Using large footprint LiDAR to predict forest canopy height and aboveground biomass in high biomass tropical forests: a challenging task.

Auteur : De Grave, Charlotte
Promoteur(s) : Lejeune, Philippe
Faculté : Gembloux Agro-Bio Tech (GxABT)
Diplôme : Master en bioingénieur : gestion des forêts et des espaces naturels, à finalité spécialisée
Année académique : 2016-2017
URI/URL : http://hdl.handle.net/2268.2/3083

USING LARGE FOOTPRINT LIDAR TO PREDICT FOREST CANOPY HEIGHT AND ABOVEGROUND BIOMASS IN HIGH BIOMASS TROPICAL FORESTS: A CHALLENGING TASK

CHARLOTTE DE GRAVE

TRAVAIL DE FIN D'ETUDES PRESENTE EN VUE DE L'OBTENTION DU DIPLOME DE MASTER BIOINGENIEUR EN GESTION DES FORETS ET DES ESPACES NATURELS

ANNEE ACADEMIQUE 2016-2017

CO-PROMOTEURS: LOLA FATOYINBO, PHILIPPE LEJEUNE

USING LARGE FOOTPRINT LIDAR TO PREDICT FOREST CANOPY HEIGHT AND ABOVEGROUND BIOMASS IN HIGH BIOMASS TROPICAL FORESTS: A CHALLENGING TASK

CHARLOTTE DE GRAVE

TRAVAIL DE FIN D'ETUDES PRESENTE EN VUE DE L'OBTENTION DU DIPLOME DE MASTER BIOINGENIEUR EN GESTION DES FORETS ET DES ESPACES NATURELS

ANNEE ACADEMIQUE 2016-2017

CO-PROMOTEURS: LOLA FATOYINBO, PHILIPPE LEJEUNE

Ce travail de fin d'études a été réalisé dans les locaux du NASA Goddard Space Flight Center, au sein du Laboratoire des Sciences Biosphériques, sous la supervision de Madame Lola Fatoyinbo.

L'université de Liège a fourni un soutien financier à l'auteur sous couvert d'un contrat de mobilité pour un stage étudiant hors Union Européenne.

Table of Contents

Using large footprint LiDAR to predict forest canopy height and aboveground biomass in high biomass tropical forests: a challenging task.

De Grave, Charlotte[1*]

1. Université de Liège, Gembloux Agro-Bio Tech

* Corresponding author, email: degravecharlotte@gmail.com

Highlights

- We used a pantropical model to compute field-based biomass estimates in a high biomass forest in Osa peninsula, Costa Rica.
- Small plot sizes cause high biomass variability between plots, leading to considerable model errors.
- Maximum tree height is a good predictor for plot level biomass in small plots.
- Applying a model based on maximum tree height generated biomass estimates with an uncertainty of ~ 30 %.

Résumé

Afin d'évaluer l'impact de la déforestation sur le changement climatique, des estimations fiables de la biomasse aérienne sont nécessaires. Les techniques de télédétection permettent d'étendre les estimations basées sur les mesures de terrain à des échelles spatiales plus vastes. Bien que les limites de sensibilité du LiDAR (Light Detection And Ranging) soient largement supérieures à celles des systèmes optiques et radars, son comportement aux densités de biomasse extrêmement élevées (≥ 500 Mg ha^{-1}) reste peu étudié. Le Parc National Corcovado (Costa Rica) représente un enjeu pour l'utilisation du LiDAR, dû aux conditions de très hautes biomasses et à la petite taille des placettes de terrain disponibles (0.07 ha). Pour ce site, les données LiDAR ne prédisent pas significativement la hauteur du couvert en raison de la faible coregistration (chevauchement spatial) des placettes de terrain et des empreintes LiDAR. Les données LiDAR prédisent par contre significativement la biomasse mais avec une faible précision (RMSE > 50%). Afin de limiter la variabilité de la biomasse entre placettes, ce qui occasionne des erreurs de modèle considérables, nous suggérons une taille de placette minimale de 0.2 ha. De plus, il semblerait que la hauteur d'arbre maximale (Hmax) soit un bon indicateur de la biomasse à l'échelle de la placette lorsque les placettes sont petites, alors que la hauteur d'arbre dominante (Hdom) et moyenne (Hmoy) sont plus performantes quand la taille des placettes augmente. Un modèle basé sur Hmax est utilisé pour prédire la biomasse à l'échelle de l'empreinte et donne lieu à des densités de biomasse moyennes à l'échelle de la bande LiDAR de 281.5 Mg ha^{-1} pour Corcovado et de 194.8 Mg ha^{-1} pour un autre site d'étude, la Station Biologique de La Selva

(Costa Rica). Ces valeurs sont comparables à d’autres résultats trouvés dans la région néotropicale.

