



Lidar-based estimates of aboveground biomass in the continental US and Mexico using ground, airborne, and satellite observations



Ross Nelson ^{a,*}, Hank Margolis ^{a,b}, Paul Montesano ^{a,c}, Guoqing Sun ^{a,d}, Bruce Cook ^a, Larry Corp ^{a,c}, Hans-Erik Andersen ^e, Ben deJong ^f, Fernando Paz Pellat ^g, Thaddeus Fickel ^h, Jobriath Kauffman ⁱ, Stephen Prisley ⁱ

^a Biospheric Sciences Lab, Code 618, NASA-GSFC, Greenbelt, MD 20771, USA

^b Centre d'Étude de la Forêt, Université Laval, Québec City, Québec, G1V0A6, Canada, and Earth Science Division, NASA Headquarters, 300 E Street SW, Washington, D.C. 20546-0001, USA

^c Science Systems and Applications, Inc., 10210 Greenbelt Road, Lanham, MD 20706, USA

^d Department of Geographical Sciences, Lefrak Hall, University of Maryland, College Park, MD 20742, USA

^e US Forest Service, Pacific Northwest Research Station, 107 Anderson Hall, University of Washington, P.O. Box 352100, Seattle, WA 98195-2100, USA

^f El Colegio de la Frontera Sur, Av. Ranco Polígono 2-A, Ciudad Industrial, 24500 Lerma, Campeche, Mexico

^g Colegio de Postgraduados en Ciencias Agrícolas, Km 36.5, Texcoco 5, México-Texcoco 56230, Montecillo, Estado de México, Mexico

^h Infrared Baron, Inc., 2372 North 1st St., Suite B, Hermiston, OR 97838, USA

ⁱ Dept. of Forest Resources and Environmental Conservation, Cheatham Hall, 310 West Campus Drive, Virginia Tech, Blacksburg, VA 24061, USA

ARTICLE INFO

Article history:

Received 13 May 2016

Received in revised form 14 October 2016

Accepted 28 October 2016

Available online 14 November 2016

Keywords:

ICESat/GLAS

Hybrid 3-phase sampling

Model-based

Forest biomass

ABSTRACT

Existing national forest inventory plots, an airborne lidar scanning (ALS) system, and a space profiling lidar system (ICESat-GLAS) are used to generate circa 2005 estimates of total aboveground dry biomass (AGB) in forest strata, by state, in the continental United States (CONUS) and Mexico. The airborne lidar is used to link ground observations of AGB to space lidar measurements. Two sets of models are generated, the first relating ground estimates of AGB to airborne laser scanning (ALS) measurements and the second set relating ALS estimates of AGB (generated using the first model set) to GLAS measurements. GLAS then, is used as a sampling tool within a hybrid estimation framework to generate stratum-, state-, and national-level AGB estimates. A two-phase variance estimator is employed to quantify GLAS sampling variability and, additively, ALS-GLAS model variability in this current, three-phase (ground-ALS-space lidar) study. The model variance component characterizes the variability of the regression coefficients used to predict ALS-based estimates of biomass as a function of GLAS measurements. Three different types of predictive models are considered in CONUS to determine which produced biomass totals closest to ground-based national forest inventory estimates - (1) linear (LIN), (2) linear-no-intercept (LNI), and (3) log-linear. For CONUS at the national level, the GLAS LNI model estimate (23.95 ± 0.45 Gt AGB), agreed most closely with the US national forest inventory ground estimate, 24.17 ± 0.06 Gt, i.e., within 1%. The national biomass total based on linear ground-ALS and ALS-GLAS models (25.87 ± 0.49 Gt) overestimated the national ground-based estimate by 7.5%. The comparable log-linear model result (63.29 ± 1.36 Gt) overestimated ground results by 261%. All three national biomass GLAS estimates, LIN, LNI, and log-linear, are based on 241,718 pulses collected on 230 orbits. The US national forest inventory (ground) estimates are based on 119,414 ground plots. At the US state level, the average absolute value of the deviation of LNI GLAS estimates from the comparable ground estimate of total biomass was 18.8% (range: Oregon, -40.8% to North Dakota, 128.6%). Log-linear models produced gross overestimates in the continental US, i.e., >2.6x, and the use of this model to predict regional biomass using GLAS data in temperate, western hemisphere forests is not appropriate. The best model form, LNI, is used to produce biomass estimates in Mexico. The average biomass density in Mexican forests is 53.10 ± 0.88 t/ha, and the total biomass for the country, given a total forest area of 688,096 km², is 3.65 ± 0.06 Gt. In Mexico, our GLAS biomass total underestimated a 2005 FAO estimate (4.152 Gt) by 12% and overestimated a 2007/8 radar study's figure (3.06 Gt) by 19%.

Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

E-mail addresses: rfn104@gmail.com (R. Nelson), hank.a.margolis@nasa.gov (H. Margolis), paul.m.montesano@nasa.gov (P. Montesano), guoqing.sun-1@nasa.gov (G. Sun), bruce.cook@nasa.gov (B. Cook), lawrence.a.corp@nasa.gov (L. Corp), handerse@fs.fed.us (H.-E. Andersen), bjong@ecosur.mx (B. deJong), pellat@colpos.mx (F.P. Pellat), tad@infraredbaron.com (T. Fickel), jkauffman@vt.edu (J. Kauffman), prisley@vt.edu (S. Prisley).

1. Introduction

Two NASA space lidars are planned for launch over the next few years, (1) a near-polar photon-counting instrument, ICESat-2/ATLAS (Gwenzi and Lefsky, 2014; <http://icesat.gsfc.nasa.gov/icesat2/instrument.php>), and (2) a waveform instrument in an equatorial orbit on board the International Space Station, GEDI (Global Ecosystem Dynamics Investigation, Dubayah et al., 2014a, 2014b; <http://science.nasa.gov/missions/gedi>). Both will be multi-beam profilers that will intensively sample the Earth's terrestrial surface. Unlike the single-beam ICESat/GLAS (Ice, Cloud, and land Elevation Satellite/Geoscience Laser Altimeter System) which collected useful vegetation data from 2003 to 2007, these new space ranging instruments will collect sequential measurements along multiple near-nadir tracks beneath their orbital paths. The challenge to researchers will be to transform these linear, sequential measurements into robust, repeatable, statistically defensible, area-based estimates of biomass and carbon. To do this, researchers will have (1) to develop models to predict ground-based biomass using the space lidar measurements, and (2) to process these ground-space biomass estimates within a sampling framework that makes sense with respect to the observational strategies used for the collection of the space data.

To this end, much work has been done with GLAS data to provide such a conceptual framework, and we continue this effort with the current study. The purpose of this investigation is to further refine techniques which might be employed by researchers to develop large-area estimates of biomass. In particular the objectives of this study are:

1. Determine which of 3 predictive models – linear, linear-no-intercept (forced through origin), and log-linear – best serves to predict total biomass at the state and national level.
2. Quantify how well the best model does at the state and national level with respect of assessing biomass density, total biomass, and the variability of these estimates.
3. Compare satellite-lidar-based estimates of biomass density and total biomass to a comprehensive ground survey in the continental United States (CONUS).
4. Extend those findings to inventory Mexico at the state and national levels.

One intention of this study is to empirically characterize how well we might expect to estimate regional biomass using a space laser system. We do not argue that GLAS-, ICESat-2-, or GEDI-like systems should replace existing regional or national forest inventories. Comprehensive sets of ground observations are the gold-standard, and these ground-based arrays of field plots typically deliver regional and site-specific information concerning not only biomass, volume, basal area, canopy height and density, but also information concerning forest health, downed-woody debris, tree and shrub species, understory cover, herbaceous and moss/lichen cover, and soils information, observations that are difficult if not impossible to reliably infer from space remote sensing platforms. For instance, over 130 separate metrics are recorded on each USFS-FIA (US Forest Service – Forest Inventory & Analysis) level-3 plot. Thus far, GLAS can be used to reliably measure or estimate, arguably, six of them – canopy height, canopy density, aboveground biomass, aboveground carbon, volume, and basal area. We note that GLAS has been used to estimate LAI (Tang et al., 2014a) and vertical vegetation profile (Tang et al., 2014b), but these measurements are not typically acquired on national forest inventory (NFI) ground plots hence are not included in this list. However, space lidars can play a critical role in that they provide a repeated measurement and estimation capability on areas where access is limited, difficult, expensive, or dangerous, e.g., Siberia, northern Canada, northern and western Alaska, Africa, portions of the Amazon, SE Asia, and the Middle East and thus will be critical for developing a robust, global forest-monitoring capability. These space lidars can also provide useful ground surface elevation measurements, that, when linked with canopy surface elevation from

spaceborne image stereogrammetry, can provide spaceborne estimates of canopy height in sparse forests (Montesano et al., 2014a). Lidars can augment but not replace ground observations, and lidar studies must depend on representative, ground-based measurements to develop predictive models used to estimate the response variable of interest, e.g., biomass, carbon.

2. Background

Numerous regional biomass studies that integrate ground observations and space lidar measurements have been conducted in an effort to develop techniques that can be used to better estimate forest biomass and carbon across large areas. These studies have all depended on the ICESat/GLAS satellite to provide the space-based ranging measurements needed to develop and refine large-area forest inventory approaches. These data, available for download via the National Snow and Ice Data Center (<http://nsidc.org/data/icesat>), have been used to inventory Quebec (Boudreau et al., 2008; Nelson et al., 2009a), central Siberia (Nelson et al., 2009b), the North American boreal forests (Margolis et al., 2015), and the circumpolar boreal forests (Neigh et al., 2013). The current study relies on the three-phase sampling techniques, a hybrid estimation framework (Ståhl et al., 2016), and two-phase estimators proposed by Ståhl et al. (2011) and employed by Neigh et al. (2013) and Margolis et al. (2015). The term “hybrid estimation framework” or “hybrid” refers to a situation where the 1st phase, in our case the GLAS orbital acquisitions, is a probability sample, i.e., a systematic-like sample assumed to have been randomly acquired, and the 2nd phase rests on model-based principles, in our case the models which predict ALS estimates of biomass as a function of GLAS measurements.

To develop regression models to predict biomass using satellite lidar data, the analyst must tie the ranging measurements acquired by the space lidar to ground observations of biomass. Currently, there are three ways that this can be done.

1. An analyst can hope that, by luck, one or more of the space lidar pulses image(s) an existing ground plot. With GLAS, across large areas in the temperate latitudes and given the duty cycle of the GLAS system, this probability turns out to be exceedingly small. For instance, in CONUS, we estimate that only ~50 of the 119,414 nonzero biomass plots, (i.e., FIA plots where at least one subplot has measureable forest biomass) scattered across the US would have been measured by the GLAS acquisitions considered in this study. This situation should markedly improve when the next generation of profiling space lidars are launched in 2017/2018, for these future missions will sample the Earth's surface much more intensively.
2. An analyst can establish ground plots directly on GLAS pulses previously collected. GLAS pulse center locations are known to within 3.1 m, 1σ (Magruder et al., 2010), enabling researchers to use terrestrial GPS receivers to navigate to pulse locations with, admittedly, some unknown but constrained spatial error. This technique was used by joint Russian/US field teams in Siberia (Nelson et al., 2009b; Montesano et al., 2014b) to directly relate ground estimates of height and biomass to GLAS measurements. The benefit of this approach is that the Ståhl et al. (2011) estimators can be appropriately applied since their estimators account for both the GLAS sample variability and ground-satellite model variability. The downside is that the work is labor-intensive and, in remote areas where it is most logical to use satellite observations to extend a sparse network of ground plots, very expensive and logistically challenging.
3. An airborne lidar can be used as an intermediary to tie ground observations to satellite measurements. This approach was used in the Quebec, North American boreal, and circumpolar boreal studies noted above. The approach involves 3 phases (ground, airborne lidar, space lidar) and two sets of predictive models. During a mission, the airborne lidar is flown over existing ground plots and also over/along a subset of the existing GLAS orbital acquisitions.

