.*, vol. 52, pp. 636–649, 2006.
[30] G. M. S. Vianna, M. G. Meekan, T. H. Bornovski, and J. J. Meeuwig, "Acoustic telemetry validates a citizen science approach for monitoring sharks on coral reefs," *Plos One*, vol. 9, pp. 1–12, 2014.
[31] M. A. Wulder *et al.*, "LiDAR sampling for large-area forest characterization: A review," *Remote Sens. Environ.*, vol. 121, pp. 196–209, 2012.
[32] D. A. Zimble, D. L. Evans, G. C. Carlson, R. C. Parker, S. C. Grado, and P. D. Gerard, "Characterizing vertical forest structure using small-footprint airborne LiDAR," *Remote Sens. Environ.*, vol. 87, pp. 171–182, 2003.
[33] BCAL LiDAR Tools ver 1.5.3, Idaho State Univ., Dept. Geosci., Boise Center Aerospace Lab. (BCAL), Boise, ID, USA, 2013 [Online]. Available: http://bcal.geology.isu.edu/envitools.shtml



Kunwar K. Singh received the B.S. degree in biology in 1999 and the M.S. degree in environmental science in 2001, both from the University of Lucknow, India, the M.Tech. degree in remote sensing and geographic information science with an emphasis on forest and ecological applications from Indian Institute of Remote Sensing, Dehradun, India, in 2005, and the Ph.D. degree in forestry and environmental resources from the North Carolina State University, Raleigh, NC, USA, in 2014.

He is currently a Postdoctoral Fellow with the Northern Plant Ecology Laboratory, Department of Biology, University of Saskatchewan, Saskatoon, SK, Canada and a visiting research scholar in the Center for Geospatial Research at the University of Georgia, Athens, GA, USA. His research focuses on disturbance regimes and forest composition and structure. He uses air- and space- borne remote sensing data, ground observations, and statistical methods to both describe and predict the impacts of disturbances on ecosystems.

Gang Chen receieved the B.S. degree in geomatics engineering and the M.S. degree in photogrammetry and remote sensing from Wuhan University, Hubei, China, in 2003 and 2006, respectively, and the Ph.D. degree in geography from the University of Calgary, Calgary, AB, Canada, in 2011.

He is an Assistant Professor with the University of North Carolina at Charlotte, Charlotte, NC, USA. His research interests include the application of remote sensing, geospatial sciences, and machine learning to assess environmental change (with an emphasis on the dynamics of forest landscapes) in response to anthropogenic and natural disturbances. He has authored more than 30 articles in peer-reviewed journals, and served as a peer-reviewer for over 30 different international journals, funding agencies, and scientific conferences.

Dr. Chen was the recipient of the Best Ph.D. Thesis Award from the Canadian Remote Sensing Society in 2011 and the North Carolina Space Grant New Investigators Program Award in 2014.

John B. Vogler received the B.A. degree in geography from the University of North Carolina, Greensboro, NC, USA, in 1994, and the M.A. degree in geography from the University of North Carolina, Chapel Hill, NC, USA, in 1998.

He is currently a Research Scholar and Geographer with the Center for Geospatial Analytics. His research interests include dynamic spatial-temporal modeling, environmental remote sensing, and using citizen science and volunteered geographic information to find collaborative solutions to societal and environmental problems. He has authored more than 20 research articles in peer-reviewed journals and co-authored a book on the Global Positioning System for the social sciences.

Ross K. Meentemeyer received the B.S. in physical geography from the University of Georgia in 1994 and the Ph.D. degree in geography from the University of North Carolina Chapel Hill, Chapel Hill, NC, USA in 2000.

He is President of the US-International Association of Landscape Ecology (2016–18), the Director of the Center for Geospatial Analytics (CGA), and a Chancellor's Faculty Excellence Program Professor at North Carolina State University. He applies tools from geospatial analytics to understand and communicate the vital roles that place, spatial interaction, and perspectives of scale play in models of environmental change and decision-making. His applied and participatory research projects are engaging stakeholders for collaborative solutions to natural resource challenges. His research interests include the application of dynamic epidemiological models to know when, where and how to collectively manage the spread of emerging forest disease in heterogeneous landscapes, and simulating alternative futures of land-use change to explore trade-offs in the provisioning of ecosystem services in rapidly urbanizing regions.