Mots clés : Biomasse forestière, LiDAR, LVIS, hauteur de la canopée, haute biomasse, taille de placette

Abstract

In order to assess the impact of deforestation on climate change, reliable estimates of aboveground biomass are needed. Estimates based on field measurements can be extended over broader spatial scales using remote sensing techniques. Although LiDAR (Light Detection And Ranging) shows no saturation at the biomass levels that represent the limits for optical and radar systems, it is not clear how it behaves at extremely high biomass densities (500 Mg ha^{-1} and above). Our study site in Corcovado National Park (Costa Rica) presents challenges for LiDAR use because of very high biomass conditions and the small size of the plots (0.07 ha). Because of the low co-registration (spatial overlap) between field plots and LiDAR footprints, LiDAR metrics could not significantly predict canopy heights. Biomass on the other hand was significantly predicted but with low accuracy (RMSE above 50%). We suggest that a plot size of at least 0.2 ha is needed to limit the biomass variability between plots, which may otherwise cause considerable model errors. Additionally, field maximum tree height (Hmax) proved a good predictor of plot level biomass in plots of small size, while dominant tree height (Hdom) and mean tree height (Hmean) seemed to outperform Hmax as plot size increased. We used a model based on Hmax to predict biomass at footprint level and obtained mean biomass densities at swath level of 281.5 Mg ha^{-1} for Corcovado and 194.8 Mg ha^{-1} for our other field site, the La Selva Biological Station in Costa Rica. These values are comparable to other results found in the Neotropics.

Keywords: Forest biomass, LiDAR, LVIS, canopy height, high biomass, plot size

1.0 Introduction

One of the greatest threats facing our planet is climate change, which is primarily caused by elevated atmospheric concentrations of greenhouse gases such as carbon dioxide (CO_2). While forests help in mitigating climate change through carbon sequestration, deforestation causes the stored carbon to be released as CO_2 into the atmosphere (Le Toan *et al.*, 2011). In order to assess the impact of deforestation on the climate, estimates of the forest carbon stocks before disturbance are needed (Drake *et al.*, 2003). Aboveground biomass (hereafter biomass) is a direct indicator of forest carbon stocks and is often used to estimate other terrestrial carbon pools (e.g. litter, dead wood and below ground biomass; Goetz and Dubayah, 2011). Biomass can be estimated with allometric equations that relate field measurements, such as DBH and tree height. Remote sensing techniques can extend these field-based estimates over broader spatial scales (Huang *et al.*, 2013). Passive optical and Synthetic Aperture Radar (SAR) instruments tend to be insensitive to changes in forest biomass above certain biomass levels, around 150Mg ha^{-1} for radar systems and at even lower levels for the optical sensors (Mitchard *et al.*, 2012; Zolkos, Goetz and Dubayah, 2013). LiDAR (Light Detection And Ranging), which is an active remote sensing technique using laser light, is able to overcome these saturation problems thanks to its high sensitivity to forest structure (Drake *et al.*, 2002). This technique enables indeed to capture the complex three-dimensional (3-D) structure of forest canopies and the underlying ground surface topography at very high spatial resolutions (Frazer *et al.*, 2011), even when canopy cover is up to 99% (Dubayah *et al.*, 2010). LiDAR instruments record the time between pulse emission and its return to the sensor after reflection by the objects within the area illuminated by the laser (Drake *et al.*, 2002). It thus measures the range, i.e. the direct distance from the laser emitter to the reflecting surfaces (ground, vegetation, …; Dashora, Lohani and Deb, 2013).

LiDAR sensors can be spaceborne (e.g. the Geoscience Laser Altimeter System = GLAS) or airborne (e.g. Laser Vegetation Imaging Sensor = LVIS). Airborne LiDAR systems use either discrete return or full return sensors. Discrete return LiDAR systems represent forested areas as three-dimensional point clouds, from which canopy height and canopy density estimates can be derived (Duncanson, Niemann and Wulder, 2010). However, these systems generally yield only between two (first and last returns) and six reflection points per laser shot (Magruder, 2010), which is insufficient to generate a detailed outline of the within-canopy and understory structure (Hancock *et al.*, 2017). Full return LiDAR systems on the other hand record the energy of the reflected signal over time since pulse emission. The form of the resulting energy wave reflects the vertical distribution of the vegetation (Duncanson, Niemann and Wulder, 2010), and allows estimation of various metrics such as top canopy height and relative heights (RH) to the ground elevation, at which different proportions of the total reflected energy are returned to the sensor (Zolkos, Goetz and Dubayah, 2013). For example, the RH75 metric is the height above the ground elevation below which 75% of the returned energy is situated in the waveform. These metrics have shown to be useful predictors of canopy vertical structure and biomass (Dubayah *et al.*, 2010). The footprint of the full waveform LiDAR system refers to the size of the area sampled by a single pulse (Pirotti, 2011).