Assuming that the area of interest (AOI) is stratified into cover types, then one set of models, one for each cover type, is developed that relates ground estimates of aboveground dry biomass/ha to airborne lidar measurements. These ground-air models are used, in turn, to estimate biomass on each of the GLAS shots overflowed with the same airborne lidar. A second set of stratified equations is developed that relate airborne estimates of biomass (Y) to GLAS measurements (X). Once these airborne lidar - GLAS equations are developed, then GLAS can be used as a regional sampling tool.

The current study employs option #3. Currently, there are downsides to using this relatively efficient method of inventorying forests within a hybrid estimation framework using space lidar measurements. The most important limitation is that a hybrid variance estimator does not yet exist that takes into account both the 1st phase sampling variability and the model variability inherent in a three-phase design, i.e., a design that incorporates not one set of models but two. This limitation has affected those employing option #3 for the last decade, and the net effect of this limitation is that we employ a hybrid two-phase variance estimator (Ståhl et al., 2011) that accounts for the model variance of only one of the two models used to generate GLAS estimates of biomass. Necessarily then, we underestimate the model variance component of our estimates by an unknown amount. The good news is fourfold: (1) We know that we have this problem and acknowledge it in our studies (Neigh et al., 2013; Margolis et al., 2015). (2) Empirically, we know the order of magnitude of this model variance underestimation problem and have ways to mitigate it, though the mitigation strategy is far from ideal. (3) A hybrid, three-phase variance estimator is currently being developed and validated by colleagues at the Swedish University of the Agricultural Sciences (Saarela et al., 2016; Sören Holm, personal communication). (4) Until that estimator is ready for use, we have in previous studies and will in this study employ a variance estimator that may mitigate but not correct the effects of the model variance underestimation problem. The reader should view all standard errors related to GLAS estimates reported here-in as approximations.

Ståhl et al. (2011) developed hybrid, two-phase variance estimators that take into account the uncertainties associated with both sampling error and model error. These two error sources are additive. The Ståhl et al. estimators are applicable in situations where an airborne lidar is used to sample extensive areas that may have only a limited number of ground observations on or near the AOI. Given the model-based nature of the estimators, the ground plots do not need to be probabilistically allocated. Practically speaking, this means that (1) an existing NFI is not a precondition to estimating regional biomass with an airborne lidar, and (2) ground plots do not have to be overflowed as an integral part of the lidar sample survey, i.e., they can be located off the sample flight line grid. These two characteristics free the researcher to purposefully select existing ground plots anywhere, with best effort made to collect a representative sample of forest conditions within the AOI. The Ståhl estimators can also be appropriately applied in situations where satellite lidar measurements are tied directly to ground biomass estimates, i.e., options #1 and 2 above. Similar to the approach taken by Nelson et al. (2009b) in Siberia, this allows researchers to gather ground observations of biomass on existing GLAS pulses in clusters throughout the AOI. It is important that these ground-satellite observations are representative of the entire AOI; in particular it is most important that the full range of the GLAS measurements are characterized in order to avoid, as much as possible, extrapolation.

Gregoire et al. (2016) encourage scientists who employ lidar as a regional sampling tool to make clear the survey design, the estimators employed, assumptions which underlie the sampling framework, and the limitations of inference. To this end, we describe the foundations on which our US and Mexico surveys rest. In this study, both surveys rest on a hybrid 3-phase sampling framework (1) where the 1st phase, i.e., the GLAS acquisitions, are acquired probabilistically and the 2nd (ALS) and 3rd (ground) phases are acquired within model-based

inferential frameworks, (2) where the ground-ALS observations and the ALS-satellite lidar observations are not spatially coincident, (3) where a GLAS ascending or descending orbit is considered a cluster, (4) where GLAS clusters are assumed to randomly sample CONUS, Mexico, and the states within, (5) where individual GLAS pulses within these clusters (orbits) are post-stratified, and (6) where we assume that ALS estimates of biomass density on each GLAS pulse overflowed are correct and calculated without error. This last assumption, patently untrue, leads directly to our underestimation concerns regarding biomass variance estimates.

This underestimation problem with the model variance component of the Ståhl et al. (2011) variance estimator, expected to be on the order of 2–4x given the ground-ALS and ALS-GLAS regression sample sizes in this study (Saarela et al., 2016, their case C.2 versus C.3) can be partially addressed by knowingly using an estimator of the sample variance component known to be inflationary. As noted above, the sample variance component of Ståhl's variance estimator rests on the assumption that the GLAS orbits randomly sample the AOI. Though GLAS orbits are not a true systematic sample, they have the appearance of a systematically-acquired transect sample. It is well documented that sampling variability may be overestimated when systematically-acquired samples are assumed to be randomly collected (Osborne, 1942; Nyssönen et al., 1967, 1971). For instance, Ene et al. (2012), using a simulated forest population in Hedmark County, Norway, demonstrates that the Ståhl et al. hybrid two-phase variance estimator performed well, i.e., was unbiased, when the assumptions upon which the estimator rest are met. However, when the assumption of randomness is not met, for instance when a systematic sample of flight lines is assumed to be a random sample, then significant variance inflation can occur. In Hedmark, treating a systematic sample of flight lines as a random sample lead to a variance inflation of 4.0x due in large part to an overestimation of the sampling variance in an ordered population. Hedmark, in fact, might be regarded as a worst-case since (1) there is an obvious N-S biomass gradient, and (2) the flight lines were flown E-W. This sample variance inflation is especially significant in situations where the population of interest, in our case the biomass in the state and national forests of CONUS and Mexico, exhibits spatial trends approximately perpendicular to the GLAS near-polar orbital inclination. Both the US and Mexico exhibit such E-W spatial trends in many states and at the national level. While large-scale biomass trends associated with geography and topography may inflate GLAS sample variances, a 2nd orbital characteristic may reduce this inflationary tendency. ICESat/GLAS was a repeat-track satellite; it flew the same near-polar orbits within 1 km of the reference ground tracks, 1σ , every 91 days. The repeat-track nature of the coverage occasionally lead to closely spaced, parallel orbits 100 s of meters apart across-track, and this will tend to decrease sampling variability. Further, in order to conserve laser resources, GLAS was typically tasked to acquire data 2 or 3 times a year in ~35 day blocks. Given the 91 day repeat cycle, this tasking leads to occasional random, large gaps in coverage up to ~130 km in CONUS and ~88 km in Mexico. These gaps, along with cloud cover/obscuration, can significantly affect biomass estimates and sampling variability in small states by limiting the number of GLAS orbits acquired over a particular area. In short, we expect sampling variance to be inflated, but the amount of inflation is unknown, and will vary by state.

The GLAS orbital characteristics detailed above lead to concerns about the accuracy of the biomass means and biomass totals. The near-polar GLAS orbits and the random nature of the extensive E-W gaps place states such as Vermont (4 orbits), New Hampshire (7 orbits), Rhode Island (3), New Jersey (7), and Delaware (0), at risk of estimation based on small, possibly nonrepresentative samples. Likewise, larger states that might be measured by many GLAS orbits but which support little forest (e.g., the states of the US Central Plains) where those forests are spatially clustered, may exhibit estimation instability. Certainly the GLAS orbital sampling pattern, a pattern which excludes from measurement large chunks of the Earth's terrestrial surface, is not ideal.

Near term, future space lidar multibeam profilers (GEDI, ICESat-2), with increased orbital sampling density and randomization, will mitigate this sampling problem.

3. Methods

3.1. Sampling framework and estimators

The specific estimators employed in the current study have been reported in a number of publications including Ståhl et al. (2011), Eqs. (11), (12), (14), (15); Nelson et al. (2012) Eqs. (11), (12), (14), (15), Ene et al. (2012), Eqs. (13), (14), and (15), Neigh et al. (2013), Eqs. (1)–(4), and Gregoire et al. (2016), Section 4.7, Eqs. (10)–(16). The sampling approach and regression estimation rest on a few important assumptions, some which are outlined in Ståhl et al. (2011), in standard regression textbooks, or in publications noted elsewhere in this report:

- (1) The linear regression model form is correctly specified. The models are unbiased.
- (2) All regressions are homoskedastic, i.e., regression errors are iid (independent and identically distributed).
- (3) The airborne lidar observations acquired to develop the ground-air regressions are independent of the airborne lidar observations acquired along GLAS transects. This condition does not affect our ability to purposively sample ground plots or GLAS transects with the airborne lidar (Royall, 1970; Royall and Herson, 1973) as long as the aforementioned spatial independence is respected.
- (4) Although not a true systematic sample, GLAS ascending and descending orbits are systematic-like, and we assume that these orbits, when treated statistically as a random sample, will have similar inflationary tendencies with respect to variance estimation. The degree of this inflationary tendency in CONUS and Mexico is unknown.

3.2. US and Mexico ground plot data

In CONUS, US Forest Service - Forest Inventory and Analysis (FIA) ground plots served as our ground reference data set both with respect to (1) developing the predictive models that relate aboveground dry biomass on ground plots to coincident ALS measurements, and (2) developing ground reference estimates of biomass for five land cover types in each of the 48 states in the continental US. Of the 895 FIA plots overflowed by our ALS sensor, 842 were used to develop the generic and stratum-specific equations used to inventory CONUS forests. The remaining 55 FIA plots were excluded from the model development process based on inconsistent ground biomass-ALS height relationships which suggested that the plots had been either mislocated or cut/cleared between the time of the FIA measurement and the laser overflights.

With respect to state and national ground reference estimates, USFS software (EVALIDator, www.fia.fed.us/tools-data/tutorials_training/docs/EVALIDator_user_guide_1.5.1.05.pdf) was used to estimate forest biomass in each of the five strata defined by the Landsat land cover map based on the predominant species classification of each FIA plot as determined on the ground at time of measurement. The FIA species-specific categories (e.g., longleaf-shortleaf pine, oak-hickory, maple-beech-birch, pinyon pine-juniper) were cross-walked into one of the five forest classes - wetlands, hardwood, conifer, mixedwood, and burn - in order to compare satellite lidar estimates to ground-based estimates. The EVALIDator software was used to query ground measurements on 119,414 FIA plots scattered across the entire US to develop the ground-based stratum, state, and national estimates. FIA plot remeasurement dates varied from 1998 to 2013 (mean year ± 1 standard deviation: 2004.7 ± 2.5 years), with measurement dates varying by state.

In Mexico, the Mexican NFI, *Inventario Nacional Forestal y Suelos* (INFyS, www.cnf.gob.mx:8090/snif/portal/infys), provided the necessary ground reference observations of biomass. Mexico's NFI includes ~26,000 ground plots established between 2004 and 2007 scattered across the forested areas of the country. Our ALS system measured 294 INFyS ground plots in northern and central dry forests, in the Sierra Madre pine/oak forests, in the Veracruz and Yucatan moist forests, and in the Yucatan dry forests.

3.3. US and Mexico land cover maps used for stratification and reporting

Landsat and MODIS data were used to categorize both ground plot data and GLAS data overflowed by the ALS into five forest cover strata in CONUS. A 30 m Landsat land cover map of CONUS, resampled to 90 m, the NLCD (National Land Cover Database) - 2006 (www.mrlc.gov/nlcd2006.php), was used to identify each FIA ground plot and each CONUS GLAS pulse as being in one of four land cover categories: wetlands (NLCD class 90,95), hardwood (NLCD class 41), conifer (NLCD 42), and mixedwood (NLCD 43). Wickham et al. (2013) state that the NLCD-2006 map used in this study to stratify the US has a classification accuracy of 78%. A fifth land cover class - burn - was identified by joining FIA plot locations and GLAS pulse locations with 2000–2006 MODIS burn data (Giglio et al., 2009). This MODIS multitemporal burn product was used to report the most recent burn date of a given 500 m pixel identified as burned within the seven year window.