Most commercial systems have a small-footprint (0.2 – 3 m diameter, depending on flying height and beam divergence) with a high point density (Mallet and Bretar, 2009). This allows the vegetation geometry to be modeled with greater detail as each laser pulse is reflected by a different part of the tree (Pirotti, 2011). Nevertheless, the laser beam has a high probability of missing the ground and the treetop which may lead to biased estimates of tree heights. On the other hand, large footprints systems (10 – 70m diameter) increase the probability of the laser beam to hit both the ground and the canopy top. However, as each echo results from the integration of several targets at different locations and with different properties (Mallet and Bretar, 2009), larger footprints lead to less detailed models of the vegetation geometry.

The Global Ecosystem Dynamics Investigation (GEDI) mission from NASA and from the University of Maryland, due to launch in 2018, will deploy a multi-beam full return LiDAR on the International Space Station (ISS) and provide billions of 25m-footprints of forest structure per year (Dubayah, 2015). The mission will cover areas between 50° north and 50° south and thereby include all tropical and subtropical forests (Qi and Dubayah, 2016). In anticipation of the mission, LVIS, the GEDI precursor airborne instrument (Mountrakis and Li, 2017), has been collecting large footprint LiDAR data over field plots in multiple forest types. However, while LiDAR has been shown to accurately estimate canopy height (e.g. Duncanson, Niemann and Wulder, 2010; Fatoyinbo and Simard, 2013), detailed analysis of its accuracy to estimate biomass in forests of very high biomass is still lacking.

In 2013, Zolkos, Goetz and Dubayah evaluated 71 different studies using LiDAR data to estimate forest biomass, with mean field-estimated biomass values varying from 15 to 602 Mg ha^{-1}. These studies concerned both discrete return LiDAR (62% of the studies) and full return (either airborne or spaceborne) LiDAR (38%) and 89% of the studies were able to successfully predict biomass from LiDAR data (multiple R-squared "R^2" value of 0.6 or more; mean R^2 of 0.75 across all studies). The authors showed that model performance decreases with increasing biomass. At mean field-estimated biomass values from 300 to 500 Mg ha^{-1}, the Root Mean Square Error (RMSE) not only increases but also becomes more fluctuating (see fig. 3A in Zolkos, Goetz and Dubayah, 2013). Above 500 Mg ha^{-1}, we can only assume that the pattern is, if not intensifying, at least of the same order.

Considering the imminent GEDI mission, the present study aims to highlight the challenges of using large footprint LiDAR data to produce accurate estimates of canopy height and biomass in tropical forests with very high biomass densities, in real world scenarios. Although LiDAR shows no saturation at the biomass levels that represent the limits for optical and radar systems (Mitchard *et al.*, 2012), it is not clear how it behaves at extremely high biomass densities (500 Mg ha^{-1} and above). In addition, much of the tropics is persistently cloud-covered which makes the data acquisition sometimes very challenging (pers. comm. Michelle Hofton).

Here, we (1) estimate biomass densities at plot level for three different field sites – including a very high biomass tropical forest - using existing allometric equations and assess how these estimates can be predicted by field measurements, (2) evaluate airborne LiDAR's ability to estimate forest height and biomass in high biomass and high cloud cover conditions (via “area-based” models), and (3) estimate biomass densities at swath level and assess the associated uncertainties in high biomass tropical areas.

2.0 Materials and Methods

2.1 Study sites and field data

Our primary study site is the Corcovado National Park, which is situated in the Osa Peninsula (Southwest Costa Rica) and is known to harbor one of the densest rain forests in Central America. The peninsula is also home to over 700 tree species, making it the most botanically diverse region in all Central America (Ankersen, Regan and Mack, 2006). The annual average temperature and average rainfall are 25°C and 6000 mm per year respectively. The rainy season, which last from August until December, is followed by a 4-month period of reduced rainfall, which last until April (Cornejo *et al.*, 2012). The vegetation of the peninsula is classified according to Holdridge’s life zone system (Holdridge, 1967) as a “tropical wet forest”. Field data were collected for seventeen 15-m radius plots (0.07 ha) in the southern part of the Corcovado National Park in 2014 (see fig.1). These plots were located near the coast (at a maximum distance of 2.5 km) and the altitudes range from 15 to 150 m above the ellipsoid. At each plot, the species, the diameter at breast height (DBH) or, when necessary, above buttresses, and the height were recorded for all trees with a DBH above 5 cm.