Similarly in Mexico, a 90 m land cover map produced by INEGI (*Instituto Nacional de Estadística y Geografía*, INEGI, 2010, <http://www.inegi.org.mx>), Version 4, based on photointerpretation of SPOT data, with 59 land cover classes, was recoded into the 4 NLCD-like forest cover classes of interest - wetlands, hardwood, conifer, and mixedwood. The actual recoding, i.e., crosswalk, used in this study is provided in the Appendix, Table S.1. This land cover map of Mexico was joined with both the Mexican NFI ground data and the GLAS acquisitions. As in the US, the MODIS multitemporal burn data was used to identify areas that burned between 2000 and 2006, thereby adding a fifth forest land cover class.

Of the 842 US ground plots, 66 were located in NLCD nonforest, 103 in shrubs, 39 - wetlands, 181 - hardwood, 404 - conifer, 38 - mixedwood, and 11 - burn. Given the paucity of burned plots overflowed by the ALS sensor, the nonforest and shrub plots were used along with the 11 burn plots to derive the FIA-ALS biomass equation for burn ($n = 180$). In Mexico, of these 294 NFI plots overflowed by the ALS sensor, 33 were identified as nonforest based on our simplified INEGI crosswalk, 0 in shrubs, 0 in wetlands, 128 in hardwood, 120 in conifer, 0 in mixedwood, and 13 in burn. In Mexico, as in the US, land cover strata not identified as one of the five forest cover types are assumed to contain zero forest biomass regardless of the GLAS measurements obtained.

3.4. Airborne LiDAR scanning data

NASA's GLiHT (Goddard's Lidar, Hyperspectral, & Thermal imager, Cook et al., 2013, <http://gliht.gsfc.nasa.gov>) ALS system was used to acquire laser ranging measurements over the US and Mexico in 2011, 2012, and 2013 (Table 1). The heart of the ranging system is a Riegl VQ-480, a 300 kHz, 1550 nm Class 1 (eye-safe) scanning lidar. The nominal flight altitude was 335 m AGL and the usable scan angle was defined to be $\pm 15^\circ$ for these overflights, yielding an effective ground swath width of 180 m, though the laser does measure out to $\pm 30^\circ$. The flight speed varied with topography but was typically 110–120 knots (205–223 km/h, 57–62 m/s). At a 335 m AGL flight altitude, given a laser beam divergence of 0.3 mrad, the footprint size is 10 cm. The GLiHT lidar is a multistop system, recording up to 6 significant returns per pulse, though the great majority of the pulses acquired in the US and Mexico generated 1 or 2, rarely 3 returns. On average, the pulse density on all 3 missions was 4–5 pulses/m².

Table 1
GLiHT acquisitions over CONUS FIA ground plots, Mexican NFI ground plots, and GLAS acquisitions.

Location	Dates (flight days)	Number of ground plots	length (km) of GLAS transects	Cost(\$)
Eastern US	Aug 2-Sept 17, 2011 (15)	348	3150	61,754 ^a
Western US	Aug 25-Sept 8, 2012 (13)	518	3647	39,155 ^b
Mexico	Apr 13-May 10, 2013 (22)	294	3160	50,380 ^b

^a NASA Cessna 206.

^b Commercial aircraft, Piper Cherokee PA-32.

The US and Mexico flights were scheduled to try to measure canopies leaf-on. In CONUS, flights were conducted mid-summer - July and August. In Mexico, scheduling was more challenging given that leaf-on/off conditions vary with latitude, elevation, and the timing of the regional rainy/dry seasons. All of Mexico was flown in April and early May 2013 in an attempt to avoid clouds associated with the rainy season in southern Mexico (May–September) while also avoiding the leaf-drop associated with the dry season (November–March). Maass et al. (1995, Fig. 5) provides some guidance for the state of Jalisco in western Mexico (latitude $\sim 20^\circ\text{N}$), showing maximum leaf drop in November and December. In Mexico, we were not successful as regards acquiring consistently leaf-on ranging measurements. In southern Mexico, especially in the pine/oak and dry forests of Oaxaca ($\sim 17^\circ\text{N}$) and in the dry forests of the northwestern Yucatan peninsula ($\sim 20.5^\circ\text{N}$), extensive stands of hardwoods were leaf-off, and these leaf-off stands were frequently just a few kilometers from similar hardwood stands which, from the aircraft, looked to be in full canopy. In hindsight, it may have been better to fly Mexico perhaps 20–30 days later, trading increased cloud cover for leaf flush. Given GLiHT's relatively low design flight altitude, data collection missions can be conducted under overcast skies. However, given the 1550 nm wavelength of the ALS transmitted pulse and concerns about pulse attenuation/absorption due to water, care was taken not to fly wet targets.

The GLiHT aircraft transited 895 USFS-FIA plots in the United States and 294 INFyS plots in Mexico. GLiHT also transited 6797 km of GLAS orbits in the continental US and 3160 km of GLAS orbits in Mexico. The FIA and INFyS ground plots and GLAS transects overflown by GLiHT in the US and Mexico are shown in Fig. 1, as well as the GLAS orbits used to sample both countries.

3.5. GLAS data

GLAS served as our large-scale sampling tool; the Landsat land cover and MODIS fire products in the US and the SPOT-photointerpreted and MODIS fire products in Mexico served only to stratify the two countries for reporting purposes and, arguably, to improve the predictive biomass models. In the US, 3 leaf-on GLAS acquisitions incorporating 230 orbits are used to estimate biomass in 47 of the 48 states. In the US, none of the 230 GLAS orbits transected Delaware nor the District of Columbia, hence biomass estimates, like the sample size, are zero for these two locales. In Mexico, 5 GLAS acquisitions were accessed to inventory the 32 states and the entire country. Three-hundred-eighty-four orbits were acquired over Mexico and each state was measured during at least one of those acquisitions (Table 2).

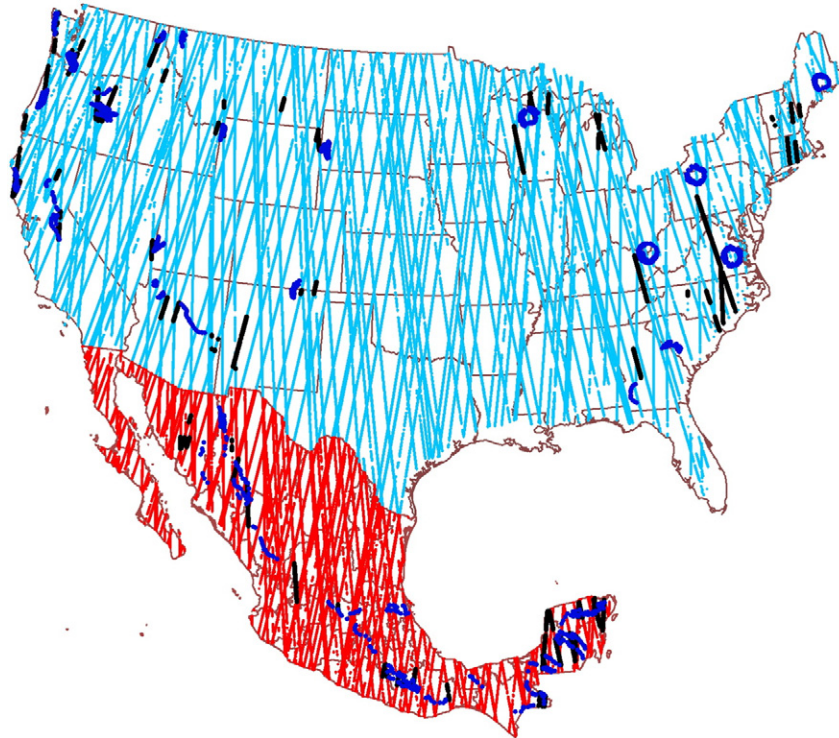


Fig. 1. GLAS transects, ALS coverage of GLAS segments, and ground plots overflown by the ALS in the continental US and Mexico. The light-blue lines are the 230 L2a, L3c, and L3f GLAS acquisitions used to sample the US. The red lines are the 384 L2a, L3a, L3c, L3d, and L3f GLAS acquisitions used to sample Mexico. Black lines are ALS overflights of selected GLAS orbits. Dark blue points are ALS overflights of FIA ground plots (US) or NFI ground plots (Mexico).

Table 2
Description of GLAS data used to inventory the continental US and Mexico. Acquisition dates are truncated in an attempt to select only leaf-on data.

Acquisition	Dates	#Orbits	#Pulses
United States			
L2a	Sept 24–Sept 30, 2003	24	103,620
L3c	June 1–June 23, 2005	94	341,421
L3f	June 1–June 25, 2006	112	494,007
All Acq.		230	939,048
Mexico			
L2a	Sept 24–Nov 17, 2003	111	153,580
L3a	Oct 3–Nov 8, 2004	85	118,409
L3c	June 1–June 23, 2005	47	84,639
L3d	Oct 21–Nov 24, 2005	81	145,495
L3f	June 1–June 25, 2006	60	97,262
All Acq.		384	599,385

3.6. Ancillary data

Ninety meter SRTM (Shuttle Radar Topography Mission) data acquired in February 2000 were used to approximate local slope on a 3×3 window centered on GLiHT-GLAS observations. Only GLAS pulses acquired on slopes $\leq 20^\circ$ were used to (1) formulate regressions and (2) to inventory US and Mexico forests. This slope criterion was established to mitigate the effects of GLAS waveform pulse spreading on steep slopes (Lefsky et al., 2005) which leads to inflated canopy height values.

3.7. Regression analysis

The LIN, LNI, and log-linear models were developed using SAS Proc Reg (SAS Institute, Inc., 1989). The Reg procedure's All Possible Subsets Regression (APSR) feature was used to select significant variables for each model. APSR considers all possible variable combinations and reports the best single variable models, the best two-variable models, three-variable models, etc., according to a user-specified criterion, in our case, R^2 . These best models are then tested individually and down-selected based on variable significance and VIF (variance inflation factor) metrics. VIF values approaching or exceeding 10 indicate multicollinearity issues and identify those independent variables that should be trimmed in order to avoid overfitting. The log-linear model estimates of $\ln(\text{biomass})$ were backtransformed to real-world units, i.e., biomass density, in t/ha, using a ratio estimator for bias correction (Snowdon, 1991).

3.8. Temporal concerns

As reported above, this study integrates ground and GLAS data sets acquired circa 2005 (ca. 2005) and ALS data acquired ca. 2012. The ALS is used to tie GLAS to ground. The ground-ALS models relate ca. 2005 ground biomass density estimates (Y) to ca. 2012 ALS measurements (X). The ca. 2012 ALS measurement are then used to estimate ca. 2005 biomass conditions on the ca. 2005 GLAS pulses overflowed for ALS-GLAS model development. No growth or mortality adjustments were made to the ground plot data to harmonize dates. The important factor here is that the ground data collects were, at least approximately, temporally coincident with the GLAS acquisitions.

4. Results

4.1. Continental United States

In CONUS, three sets of explicitly or implicitly linear models, (1) linear (LIN), (2) linear-no-intercept (LNI), and (3) log-linear, were used to develop stratum-within-state, state, stratum, and national-level estimates of aboveground total dry biomass for the contiguous US.

A summary of CONUS LIN and LNI models are given in Table 3. An expanded version of Table 3 that includes the specific regression coefficients and the variance-covariance matrices can be found in the Appendix, Table S.2.