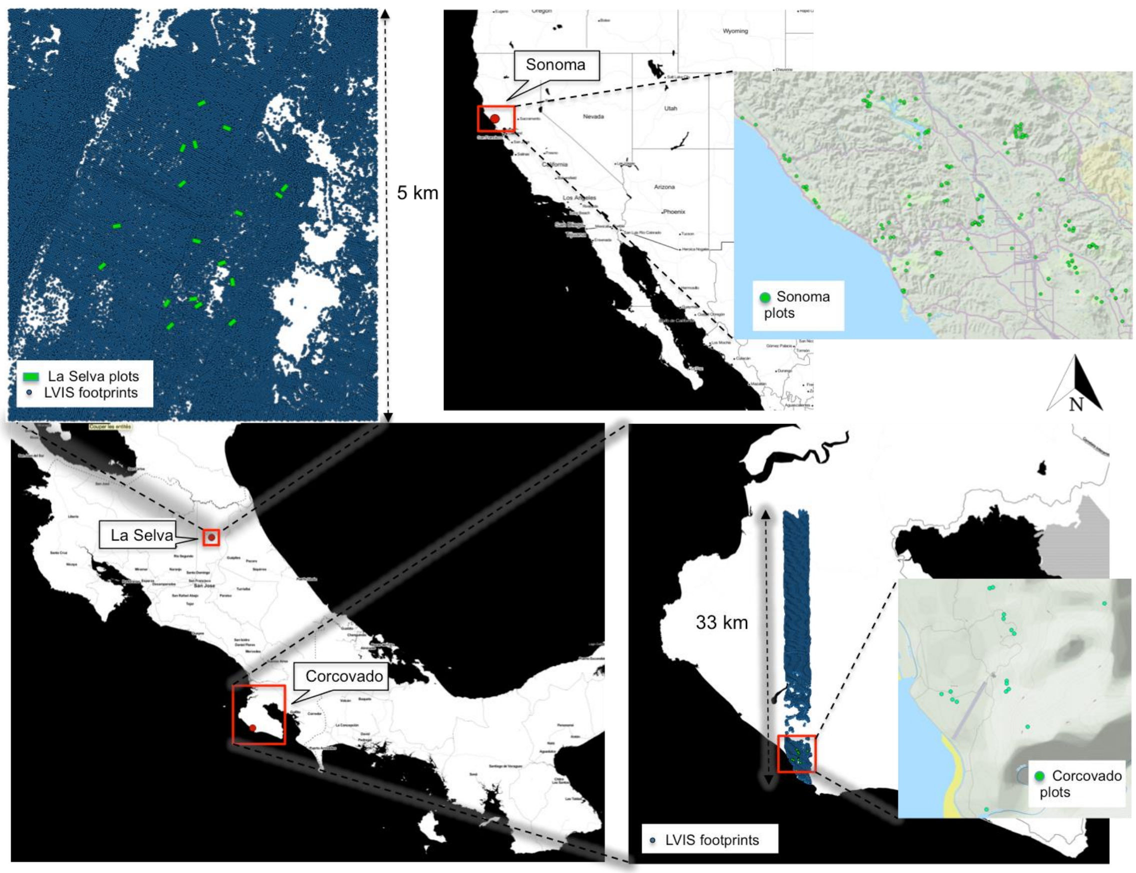

Fig. 1. Field data were collected in three different sites: Corcovado National Park (Costa Rica), La Selva Biological Station (Costa Rica) and Sonoma County (California). Waveform LiDAR data were collected with the Laser Vegetation Imaging Sensor (LVIS) for Corcovado (in blue at the bottom right corner of the image; laser swath of 33 km long) and La Selva (in blue at the top left; laser swath of 5 km long). Airborne Laser Scanner (ALS) sensors collected wall to wall discrete return LiDAR data for Sonoma (not shown). See section 2.3 for more details on the LiDAR data.