The third set of models, log-linear, i.e., $\ln(b_{\text{ground}}) = f(ht, den)_{\text{ALS}}$ and $\ln(\hat{b}_{\text{ALS}}) = f(ht, den)_{\text{GLAS}}$ were likewise formulated, but these equations produced state and national estimates that grossly overestimated the ground-based FIA ground reference values by factors of 2–3. Though log-linear ground-ALS models have been used successfully in the predominantly coniferous Scandinavian countries to estimate volume (Næsset, 1997) and biomass (Næsset et al., 2011), airborne-GLAS log-linear models have consistently failed to produce reasonable biomass estimates in Quebec (Boudreau et al., 2008; Nelson et al., 2009a) and now, in CONUS. As illustrated in Fig. 2, the failure is broad-spectrum. Fig. 2 illustrates comparable GLAS biomass estimates from the three models made on 241,718 US GLAS pulses located in the 5 forest strata in CONUS. For whatever reason, e.g., inappropriate model form, sparse, nonrepresentative ALS data collects over ground plots and GLAS pulses, noisy GLAS data - the log-linear models turn out to be much less forgiving than the linear model forms, especially at the high end of the AGB spectrum. As such, the log-linear equations and resultant estimates derived based on those equations are not reported.

The LIN and LNI model results were compared to USFS-FIA ground-based estimates of biomass density at the state and national levels. Table 4 summarizes the results of the national comparison. The GLAS results are based on analysis of those 241,718 GLAS pulses acquired along 116 descending and 114 ascending ICESat/GLAS orbits over CONUS during acquisition periods L2a (2003), L3c (2005), and L3f (2006). The USFS-FIA biomass estimates are based on analysis of 119,414 CONUS FIA plots measured between 1998 and 2013. By all metrics presented in Table 4, the linear-no-intercept models were better at reproducing FIA results which, in this study, serve as our ground reference data. The linear models produce a national total biomass estimate 7% larger than the FIA estimate whereas the linear-no-intercept models reproduced the national FIA estimate within <1%. State-level LNI estimates also exhibited less scatter about the 1:1 line that denotes equivalence between FIA and GLAS biomass estimates. The average absolute value of the departure from the 1:1 line at the state level was 19%; the comparable linear model departure was over 30%. Based on the results presented in Table 4, the decision was made to inventory Mexico using LNI models only.

Both the GLAS LIN and LNI estimates of national biomass density were greater than the FIA average biomass density, however the NLCD-2006 map used to stratify the US GLAS pulses into forest and nonforest reported less forest area relative to the FIA. The FIA reports that the US is 34.5% forested; the NLCD-2006 Landsat map reports that the US is 31.4% forested if we judge that all pixels identified as wetland, hardwood, conifer, mixedwood, and burn are forest. Fig. 3 illustrates state-to-state differences between FIA and NLCD forest area estimates.

Tabular estimates of total aboveground dry biomass based on LNI equations, by state, are provided in Table 5. The ground-based USFS-FIA estimates are provided for comparison. Ground-GLAS state comparisons of biomass density and total aboveground dry biomass are illustrated in Fig. 4.

4.2. Mexico

Based on the US results, Mexico was inventoried using linear-no-intercept models. Unlike the US, we did not inventory Mexico using separate equations for wetland, hardwood, conifer, mixedwood, and burn forest. Two significant characteristics of the Mexican data set suggested that a different analysis approach in Mexico was needed. The first characteristic was that the distribution of forest cover types flown by the ALS varied significantly from north to south, affecting our ability to generate cover-type-specific ground-ALS equations throughout the country.

Table 3

CONUS linear models and linear-no-intercept models used to predict AGB for each of the 5 forest strata within US states. Note: While the linear versus linear-no-intercept ground-ALS models may be directly compared within cover type, the ALS-GLAS models may not. Within cover type, the linear and linear-no-intercept Ground-ALS models use the same dependent (b_{grnd}) observations; the ALS-GLAS models do not. VIF = variance inflation factor.

Variable definitions:						
<i>Ground-ALS variables:</i>						
Variable	Description					
b_{grnd}	Plot-level FIA (in US) or INFyS (in Mexico) estimate of total aboveground dry biomass per hectare (t/ha).					
\bar{h}_a	Average height, all pulses (m).					
\bar{h}_c	Average height, canopy hits only (m) where a canopy hit = any return > 1.3 m above ground level (AGL).					
cc	Canopy closure, the fraction of first returns intercepted by a tree, i.e., first return height > 1.3 m AGL ($0 < cc < 1.0$).					
$h10_a, h30_a, h50_a, h70_a, h90_a$	Height deciles, all pulse returns, i.e., height below which 10% of the returns occur, 30% occur, etc.; ground hits included (m).					
$d0_a, d1_a, d2_a, d4_a, d5_a, d7_a$	Density deciles, all pulse returns, i.e., the proportion of returns that occur in equally spaced height bins between ground and $h95_a$, $\sum_{x=0}^9 dx_a = 1.0$.					
<i>ALS-GLAS variables:</i>						
\hat{b}_{ALS}	ALS estimate of biomass density (t/ha) based on ground-ALS equation.					
centroid	Height of the GLAS waveform where half of the waveform energy is above and half below, signal start to signal end (m), relative to signal end.					
medh	Median waveform height above ground peak (m).					
h14	The height from ground, i.e., the last Gaussian peak, to signal start (m).					
h10, h25, h30, h75, h90	Height deciles or quartiles; e.g., h75 is that waveform height below which 75% of the waveform energy resides (m), relative to signal end.					
$h75_{\text{srtm}}, h90_{\text{srtm}}$	Height deciles corrected for SRTM ground slope (m).					
trail	Vertical distance, ground peak to the end of the Gaussian ground signal (m).					
lead	Vertical distance, start of the Gaussian ground signal to the ground peak (m).					
fslope	Front slope angle, that acute angle formed by a vertical line denoting ambient waveform noise and a 2nd line joining signal-start and the first (highest) vegetation peak (degrees). See Boudreau et al. (2008), Fig. 1.					
elev	Elevation of waveform ground return relative to Topex/Poseidon geoid (m).					
acq2	US only, dummy variable, i.e., 0/1, identifying L3f pulses.					
acq3	Mexico only, dummy variable identifying L3d pulses.					
ct2	Mexico only, dummy variable identifying pulses intercepting conifer.					
Ground - ALS models: $b_{\text{grnd}} = f(\text{ALS variables})$						
Stratum	Model	R ²	RMSE (t/ha)	Largest VIF	n	Significant ALS variables
Wetland	LIN	0.75	36.53	3.86	39	$h10_a, h50_a, d0_a$
	LNI	0.73	36.82	1.54	39	$\bar{h}_a, d1_a$
Hardwood	LIN	0.57	44.49	2.47	181	$\bar{h}_a, h10_a, d7_a$
	LNI	0.54	45.53	3.73	181	$h70_a, d7_a$
Conifer	LIN	0.72	69.29	2.39	404	$\bar{h}_a, d5_a$
	LNI	0.72	69.32	5.29	404	$\bar{h}_a, d5_a$
Mixedwood	LIN	0.71	50.00	NA	38	\bar{h}_a
	LNI	0.71	49.60	NA	38	\bar{h}_a
Burn	LIN	0.56	39.54	2.87	180	$h70_a, cc$
	LNI	0.56	39.55	1.42	180	$h70_a, d2_c$
ALS - GLAS models: $\hat{b}_{\text{ALS}} = f(\text{GLAS variables})$						
Stratum	Model	R ²	RMSE (t/ha)	Largest VIF	n	Significant GLAS variables
Wetlands	LIN	0.50	47.85	4.47	1037	centroid, elev, $h75_{\text{srtm}}$
	LNI	0.59	42.43	5.55	1037	medh, $h90_{\text{srtm}}$
Hardwood	LIN	0.47	38.69	8.95	3996	$h75_{\text{srtm}}, h30, h10, medh, trail$
	LNI	0.42	41.96	4.20	3996	$h75, h10, trail$
Conifer	LIN	0.60	60.73	3.54	5580	$h75, medh, h14, elev$
	LNI	0.57	65.10	5.64	5580	$h75_{\text{srtm}}, h14, elev$
Mixedwood	LIN	0.43	39.58	5.58	817	centroid, h10, h50, elev, acq2
	LNI	0.35	40.81	4.48	817	$h75, h10, lead, trail$
Burn	LIN	0.68	19.45	3.54	35	$h90, h75_{\text{srtm}}$
	LNI	0.75	16.85	2.39	35	$h90, h75_{\text{srtm}}$

Sixty-four of the 79 Mexican NFI plots overflow in the 12 northern states were pine; only 6 were hardwood, and none were mixedwood. Northern Mexico (excluding Veracruz and southern Tamaulipas), primarily steppe/desert, tends to support dry pine or pine-oak forests. In this northern zone, one general NFI-ALS equation was generated. In central and southwestern Mexico, states on or west of the Sierra Madre support mountain pine-oak forests and to the west, the dry Pacific forests. In this southwest zone, two NFI-ALS biomass equations were generated, one for hardwood, one for conifer. The third zone, the Atlantic Crescent states which include Veracruz, Tabasco, Campeche, Yucatan, and Quintana Roo, are a mixture of moist and dry hardwood forests with little conifer. Of the 112 NFI plots transited in this third zone, 86 were hardwood, 2 were conifer. One general NFI-ALS equation was generated for this Atlantic Crescent zone, primarily characterizing the moist and dry tropical hardwood cover types.

The second characteristic that suggested that Mexico should be analyzed as three independent zones - Northern, Southwest, Atlantic Crescent - concerned the Atlantic Crescent data specifically. The problem is illustrated in Fig. 5, a scatterplot of GLAS $h90$ (X) versus \hat{b}_{ALS} (Y) for the entire country, GLAS acquisitions L2a, L3a, L3c, L3d, and L3f. The bi-lobed nature of the data cloud is obvious and troublesome from the standpoint of fitting linear models; at a minimum, at least two distinct subpopulations are seen in this plot.

Additional work indicated that the near-vertical lobe closest to the Y-axis was tied to the L2a, L3a, and L3d GLAS acquisitions, i.e., the autumn measurement periods, in the Atlantic Crescent. We hypothesize that Fig. 5 illustrates the effects of dry season deciduous leaf drop on the upper GLAS height deciles, e.g., $h70, h80, h90$. Based on this hypothesis, a decision was made to inventory the five Atlantic Crescent states using early summer GLAS acquisitions only, i.e., L3c and L3 f.

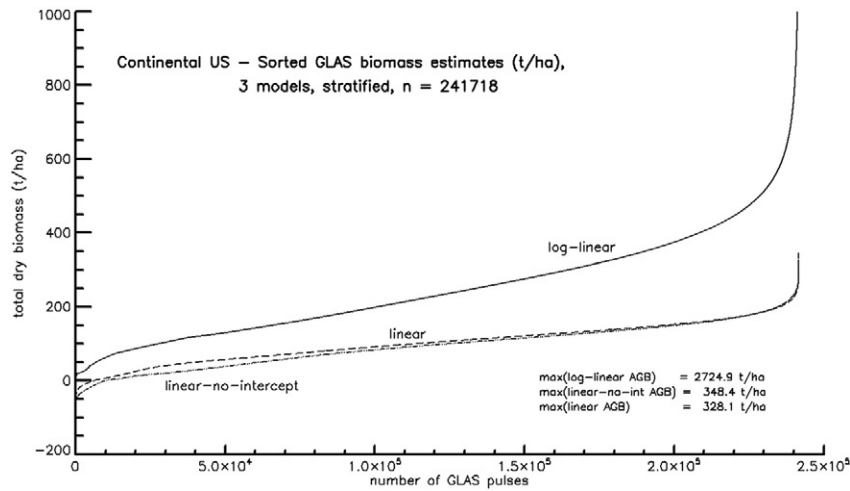


Fig. 2. A comparison of the range of aboveground dry biomass estimates generated using (1) stratified linear, (2) stratified linear-no-intercept, and stratified log-linear models for 241,718 GLAS pulses in CONUS.