For a comparison, two other sites were included in the analysis. The La Selva Biological Station in northeast Costa Rica is also classified as a tropical wet forest by

Holdridge (1967) but has lower biomass densities as it comprises a mixture of old growth and secondary lowland rainforests, along with remnant plantations and various agroforestry treatments. Although its topography is similar to that of Corcovado (< 150 m), the region receives less rain (4000 mm on average per year; Dubayah *et al.*, 2010). Field data are collected each year in eighteen 0.5-ha plots as part of the Carbono project, a long-term landscape-scale monitoring of tropical rainforest productivity and dynamics (Clark and Clark, 2000). We used the field data of 2005 to fit the LVIS data timewise. At each plot, the DBH (or when necessary the diameter above buttresses) and the species were collected for all trees with a DBH above 10 cm. As tree heights weren't measured, we computed them using the pantropical diameter–height allometric model of Chave *et al.* (2014), which is based on the DBH and an environmental stress factor E, which integrates three bioclimatic variables (temperature seasonality, precipitation seasonality and climatic water deficit; see equation 6a in Chave *et al.*, 2014).

We compared our results from the two tropical sites with those from a temperate site located in California. Field data were collected during 2014 in 179 variable radius plots across the Sonoma County, as part of a project included in NASA's Carbon Monitoring System (CMS) program. This project focusses on developing empirical models relating field estimates of forest biomass to LiDAR metrics and on producing county-level biomass maps. Plot locations were distributed along various vegetation types among which conifer, deciduous and mixed forests but also non-forest ecosystems like wetlands, herb and shrub vegetation. In each plot, DBH and species were recorded for all trees with a DBH above 5 cm, as well as the height of the tallest 1 to 3 trees (Duncanson et al., in revision).

We used the stem diameters and heights measured on field to calculate quadratic mean stem diameter (QMSD), basal area (BA) and Lorey's mean height (LH). QMSD[1] was calculated to compensate for the different diameter thresholds (i.e. 5 vs. 10 cm) that were used in the different field campaigns. This mean gives greater weight to larger trees and is greater than the arithmetic mean by an amount that depends on the variance of diameters (Curtis and Marshall, 2000). We also calculated LH[2] which is a basal area weighted mean height and has often shown high significant relationships with LiDAR metrics (Lefsky, 2010; Asner and Mascaro, 2014).

2.2 Field biomass estimation

For the two tropical sites, we calculated field biomass using the R package "BIOMASS". The package allows to correct the taxonomy of the trees, to retrieve an estimate of their wood density using the global wood density database and to compute their biomass and associated uncertainty (Réjou-Méchain *et al.*, 2015). To estimate biomass, the package uses the pantropical allometric model of Chave *et al.* (2014) which is based on the DBH, the height and the wood density (WD) of individual trees (see equation 4 in Chave *et al.*, 2014).

---

[1] $QMSD\ (cm) = \sqrt{(\frac{\sum_{i=1}^{j} D^2{}_i}{N})}$ with D_i = diameter of tree i (cm) N = number of trees (Rondeux, 1993).

[2] $Lorey's\ mean\ height\ (m) = \frac{\sum_{i=1}^{j} g_i\, h_i}{G}$ with g_i = basal area of tree (m^2) i ; h_i = height of tree (m) i ; G = total basal area (m^2) (Rondeux, 1993).

For the Californian plots, we used the allometric equations developed by Jenkins *et al.* (2003) for different hardwood and softwood species groups. These models are based solely on the DBH of the trees.

We then calculated plot level biomass by summing the biomass of individual trees.

2.3 LiDAR Data

We used LiDAR data acquired by LVIS. This is a waveform digitizing, airborne laser altimeter which operates at altitudes around 10 km above ground and scans footprints with a nominal diameter of 25 m (Blair, Rabine and Hofton, 1999; see fig. 2). The resulting waveforms first need to be processed before any height metrics can be derived. Latitude, longitude and altitude estimates are computed for each footprint by merging the laser data to the data received from the Global Positioning System (GPS) receiver that is coupled to the sensor. Footprints also receive an aircraft attitude (roll, pitch and yaw) estimate from the integrated Inertial Navigation System (INS). Various biases affect the laser ranges, one of which is linked to the refraction and velocity change of light in the atmosphere compared to a vacuum and can be accounted for by measuring the atmospheric pressure and temperature of the air during the flight. Systematic instrument biases are corrected by comparing the known elevation of a ground feature (e.g. base station antenna) with that obtained from the laser range measurement. Finally, the waveforms are geolocated by transforming the local reference system within the aircraft to the global WGS-84 ellipsoidal system. For a complete description of the processing procedures, see Hofton, Minster and Blair, 2000.