Table 4
USFS-FIA and GLAS linear (LIN) and linear-no-intercept (LNI) estimates of aboveground dry biomass density and total aboveground dry biomass, with FIA - GLAS percent differences at the national and state level.

	AGB density (t/ha)	SE (t/ha)	Forest area (km ²)	Total biomass (Gt)	SE (Gt)	GLAS/FIA		
						CONUS tot biom, % diff	Avg state tot biom, % diff	Avg state tot biom, [%diff]
USFS-FIA	88.21	0.23	2,739,384	24.165	0.063	–	–	–
GLAS - LIN	103.99	1.98	2,487,904	25.871	0.492	7.06	20.74	30.83
GLAS - LNI	96.25	1.81	2,487,904	23.946	0.451	-0.91	6.96	18.82

The other two zones were inventoried using all five GLAS acquisitions, none of which exhibited this distinct bi-lobe pattern. The Mexico linear-no-intercept models used to estimate stratum-within-state and national biomass are given in Table 6. An expanded version of Table 6 that includes the specific regression coefficients and the variance-covariance matrices can be found in the Appendix, Table S.3.

Table 7 reports the individual state estimates of biomass density, with error bounds, derived using the LNI equations reported in Table 6. Table S.5 in the Appendix reports stratum-within-state biomass densities for Mexico. As expected, forest biomass density increases as one moves from northwest to southeast in Mexico, with the Baja states reporting the lowest densities (and forest areas) and the Atlantic Crescent states, Veracruz and Quintana Roo, reporting the highest biomass densities.

At the national level, a number of Mexico forest biomass estimates can be compared to this study's national estimate (Table 8). The FAO

estimates for 2000, 2005, and 2010 (FAO, 2010a) and the various biomass estimates reported by Cartus et al. (2014, p. 5573) were derived by multiplying their aboveground carbon estimates by 2 to calculate dry biomass, an inversion of the procedure employed by FAO to convert biomass estimates to carbon. A few observations can be made regarding the numbers reported in Table 8. First, as witnessed by the FAO estimates, Mexican forests are under pressure much more so than forest area and forest growing stocks in the US. In fact, between 2000 and 2010, Mexican forest area losses were offset by forest area gains in the US (FAO, 2010a, p. 19). As noted in Table 8, our national estimate for Mexico, developed based on GLAS observations acquired between 2003 and 2006, are bracketed by the FAO ground-based and the PALSAR results presented by Cartus et al. (2014).

5. Discussion

5.1. Continental United States

In a sampling study such as this where the NLCD is used to post-stratify the FIA plots overflowed by the ALS and all of the GLAS pulses, there are two moving parts: (1) an estimate of biomass density (t/ha) derived from all of the GLAS pulses that intercept a given forest cover type as characterized by the NLCD, and (2) the optical satellite/GIS-based areal estimate of that forest cover type. Fig. 4A illustrates the significant differences between FIA and GLAS estimates of biomass density. Fig. 3 illustrates the significant differences between FIA and NLCD-2006 forest area estimates. Given the fact that the FIA and GLAS national estimates are so close - LNI model results within 1%, LIN model results within 8% - how do such disparate state components produce such national agreement? Texas, the largest outlier in Fig. 3, serves as an informative example. As shown in Fig. 3, the area covered by Texas forests according to the FIA is 249,127 km².

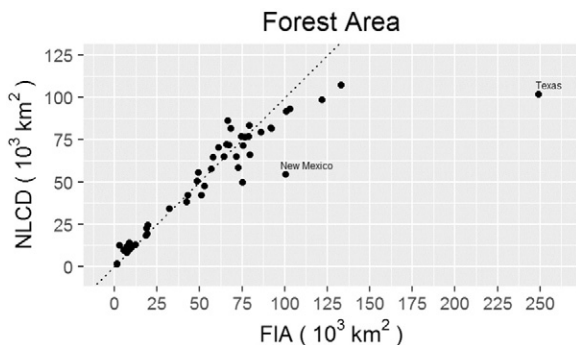


Fig. 3. FIA forest area versus NLCD forest area, by US state.

Table 5

CONUS state and national estimates of total aboveground dry biomass developed using linear-no-intercept ALS-GLAS equations in Table 3. Tot. SE = $\sqrt{\text{sample SE}^2 + \text{model SE}^2}$.

	USFS-FIA est.		GLAS estimates, biomass density							
	Biomass density		NLCD area (km ²)		GLAS Mean	#GLAS Shots	Sample SE (t/ha)	Model SE (t/ha)	Tot. SE (t/ha)	%model error
	AGB (t/ha)	SE (t/ha)	Nonforest	Forest						
Alabama	83.69	0.99	52,891	81,551	113.41	9711	2.06	0.46	2.11	4.74
Arizona	33.32	0.80	244,555	49,758	35.00	7458	4.01	1.15	4.17	7.55
Arkansas	89.81	1.10	60,696	76,825	103.70	9105	4.29	0.42	4.31	0.97
California	135.00	1.70	300,088	107,363	108.41	10,453	5.27	0.88	5.34	2.71
Colorado	65.98	1.17	187,932	82,004	57.95	7468	3.37	1.18	3.57	10.96
Connecticut	155.80	6.19	4612	8208	117.42	430	4.74	0.52	4.77	1.20
Delaware	142.69	10.00	3308	1758	0.00	0	0.00	0.00	0.00	0.00
Florida	70.29	1.24	64,460	81,743	69.81	3212	4.61	0.68	4.66	2.12
Georgia	89.61	0.98	60,633	91,840	110.80	9314	2.79	0.49	2.83	3.04
Idaho	87.92	2.01	136,366	79,604	95.40	5015	5.76	1.03	5.86	3.12
Illinois	107.90	2.38	125,578	24,449	98.70	2945	1.57	0.50	1.65	9.24
Indiana	118.87	2.41	71,999	22,691	115.17	2972	4.24	0.57	4.28	1.80
Iowa	84.21	2.76	132,228	12,833	80.41	1532	4.16	0.45	4.18	1.16
Kansas	80.90	3.49	197,587	14,231	43.03	1509	1.91	0.30	1.93	2.39
Kentucky	113.59	1.68	48,581	55,507	122.90	5935	3.90	0.61	3.95	2.38
Louisiana	88.28	1.54	54,292	64,754	121.98	4177	2.00	0.93	2.21	17.78
Maine	83.73	0.83	18,303	64,927	95.97	5291	2.96	0.53	3.00	3.12
Maryland	163.09	7.06	14,246	11,298	135.33	619	6.94	0.64	6.97	0.85
Massachusetts	150.39	4.15	7994	13,039	118.44	1075	6.45	0.50	6.47	0.59
Michigan	89.93	0.63	166,327	83,279	99.73	10,592	1.93	0.49	1.99	5.98
Minnesota	60.84	0.65	138,010	86,183	71.94	9054	1.94	0.47	1.99	5.66
Mississippi	86.51	1.12	58,050	65,973	112.54	7556	2.74	0.52	2.78	3.50
Missouri	88.92	0.85	109,712	70,552	104.87	9818	3.25	0.54	3.30	2.67
Montana	72.92	1.29	286,832	93,087	81.53	8472	3.99	0.96	4.10	5.49
Nebraska	70.80	4.64	190,191	9648	42.57	1127	1.68	0.43	1.73	6.22
Nevada	23.09	0.59	244,262	42,157	32.88	5435	3.52	1.18	3.71	10.07
New Hampshire	128.26	2.54	4473	19,635	117.68	2592	6.28	0.60	6.30	0.91
New Jersey	119.56	5.58	10,046	9647	111.15	874	5.64	0.59	5.67	1.09
New Mexico	28.71	0.59	260,593	54,589	37.81	5243	4.03	1.38	4.26	10.55
New York	127.60	1.45	59,496	76,698	112.72	6990	3.45	0.46	3.48	1.72
North Carolina	108.72	1.32	56,265	71,589	119.33	4628	3.51	0.49	3.54	1.93
North Dakota	56.99	4.87	169,564	12,444	29.72	1332	1.29	0.39	1.34	8.32
Ohio	130.43	2.22	81,701	34,318	115.40	3657	2.40	0.59	2.47	5.62
Oklahoma	50.99	1.35	138,742	42,404	78.50	5977	3.95	0.39	3.97	0.98
Oregon	147.12	1.66	152,050	98,783	107.23	8034	8.81	0.82	8.85	0.86
Pennsylvania	134.50	1.37	47,284	71,849	124.08	4839	3.24	0.57	3.29	3.04
Rhode Island	137.87	8.29	1136	1595	123.81	281	4.47	0.56	4.51	1.56
South Carolina	94.52	1.41	32,900	47,629	113.57	3113	2.74	0.58	2.81	4.32
South Dakota	56.54	2.40	186,359	11,949	45.01	939	2.87	0.35	2.89	1.48
Tennessee	120.55	1.48	51,032	57,923	123.72	7137	2.30	0.57	2.37	5.70
Texas	31.56	0.47	583,193	101,835	78.28	13,506	4.60	0.40	4.62	0.75
Utah	38.56	0.78	161,162	58,384	48.94	6913	2.18	1.04	2.42	18.66
Vermont	128.72	2.59	6087	18,454	119.49	467	5.18	0.53	5.20	1.02
Virginia	120.30	1.42	38,797	64,970	125.51	3699	3.65	0.53	3.69	2.05
Washington	169.65	2.43	96,890	76,899	128.72	4931	5.84	0.82	5.90	1.95
West Virginia	143.88	2.27	11,795	50,678	137.10	2755	3.97	0.73	4.04	3.30
Wisconsin	80.73	0.67	97,154	72,216	97.33	8956	3.20	0.46	3.23	2.06
Wyoming	59.20	3.12	215,782	38,153	54.68	4580	2.67	1.20	2.93	16.86
All 48 states	88.21	0.23	5,442,232	2,487,904	96.25	241,718	1.76	0.44	1.81	5.95

The corresponding NLCD estimate of the area of the five forest strata is 101,835 km², a –59% difference $((1 - 101,835/249,127) * 100)$. The FIA biomass density in Texas is 31.56 ± 0.47 t/ha; the corresponding GLAS estimate is 78.28 ± 4.62 t/ha, a 248% difference $((78.28/31.56) * 100)$. Multiply each biomass density by the area that supports that density to get total biomass - FIA: 786.2 ± 11.7 Mt, NLCD/GLAS: 797.2 ± 47.0 Mt, a difference of 1.4%, a point in Fig. 4B very close to the 1:1 line.

The scatter about the 1:1 line in Fig. 4B illustrates an average FIA-NLCD/GLAS biomass difference of 18.8% at the state level. These large biomass differences are primarily due to the factors listed directly above, i.e., the interplay between forest area (and the associated differences in the definition of forest) and forest biomass density. Perhaps as important is the fact that the estimator used to calculate the mean biomass density of each stratum within state assumes that the area is randomly sampled. As discussed previously, GLAS does not sample the Earth's surface randomly; large areas between

orbits literally have no chance of being sampled. This characteristic, coupled with the small orbital, i.e., cluster, sample sizes associated with small states or states with few forest resources, can result in nonrepresentative samples which poorly estimate mean biomass density or total biomass. Though an imperfect sampling tool, GLAS does permit analysts to hone large-area sampling techniques, and future space lidars will mitigate these GLAS weaknesses by greatly increasing orbital sampling density.

Noted in Fig. 4B are the outliers on the far right-hand side of the graph, 3 points representing California, Oregon, and Washington. In these three states, FIA estimates of forest area are 4–24% larger than the corresponding NLCD areas and FIA biomass densities are 25–35% larger than GLAS estimates. These states support the largest trees in the conterminous US, with the tallest trees reaching heights of almost 100 m. The GLiHT data used in this study were post-processed with a maximum upper height limit of 50 m; any pulse with a first return height-above-ground of >50 m was discarded.

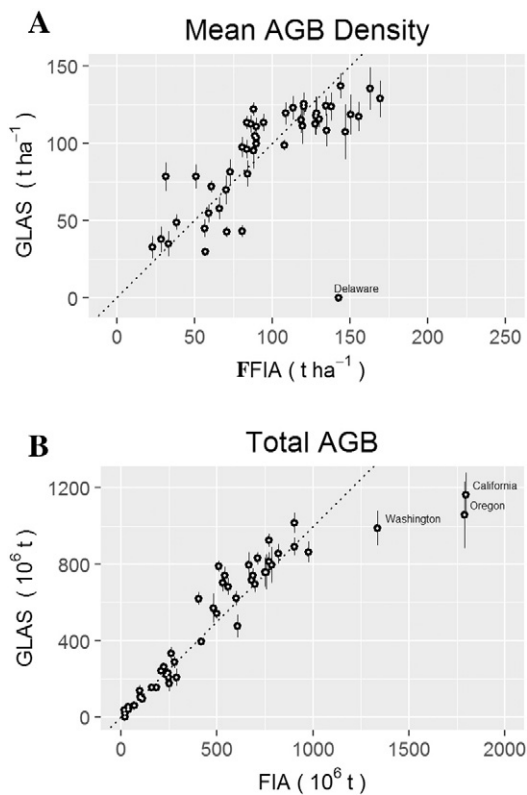


Fig. 4. A comparison of 4A) aboveground biomass density, t/ha, and 4B) total aboveground biomass, million tons, for each of the 48 states in CONUS. These results were developed using the linear-no-intercept models in Table 3. 2σ error bars ($\sim 95\%$ confidence interval) are shown on each GLAS estimate.

This was done nationally to mitigate the effects of spurious, above-canopy returns, but in these three western states this may have led to legitimate ALS measurements being ignored. This truncation was expected to have, at most, only a minor effect, an effect worth suffering to improve overall data quality nationally given that trees exceeding 50 m on FIA plots can be considered rare events. More likely, we expect that the GLAS underestimation problem on the west coast is due to the fact that certain areas of the US such as the west coast should be considered separately with respect to the development of predictive ALS - GLAS equations. Subsequent studies

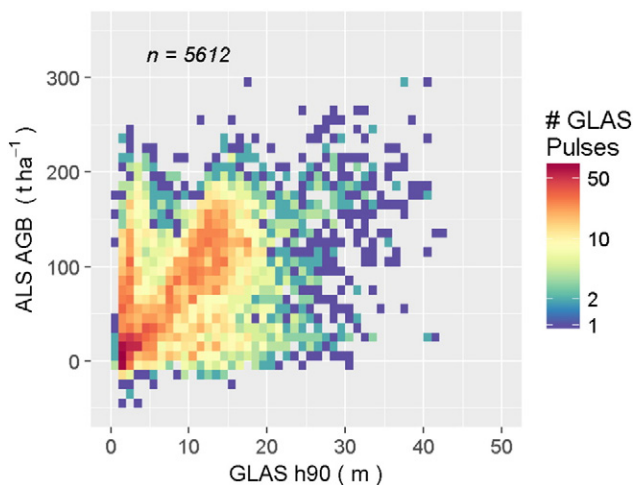


Fig. 5. A scatterplot of GLAS h_{90} versus ALS estimates of total aboveground dry biomass density (agb4) for Mexico hardwoods (entire country), GLAS acquisitions L2a, L3a, L3c, L3d, and L3f, $n = 5612$ GLAS pulses.

should consider (1) increasing the maximum allowable height threshold and (2) developing an independent set of stratified equations in, at a minimum, these three states.

Blackard et al. (2008), mapping biomass via imputation using satellite optical (MODIS, Landsat-NLCD), topographic, and climatic data sets, ran into a similar problem with these three western US states despite developing predictive equations by ecoregions. They report large mapped biomass underestimation errors relative to their FIA estimates (their Fig. 4). They underestimate total biomass in California, Oregon, and Washington by approximately 43, 35, and 41%, respectively (interpolated from their Fig. 4). Our corresponding GLAS-based estimates under-report FIA in these same states by 35, 41, and 26%, respectively. We hypothesize that the consistent underestimation in these three states in both studies is due to an inadequate characterization, measurement, and modeling of the exceptionally large trees found on the west side of the Coast Range and in the California Sierra Nevada Mountains. ALS overflights should target ground plots and satellite lidar profiles in 'big tree' areas, e.g., redwood, Sequoia, and fir/hemlock associations, so that these tall, high-biomass cover types are adequately modeled. In the current study's model-based inferential framework of the 2nd and 3rd sampling phases, such purposeful, targeted sampling is not only allowed, it would be encouraged to facilitate model development.

Stratum-level biomass density estimates for each state in CONUS are provided in the Appendix, Table S.4. In considering these stratum-within-state estimates, note that standard errors increase, reflecting the fact that the reliability and most likely the accuracy of these estimates become more suspect since national-level predictive equations are used to characterize local conditions (e.g., Huang et al., 2015). Also note that there is a general, albeit noisy, tendency for the percent-model error to decrease as a percentage of the total standard error as one drills down (spatially) into the data. This phenomenon is due to the fact that, as larger and larger areas are considered, the sampling error component decreases since more GLAS orbits hit the target. Model error tends to be relatively stable regardless of the spatial extent of the AOI. (Nelson et al., 2012, their Fig. 3). In the context of the logistic modeling of Landsat spectral data to predict forest proportion on circular AOIs 5, 10, and 15 km in radius, the stability of model standard error relative to sampling error is obvious in Tables 1–3 or Tables 4–6 in McRoberts (2006).

In this study, the continental US was specifically targeted as our primary AOI because of the existence of a country-wide, well-documented, well-measured national forest inventory, the USFS-FIA. The FIA provided ground reference, allowing us to compare our model-based, three-phase estimates of biomass density and our NLCD estimates of forest area to those of an independent, probabilistically allocated, ground-based inventory program. The analysis techniques that produced the most accurate estimates in CONUS were subsequently extended to the Mexican neotropical and subtropical forests.

The total variance of an estimate of biomass, whether it is mean density or total, additively combines (a) a sampling variance component that is a function of the number of orbits intercepting the entire AOI (not just the stratum) and (b) a model variance component that adds the uncertainty associated with the model coefficients. This model variance, which essentially characterizes how the model coefficients change with repeated sampling, assumes that the specified model form, in our case, LNI, is correct. In this study, the model error component of overall standard error of a biomass estimate played a relatively minor role, adding, on average, only 0.07 t/ha to state-level standard errors of biomass density, accounting for, on average, only 4.5% of the standard error. Previous aircraft lidar studies (Gobakken et al., 2012; Nelson et al., 2012) had noted much larger model error contributions; sometimes over 90% of the total variance of the biomass estimate was attributable to model error. There are two reasons why the contribution in this study is relatively small, though still significant: (1) The current study employs LNI models; the Y-intercept is locked at (0,0) and any uncertainty associated with the estimation of that Y-intercept is discounted.

Table 6

The Mexico linear-no-intercept models used to predict biomass density (t/ha) for each of the 5 forest strata within the states in 3 zones in Mexico (the number of NFI plots overflowed in each state are in parentheses).

Northern Mexico (NMex): includes the states of Baja California, Baja California Sur, Chihuahua (56), Coahuila, Durango (15), Nayarit, Nuevo Leon, San Luis Potosi, Sinaloa, Sonora (8), Tamaulipas, Zacatecas. NFI plots overflowed, by cover type: nonforest - 7, hardwood - 6, conifer - 64, burn - 2.

Southwestern Mexico (SWMex): includes the states of Aguascalientes, Chiapas (26), Colima, Distrito Federal, Guanajuato (12), Guerrero (1), Hidalgo (2), Jalisco (1), Mexico (5), Michoacan (1), Morelos (2), Oaxaca (45), Puebla (5), Queretaro (3), Tlaxcala. NFI plots overflowed, by cover type: nonforest - 8, hardwood - 36, conifer - 54, burn - 5.

Atlantic Crescent Mexico (ACMex): includes the states of Campeche (33), Quintana Roo (26), Tabasco (1), Veracruz (23), Yucatan (29). NFI plots overflowed: nonforest - 18, hardwood - 86, conifer - 2, burn - 6.

Ground - ALS models: $b_{gnd} = f(\text{ALS variables})$						
Zone/stratum	Model	R ²	RMSE (t/ha)	Largest VIF	n	Significant ALS variables
NMex/all strata	LNI	0.44	33.28	1.17	79	$h90_a, h50_a$
SWMex/hardwood-nf/conifer-burn	LNI	0.83	28.44	2.62	44	$\bar{h}_c, d2_a$
ACMex/all strata	LNI	0.74	29.20	5.69	59	$\bar{h}_a, h30_a, d4_a$
	LNI	0.59	47.32	2.69	112	$\bar{h}_a, h10_a, d1_a$
ALS - GLAS models: $\hat{b}_{ALS} = f(\text{GLAS variables})$						
Zone/stratum	Model	R ²	RMSE (t/ha)	Largest VIF	n	Significant GLAS variables
NMex/all strata	LNI	0.71	18.94	4.60	2186	$h25, h90, acq3, ct2$
SWMex/all strata	LNI	0.63	40.40	4.53	1293	$h75_{strm}, h10, trail$
ACMex/all strata	LNI	0.50	37.33	3.92	653	$h90, h10, fslope$

In fact, in CONUS (this study), the LNI model error contributions to state-level biomass density standard errors was 0.18 t/ha, accounting for, on average, 13.7% of the states' standard errors. At the national level, about 5.9% of CONUS standard error can be attributed to LNI model

uncertainty; in Mexico that number is 22.6%. (2) The current study ignores the model variance associated with the ground-airborne lidar models, thus ignoring a large and significant source of model variability. Preliminary calculations using linear models in CONUS suggest that a

Table 7

Mexico state and national estimates of total aboveground dry biomass density developed using the ALS-GLAS equations in Table 5.