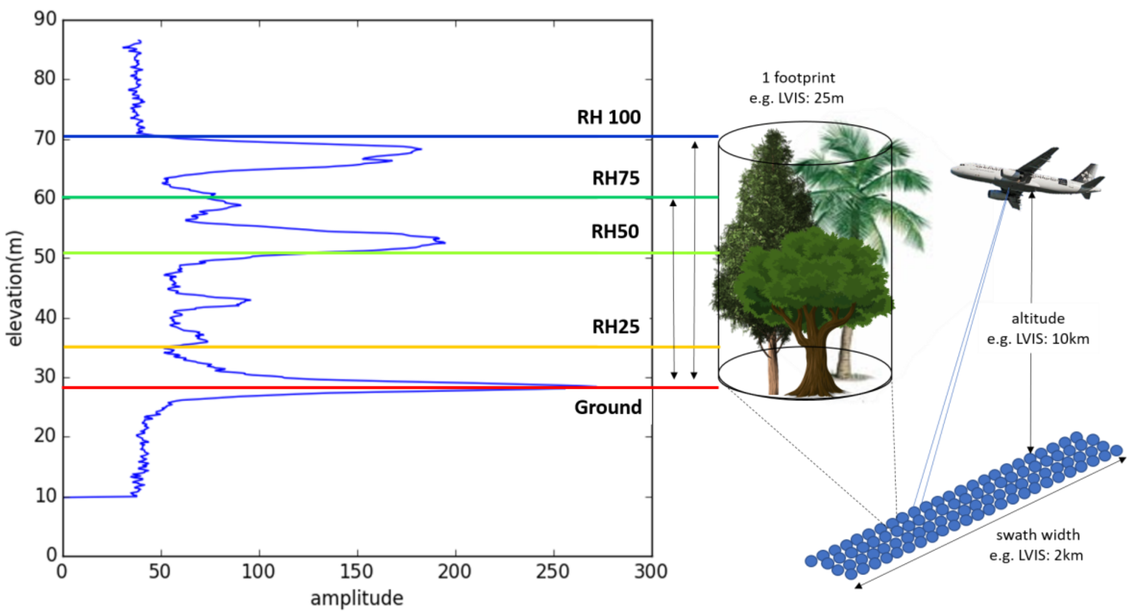

Fig. 2. An illustration of full return (waveform) LiDAR remote sensing equipped with the airborne Laser Vegetation Imaging Sensor (LVIS), from which data was used in the present study. The sensor emits laser pulses which are reflected by different surfaces (canopy, ground, …) and records the returned energy over time. Top canopy height (RH100) and other relative heights (RH), representing cumulative percentages of waveform energy (i.e. 25%, 50%, and 75%) are important metrics, which can be derived from the resulting waveform (Drake *et al.*, 2002).

After waveform processing, a noise threshold is chosen based on the background noise statistics recorded during the flight. The last mode of the waveform (or the first when starting from the trailing edge) over that noise threshold is regarded as the ground return (pers. comm. Michelle Hofton). The elevation of the ground return is defined as the center of the ground return mode. When the ground return signal is strong, there is no possible misinterpretation of the ground elevation. However, in case of weak ground returns, e.g. when the cloud cover is important or in dense multi-storey forests, the automated ground finding algorithms can misplace the ground. This causes the RH metrics to be also wrong as they are all derived from the ground (Dubayah *et al.*, 2010). The figure below (see fig.3) illustrates the problematic of finding the ground in case of weak ground returns. The left panel shows a waveform with a weak ground signal but still strong enough to be easily distinguished while on the right, a case is shown where the exact ground localization is more subject to interpretation.

The LVIS data from Corcovado were collected in September 2015 at an altitude of approximately 12 km (40,000 ft) during a transit flight of the NSF/NCAR Gulfstream-V aircraft to Chile for the purpose of another mission. The swath width was 2.7 km. The LVIS data from La Selva were collected in March 2005 on board of the DOE King Air B-200 aircraft at an altitude of 10 km and with a swath width of 2 km. For both sites, the nominal footprint diameters were 25 m with 20 to 30-m spacing along and across the track. With a reported horizontal accuracy of around 0.1 m, the geolocation of the LiDAR footprints recorded by LVIS is very accurate (Blair et al., 1999). For each footprint, the following relative height metrics were retrieved from the LVIS datasets for the Corcovado site: RH10, RH15, RH20, RH25, …, RH95, RH96, RH97, RH98, RH99 and RH100 and the following for the La Selva site: RH25, RH50, RH75 and RH100. RH50 is equivalent to the HOME metric defined by Drake *et al.* (2002), who suggest that its position in the waveform is sensitive to changes in the degree of canopy openness, including tree density. We therefore tested their HTRT metric, which is simply the HOME divided by canopy height (i.e. RH50/RH100). For the Corcovado site, the footprint density of the LiDAR data is very low (on average 3 footprints per plot; see table S1 in supplementary material) and the field plots are not perfectly aligned in space with the LiDAR footprints. We tested therefore two sets of RH metrics for each plot: from all footprints that were contained within the plot or partially overlap with it and from which the waveform had a distinguishable ground return (see fig.3), we took either the maximum RH value or the average RH value weighted by the area of the overlapping footprints.