States	State estimates						
	Nonforest Area (km ²)	Forest Area (km ²)	AGB dens (t/ha)	Sample SE(t/ha)	Model SE(t/ha)	Total SE(t/ha)	%model Error
Northern Mexico (NMex)							
Baja California	65,292	5044	20.89	2.58	0.42	2.62	2.54
Baja California Sur	67,732	3930	2.66	0.23	0.06	0.24	5.30
Chihuahua	173,132	74,605	38.17	1.88	0.43	1.93	4.89
Coahuila	141,420	8821	32.14	2.17	0.36	2.20	2.66
Durango	63,637	57,040	40.60	1.91	0.47	1.97	5.68
Nayarit	8032	19,009	45.24	2.32	0.67	2.42	7.69
Nuevo Leon	57,823	7336	43.94	3.14	0.52	3.19	2.64
San Luis Potosi	54,013	10,127	35.43	3.55	0.57	3.59	2.47
Sinaloa	25,305	31,458	42.31	2.13	0.69	2.23	9.44
Sonora	139,596	40,153	37.67	2.31	0.64	2.40	7.18
Tamaulipas	61,411	15,207	31.56	2.42	0.50	2.47	4.08
Zacatecas	59,778	15,213	32.11	2.53	0.41	2.57	2.53
Across 12 states	917,171	287,943	37.91	1.09	0.45	1.18	14.81
Southwestern Mexico (SWMex)							
Aguascalientes	4415	1194	37.29	8.22	0.99	8.27	1.44
Chiapas	33,918	39,199	70.48	4.33	0.78	4.40	3.11
Colima	2724	2962	55.64	4.33	1.12	4.47	6.23
Distrito Federal	1016	327	43.07	8.16	0.67	8.19	0.66
Guanajuato	20,039	8427	30.29	1.95	0.69	2.07	11.25
Guerrero	19,413	45,026	57.59	1.48	0.77	1.67	21.08
Hidalgo	14,989	6383	45.79	2.63	0.71	2.73	6.79
Jalisco	35,607	44,047	54.26	2.66	0.75	2.76	7.40
Mexico	14,633	7061	67.04	2.19	0.60	2.27	6.99
Michoacan	23,616	35,980	55.67	1.48	0.70	1.64	18.52
Morelos	2994	2013	46.79	3.02	0.90	3.15	8.24
Oaxaca	30,794	61,265	60.35	1.42	0.75	1.60	21.66
Puebla	22,219	12,146	49.02	2.92	0.79	3.02	6.85
Queretaro	7798	4297	37.12	2.69	0.84	2.82	8.76
Tlaxcala	3443	648	32.12	3.63	0.75	3.71	4.13
Across 15 states	239,528	270,973	57.42	1.04	0.72	1.27	32.63
Atlantic Crescent Mexico (ACMex)							
Campeche	10,491	38,930	75.89	5.27	1.31	5.43	5.81
Quintana Roo	6401	44,035	87.62	4.02	1.49	4.29	11.99
Tabasco	18,314	5202	42.91	6.81	0.96	6.88	1.94
Veracruz	56,441	13,779	109.75	9.39	1.78	9.56	3.45
Yucatan	10,449	27,234	55.47	1.86	1.41	2.33	36.37
Across 5 states	102,097	129,180	77.87	2.52	1.27	2.82	20.40
National Estimates - 3 Zones, NMex, SWMex, ACMex	1,258,796	688,096	53.10	0.77	0.42	0.88	22.59

Table 8
Comparison of published Mexico national estimates of biomass density, forest area, and total biomass.

	Biomass density				Total biomass (Gt)	
	Mean	SE	Area (km ²)		Total	SE
	(t/ha)	(t/ha)	Nonforest	Forest		
FAO - 2000 ^a	–	–	1,296,870	667,510	4.222	–
FAO - 2005 ^b	–	–	1,308,600	655,780	4.152	–
FAO, 2010a ^c	64	–	1,316,360	648,020	4.086	–
Cartus ^d	–	–	–	–	4.420	–
FAO/Mex, 2010b ^e	–	–	–	–	3.380	–
CONAFOR - 2015 ^f	–	–	–	–	3.200	–
Cartus ^g	–	–	–	–	3.060	–
This study	53.096	0.879	1,258,796	688,096	3.654	0.06

^a Forest area, Table 3, p. 233; Table 2, p. 228 for total land area (FAO, 2010a); total biomass, Table 11, p. 276 (FAO, 2010a).

^b Forest area, Table 3, p. 233; Table 2, p. 228 for total land area (FAO, 2010a); total biomass, Table 11, p. 276 (FAO, 2010a).

^c Biomass density, Table 11, p. 276, FAO (2010a); forest area, Table 3, p. 233; Table 2, p. 228 for total land area (FAO, 2010a); total biomass, Table 11, p. 276 (FAO, 2010a).

^d Cartus et al. (2014), p. 5573, no forest/nonforest mask used, C estimated for entire country using Landsat, PALSAR, topographic data, circa 2007/8.

^e FAO (2010b), biomass total based on INPhS ground plots and INEGI forest cover areas [report cited by Cartus et al. (2014), p. 5573].

^f CONAFOR estimate of total biomass based on ground plots and Landsat land cover area, circa 2015 (unpublished, as cited by Cartus et al., 2014, p. 5573).

^g Cartus et al. (2014), with Landsat forest/nonforest mask (series 4, INEGI) used to exclude nonforest and shrub biomass, circa 2007/8.

given state's *model* standard error will increase ~5x on average, increasing the *total* state standard errors by a factor of ~1.5x. Similar increases are expected at the national level.

5.2. Mexico

With respect to the estimates provided in Table 8, which estimate is correct? Certainly an argument could be made that the Cartus C estimates may be better since (1) theirs depends on wall-to-wall L-band radar observations, (2) L-band radar is sensitive to the biomass densities typically present in most Mexican forests (0–100 t/ha, Imhoff, 1995; Wagner et al., 2003), (3) the NFI-ALS sample (this study) is relatively sparse with respect to adequately representing the full range of canopy height and biomass conditions in Mexico, and (4) the GLAS sample is just that, a sample, i.e., not wall-to-wall. A counter-argument could be made that the ground-airborne lidar-GLAS sampling procedures used in CONUS and Quebec produced regional- and national-level estimates well within 10% of intensive ground-based samples.

Though we have no state-level ground information for Mexico, we expect scatter around the 1:1 line to be similar to that seen in the US (Fig. 4A and B). Fig. 12 in Cartus et al. (2014) compares their state estimates of C-density based on wall-to-wall PALSAR measurements of the entire country (no forest/nonforest mask employed) to national forest inventory ground measurements. The radar C estimates track the 1:1 (ground:radar) line very closely ($R^2 = 0.98$) and, multiplying their C estimates by 2, range from 0 to approximately 66 t/ha dry biomass, with 30 out of the 32 Mexican states reporting biomass densities at or below 35 t/ha. The current study's state-level estimates range from 2 to 109 t/ha (Baja California Sur and Veracruz, resp.) in areas defined as forest in our INEGI version 4 land cover crosswalk. Twenty-nine of our 32 state biomass density estimates were at or below 70 t/ha. Our per-state biomass density estimates are approximately double those of Cartus. This is understandable from the standpoint that our biomass numbers describe only the biomass in forest strata for a given state, whereas the biomass density estimates illustrated in Fig. 12 (Cartus et al.) report biomass averages for the entire state, including nonforested areas.

In terms of national biomass estimates (Table 8), the Cartus estimate associated with the state-level numbers in their Fig. 12 overestimates our national estimate, 3.645 ± 0.06 Gt, by 21% ($((4.42/3.645)*100)$). At the other end of the spectrum, when Cartus applied a forest/nonforest mask to exclude biomass contributions from nonforested areas, then their total forest biomass in Mexico drops to 3.06 Gt, a 16% underestimate ($((1-3.06/3.654)*100)$) relative to this study's GLAS total. Table 8 highlights two facts: (1) The areal extent of Mexican forests have been changing rapidly, with the forests experiencing extensive deforestation and degradation (FAO, 2010a, p. 19). (2) The definition of forest and the areal extent of those forests as described using satellite optical data are huge drivers with respect to calculating a country's total biomass. As demonstrated in Table 8, Mexico's total biomass can change by 44% ($((4.42 \text{ Gt}/3.06 \text{ Gt}) * 100)$) just by employing a forest/nonforest mask to control biomass contribution from areas defined as nonforest. Similarly, in a study in Alaska (Margolis et al., 2015), GLAS estimates of biomass and carbon for the entire state were generated with and without the contribution from the NLCD "shrub" land cover class (NLCD classes 51, 52). Inclusion of the shrub classes in Alaska lead to an apparent 43% increase in state estimates of total forest biomass, i.e., 3.02 Gt with shrub, 2.11 Gt without. In the final version of the North American boreal report, Margolis et al. (2015) reported Alaskan biomass without shrubs because the team had not included shrub allometry in the generation of the predictive ground-airborne lidar nor the airborne lidar-GLAS equations. A major take-home point is this: In a given country or state, a consistent definition of forest and a consistent delineation of that forest is, for bookkeeping purposes in a UN/REDD+ environment, all-important (Sexton et al., 2016).

6. Conclusion

A hybrid three-phase sampling framework is used to generate ca. 2005 biomass density estimates and total biomass estimates, with uncertainties, for CONUS and for Mexico at the state and national levels. The standard errors reported incorporate sampling variability and ALS-GLAS model variability, but due to our inability to incorporate ground-ALS model variability, all standard errors must be considered approximations and most are likely underestimated. Regardless, these results empirically demonstrate that studies that rely on models that employ airborne or satellite remote sensing assets to predict variables of interest - e.g., biomass, carbon, LAI, Lorey's height, canopy gap fraction, percent lignin, carbon flux - should report variances or standard errors that account for both sampling variability (if applicable) and model variability. Not to do so implies that the models used to generate the estimates are perfectly specified, perfectly parameterized, and without error, an entirely unrealistic assumption. To be clear, this comment applies not only to airborne and space lidar studies, but also to all studies that employ any remote sensing assets - e.g., radar, optical sensors, thermal sensors - to estimate, via modeling, a particular variable of interest.

What is the best way to keep track of forest biomass and carbon, by nation, globally? "Best" in this instance can be defined by accuracy, precision, statistical robustness, consistency over time, and repeatability. This study points up the difficulties involved when analysts try to develop analysis approaches to address this question because the question itself implies that there is, for a given nation, one correct answer. Viewed in the extreme, there is. In a study like this one, where the current state of a variable is being estimated, the correct answer, known without error, is the weight of the wood in a given country when we cut down all of the forests, dry them out, and weigh them on a particular date. As nonsensical as this solution is, it is fair to ask if there is a consistent sampling or census mechanism, preferably one dependent on global remote sensing assets, that can be employed to provide an accurate, repeatable national mean or total estimate that is unbiased or minimally biased, with a statistically defensible estimate of uncertainty.

We argue that space lidars and moderate resolution space optical systems (15 m - 30 m, e.g., SPOT, Landsat, IRS), when used in conjunction with sampling frameworks similar to the one reported in this study, might provide such a mechanism. But, regardless of the asset employed, the key to the utility of those remote-sensing-based estimates is consistency in the definitions, measurement methods, and statistical techniques over time. A concerted effort would have to be made to calibrate airborne and space sensors so that the measurements acquired are comparable across decadal time scales, similar to the calibration efforts made with Landsat. Otherwise, national and continental remote-sensing-based estimates of biomass and carbon such as those reported in this current study will continue to be one-offs, dependent on the definition of forest and on the particular measurement and observational characteristics of the sensors employed.

Acknowledgments

This research was funded by NASA's Carbon Cycle Science Program within the Science Mission Directorate Earth Science Division-NNH10ZDA001N-CARBON(2010).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.rse.2016.10.038>.