There is no LVIS data available for the temperate site in Sonoma County. For this site, two discrete return sensors (an ALS50 aboard a Cessna Grand Caravan and an ALS70 aboard a Piper PA-31 Navajo flying at 900 m above ground) collected wall to wall LiDAR across the area in 2014. The nominal pulse density was 10.66 pulses/m^2 at 105 kHz. The LiDAR point cloud was processed with the LAStools software (see https://rapidlasso.com/LAStools/). The LiDAR metrics were extracted over the field plots with a fixed 15-m radius by means of the tool “lasclip”. After classifying the ground points with the tool “lasground_new”, the height above the ground was computed for each non-ground point with “lasheight”. Finally, the forestry metrics (height percentiles “p10, p20, …, p90, p99” which are equivalent to the relative heights in full return LiDAR) were generated with “lascanopy” (Duncanson et al., in revision).

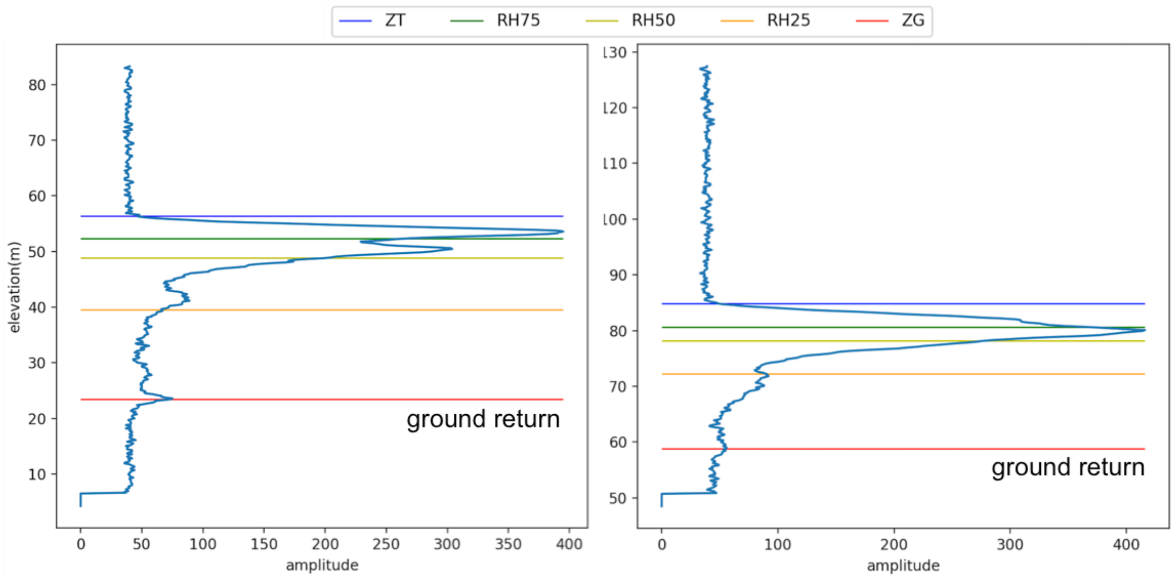

Fig.3. An illustration of the ground finding problematic in case of weak ground returns. On the left, the ground return can easily be distinguished while on the right, the placement of the ground return could be subject to interpretation. ZG: ground elevation; ZT: top canopy elevation; RH25, RH50, and RH75: LiDAR heights relative to the ground elevation, at which 25%, 50%, and 75% of the total reflected energy are returned to the sensor.