References

- Blackard, J.A., Finco, M.V., Helmer, E.H., Holden, G.R., Hoppus, M.L., Jacobs, D.M., et al., 2008. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sens. Environ.* 112 (2008):1658–1677. <http://dx.doi.org/10.1016/j.rse.2007.08.021>.
- 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 Québec. *Remote Sens. Environ.* 112 (2008):3876–3890. <http://dx.doi.org/10.1016/j.rse.2008.06.003>.
- Cartus, O., Kellndorfer, J., Walker, W., Franco, C., Bishop, J., Santos, L., Fuentes, J.M.M., 2014. A national, detailed map of forest aboveground carbon stocks in Mexico. *Remote Sens.* 6:5559–5588. <http://dx.doi.org/10.3390/rs6065559>.
- Cook, B.D., Corp, L.A., Nelson, R.F., Middleton, E.M., Morton, D.C., McCorkel, J.T., et al., 2013. NASA Goddard's LiDAR, hyperspectral and thermal (G-LiHT) airborne imager. *Remote Sens.* 2013 (5):4045–4066. <http://dx.doi.org/10.3390/rs5084045>.
- Dubayah, R., Goetz, S., Blair, J.B., Luthcke, S., Healey, S., Hansen, M., et al., 2014a. The Global Ecosystem Dynamics Investigation (Abstract). *ForestSAT 2014 Conference*, Nov. 4–7, 2014, Riva de Garda, Italy.
- Dubayah, R., Goetz, S., Blair, J.B., Fatoyinbo, T.E., Hansen, M., Healey, S., et al., 2014b. The Global Ecosystem Dynamics Investigation (Abstract). *AGU Fall Meeting*, Dec. 15–19, 2014, San Francisco, CA, USA.
- Ene, L.T., Næsset, E., Gobakken, T., Gregoire, T.G., Ståhl, G., Nelson, R., 2012. Assessing the accuracy of regional LiDAR-based biomass estimation using a simulation approach. *Remote Sens. Environ.* 123 (2012):579–592. <http://dx.doi.org/10.1016/j.rse.2012.04.017>.
- FAO (Food and Agriculture Organization), 2010a. Global Forest Resources Assessment 2010, Main Report. FAO Forestry Paper 163, FAO, Rome. (378 pgs). <http://www.fao.org/docrep/013/i1757e/i1757e.pdf> (ISBN 978-92-5-106654-6).
- FAO (Food and Agriculture Organization), 2010b. Evaluación de los Recursos Forestales Mundiales 2010—Informe Nacional México [Evaluation of Global Forest Resources 2010 - National Report Mexico] FAO, Rome.
- Giglio, L., Loboda, T., Roy, D.P., Quayle, B., Justice, C.O., 2009. An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote Sens. Environ.* 113 (2009):408–420. <http://dx.doi.org/10.1016/j.rse.2008.10.006>.
- Gobakken, T., Næsset, E., Nelson, R., Bolland, O.M., Gregoire, T.G., Ståhl, G., et al., 2012. Estimating biomass in Hedmark County, Norway using national forest inventory field plots and airborne laser scanning. *Remote Sens. Environ.* 123 (2012):443–456. <http://dx.doi.org/10.1016/j.rse.2012.01.025>.
- Gregoire, T.G., Næsset, E., McRoberts, R.E., Ståhl, G., Andersen, H.-E., Gobakken, T., Ene, L., Nelson, R., 2016. Statistical rigor in LiDAR-assisted estimation of aboveground forest biomass. *Remote Sens. Environ.* 173 (2016):98–108. <http://dx.doi.org/10.1016/j.rse.2015.11.012>.
- Gwenzi, D., & Lefsky, M.A. (2014). Prospects of Photon-Counting LiDAR for Savanna Ecosystem Structural Studies. *International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Vol. XL-1, 2014, ISPRS Tech. Commission Symposium, 17–20 November 2014, Denver CO., pgs. 141–147. doi: 10.5194/isprsarchives-XL-1-141-2014.
- Huang, W., Swatantran, A., Johnson, K., Duncanson, L., Tang, H., O'Neil Dunne, J., Hurr, G., et al., 2015. Local discrepancies in continental scale biomass maps: a case study over forested and non-forested landscapes in Maryland, USA. *Carbon Balance Manag.* 10 (19). <http://dx.doi.org/10.1186/s13021-015-0030-9> (16 pgs).
- Imhoff, M.L., 1995. Radar backscatter and biomass saturation: ramifications for global biomass inventory. *IEEE Trans. Geosci. Remote Sens.* 23 (2), 511–518.
- INEGI (Instituto Nacional de Estadística y Geografía), 2010. Conjunto Nacional de Uso del Suelo y Vegetación a escala 1:250,000, Serie IV DGG-INEGI, Aguascalientes, México. <http://www.inegi.org.mx>.
- Lefsky, M., Harding, D., Keller, M., Cohen, W.B., Carabajal, C.C., Espirito-Santo, F.D.B., Hunter, M.O., deOliveira, R., 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophys. Res. Lett.* 32, L22S02. <http://dx.doi.org/10.1029/2005GL023971>.
- Maass, J.M., Vose, J.M., Swank, W.T., Martínez-Yrizar, A., 1995. Seasonal changes of leaf area index (LAI) in a tropical deciduous forest in west Mexico. *For. Ecol. Manag.* 74 (1995), 171–180.
- Magruder, L.A., Ricklefs, R.L., Siverberg, E.C., Horstman, M.F., Suleman, M.A., Schutz, B.E., 2010. ICESat geolocation validation using airborne photography. *IEEE Trans. Geosci. Remote Sens.* 48 (6):2758–2766. <http://dx.doi.org/10.1109/TGRS.2010.2040831>.
- Margolis, H., Nelson, R.F., Montesano, P.M., Beaudoin, A., Sun, G., Andersen, H.-E., et al., 2015. Combining satellite lidar, airborne lidar, and ground plots to estimate the amount and distribution of aboveground biomass in the boreal forest of North America. *Can. J. For. Res.* 45:838–855. <http://dx.doi.org/10.1139/cjfr-2015-0006>.
- McRoberts, R.E., 2006. A model-based approach to estimating forest area. *Remote Sens. Environ.* 103 (2006):56–66. <http://dx.doi.org/10.1016/j.rse.2006.03.005>.
- Montesano, P.M., Sun, G., Dubayah, R., Ranson, K.J., 2014a. The uncertainty of plot-scale forest height estimates from complementary spaceborne observations in the taiga-tundra ecotone. *Remote Sens.* 6 (10):10070–10088. <http://dx.doi.org/10.3390/rs61010070>.
- Montesano, P.M., Nelson, R.F., Dubayah, R.O., Sun, G., Cook, B.D., Ranson, K.J., et al., 2014b. The uncertainty of biomass estimates from LiDAR and SAR across a boreal forest structure gradient. *Remote Sens. Environ.* 154 (2014):398–407. <http://dx.doi.org/10.1016/j.rse.2014.01.027>.
- Næsset, E., 1997. Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sens. Environ.* 61:246–253. [http://dx.doi.org/10.1016/S0034-4257\(97\)00041-2](http://dx.doi.org/10.1016/S0034-4257(97)00041-2).
- Næsset, E., Gobakken, T., Solberg, S., Gregoire, T.G., Nelson, R., Ståhl, G., et al., 2011. Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: a case study from a boreal forest area. *Remote Sens. Environ.* 115 (2011):3599–3614. <http://dx.doi.org/10.1016/j.rse.2011.08.021>.
- Neigh, C.S.R., Nelson, R.F., Ranson, K.J., Margolis, H.A., Montesano, P.M., Sun, G., et al., 2013. Taking stock of circumpolar forest carbon with ground measurements, airborne, and space LiDAR. *Remote Sens. Environ.* 137 (2013):274–287. <http://dx.doi.org/10.1016/j.rse.2013.06.019>.
- Nelson, R., Boudreau, J., Gregoire, T.G., Margolis, H., Næsset, E., Gobakken, T., et al., 2009a. Estimating Quebec provincial forest resources using ICESat/GLAS. *Can. J. For. Res.* 39:862–881. <http://dx.doi.org/10.1139/X09-002>.
- Nelson, R., Ranson, K.J., Kimes, D.S., Kharuk, V., Montesano, P., 2009b. Estimating Siberian timber volume using MODIS and ICESat/GLAS. *Remote Sens. Environ.* 113 (2009), 691–701.
- Nelson, R., Gobakken, T., Næsset, E., Gregoire, T.G., Ståhl, G., Holm, S., Flewelling, J., 2012. Lidar sampling – using an airborne profiler to estimate forest biomass in Hedmark County, Norway. *Remote Sens. Environ.* 123 (2012):563–578. <http://dx.doi.org/10.1016/j.rse.2011.10.036>.
- Nyyssönen, A., Kilkki, P., Mikkola, E., 1967. On the precision of some methods of forest inventory. *Acta Forestalia Fennica* 81 (60 pgs).
- Nyyssönen, A., Roiko-Jokela, P., Kilkki, P., 1971. Studies on improvement of the efficiency of systematic sampling in forest inventory. *Acta Forestalia Fennica* 116 (26 pgs).
- Osborne, J.G., 1942. Sampling errors of systematic and random surveys of cover-type areas. *J. Am. Stat. Assoc.* 37 (218), 256–264.
- Royall, R.M., 1970. On finite population sampling theory under certain linear regression models. *Biometrika* 57 (2):377–387. <http://dx.doi.org/10.1093/biomet/57.2.377>.
- Royall, R.M., Herson, J., 1973. Robust estimation in finite populations. *J. Am. Stat. Assoc.* 68 (344):880–889. <http://dx.doi.org/10.2307/2284516>.
- Saarela, S., Holm, S., Grafstrom, A., Schnell, S., Næsset, E., Gregoire, T.G., Nelson, R.F., Ståhl, G., 2016. Model-based inference for forest inventory utilizing three sources of information and hierarchical modelling. *Ann. For. Sci.* (accepted for publication, 34 pgs).
- SAS Institute, Inc., 1989. *SAS/STAT User's Guide*, Ver. 6. 4th edition. Vol. 2. SAS Institute, Inc., Cary, North Carolina USA, p. 846.
- Sexton, J.O., Noojipady, P., Song, X.-P., Feng, M., Song, D.-X., Kim, D.-H., et al., 2016. Conservation policy and the measurement of forests. *Nat. Clim. Chang.* 6:192–196. <http://dx.doi.org/10.1038/nclimate2816>.
- Snowdon, P., 1991. A ratio estimator for bias correction in logarithmic regressions. *Can. J. For. Res.* 21, 720–724.
- Ståhl, G., Holm, S., Gregoire, T.G., Gobakken, T., Næsset, E., Nelson, R., 2011. Model-based inference for biomass estimation in a LiDAR sample survey of Hedmark County, Norway. *Can. J. For. Res.* 41:96–107. <http://dx.doi.org/10.1139/X10-161>.
- Ståhl, G., Saarela, S., Schnell, S., Holm, S., Breidenbach, J., Healey, S.P., et al., 2016. Use of models in large-area forest surveys: comparing model-assisted, model-based, and hybrid estimation. *For. Ecosyst.* 3 (5). <http://dx.doi.org/10.1186/s40663-016-0064-9> (11 p).
- Tang, H., Brolly, M., Zhao, F., Strahler, A.H., Schaaf, C.L., Ganguly, S., Zhang, G., Dubayah, R., 2014a. Deriving and validating leaf area index (LAI) at multiple spatial scales through lidar remote sensing: a case study in Sierra National Forest, CA. *Remote Sens. Environ.* 143 (2014):131–141. <http://dx.doi.org/10.1016/j.rse.2013.12.007>.

- Tang, H., Dubayah, R., Brolly, M., Ganguly, S., Zhang, G., 2014b. Large-scale retrieval of leaf area index and vertical foliage profile from the spaceborne waveform lidar (GLAS/ICESat). *Remote Sens. Environ.* 154 (2014):8–18. <http://dx.doi.org/10.1016/j.rse.2014.08.007>.
- Wagner, W., Luckman, A., Vietmeier, J., Tansey, K., Blazter, H., Schmullius, C., et al., 2003. Large-scale mapping of boreal forest in SIBERIA using ERS tandem coherence and JERS backscatter data. *Remote Sens. Environ.* 85 (2003):125–144. [http://dx.doi.org/10.1016/S0034-4257\(02\)00198-0](http://dx.doi.org/10.1016/S0034-4257(02)00198-0).
- Wickham, J.D., Stehman, S.V., Gass, L., Dewitz, J., Fry, J.A., Wade, T.G., 2013. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sens. Environ.* 130 (2013):294–304. <http://dx.doi.org/10.1016/j.rse.2012.12.001>.