2.4 Field heights and biomass modeling

Before assessing the relationships between LiDAR metrics and field metrics, we examined if field metrics can predict plot-aggregated biomass by modeling biomass as a function of the following height metrics: mean tree height "Hmean", maximum tree height "Hmax" and dominant tree height "Hdom"[3]. We also tested the metric LH and the product of Hmax and BA "Hmax*BA" to investigate if model

[3] Mean height of the 100 largest trees per hectare (Rondeux, 1993).

performance improved when adding a factor that accounts for density (BA). At first sight, relating plot-aggregated biomass estimates and field metrics may appear circular because biomass densities were calculated using models based on tree measurements, which were then aggregated to the same plot level estimates. However, whereas biomass is estimated by applying tree allometry to all trees encountered in a plot, LiDAR-based models often apply "allometric" equations at the plot or stand level (the so-called plot-aggregated allometry defined by Asner and Mascaro (2014)). So, if forest structure and biomass do not follow consistent scaling patterns at plot level, we do not expect LiDAR-based "plot-aggregated allometries" to function properly either.

Regressions were performed within the R environment (R Core Team, 2016). Ordinary least squares regression was applied to model field heights as a function of LiDAR RH metrics at plot level (linear regression). To predict biomass either by field metrics or LiDAR metrics, power or exponential regression (depending on which performed better) using non-transformed variables was preferred over linear regression with log transformed variables. This avoids the bias associated with back transformation of the data. To compensate for heteroscedasticity of the residuals, we used weighted least squares (Power Variance Function in R).

We compared the model performances obtained with four different sets of plots: the Corcovado plots alone, the La Selva plots alone, all Costa Rican plots combined and the Sonoma plots alone. The Corcovado field data present challenges for LiDAR use because of the very high biomass conditions. In addition, the very dense forest canopy weakens the GPS signal causing the accurate geolocation of the field plots to be delicate, which can result in low spatial overlap between field plots and LiDAR footprints. Moreover, due to the small plot size there is an increased importance of the so-called edge effects. These effects are attributable to trees, that are not included in the field measurements, because they are located just outside the plot boundary, but have some portion of their crowns falling within the plot and therefore are measured by LiDAR (Frazer *et al.*, 2011).

In contrast, the La Selva plots are much bigger, thus minimizing potential edge effects, and situated in lower biomass conditions, providing a suitable point of comparison.

The Sonoma plots on the other hand are situated in a temperate site, where the smaller and more conical shaped canopies likely reduce the edge effects (pers. comm. Laura Duncanson). Moreover, the wall to wall LiDAR point cloud was extracted exactly over the field plots and by consequence, there is no possible geolocation error between both data sets.

2.5 Biomass estimation at swath scale

The model, which predicted plot level biomass with the best performance for the Corcovado data, is validated using the La Selva data. For both tropical sites, the model is then used to estimate biomass for each footprint composing the laser swaths (see fig.1). We then calculated the mean biomass of the forest at swath level, after filtering out the footprints that weren't situated in forested areas (footprints with a biomass value of 0 and in a non-forested location according to Google maps). For Corcovado, we considered only the footprints inside the Corcovado National

Park. Finally, we produced a biomass map by interpolating the biomass densities at footprint level to a raster of 25-m pixel resolution using the Natural Neighbor Interpolation tool of ArcGIS (3D Analyst Toolbox).

2.6 Model comparison

We compared the biomass densities of the La Selva plots predicted by our selected model with those predicted with the following model developed by Taylor *et al.* (2015) for the Osa Peninsula:

$$ACD = 3.8358\ (TCH)^{0.2807}\ (TCH * 0.6767)^{0.9721}\ (-0.0008 * TCH + 0.56)^{1.3763}$$

where ACD is aboveground carbon density (Mg ha^{-1}) and TCH is LiDAR derived top-of-canopy height (m), which is equivalent to RH100. This model is based on the general plot-aggregated allometry of Asner and Mascaro (2014) :

$$ACD = aTCH^{b1}\ BA^{b2}\ \rho_{BA}{}^{b3}$$

where BA is plot-averaged basal area (m^2 ha^{-1}) and ρ_{BA} is basal-area weighted WD. Neither BA nor ρ can be directly estimated using LiDAR, so Asner and Mascaro (2014) prescribe to develop regional relationships with TCH to replace these parameters, which was done by Taylor *et al.* (2015) for the Osa Peninsula.

We then convert the ACD in biomass by dividing it by 0.474, as in tropical forests 47.4% of biomass is carbon (Martin and Thomas, 2011).

3.0 Results

3.1 Field and biomass estimation

During the field surveys in Corcovado, researchers recorded 964 individual trees. Although they found 159 different species, almost 20 % of the measured trees belong to the following four species: *Simaba cedron* (Simaroubaceae), *Tetrathylacium macrophyllum* (Salicaceae), *Chrysochlamys glauca* (Clusiaceae) and *Nectandra umbrosa* (Lauraceae). In the much larger La Selva plots, 6240 trees were measured and 344 tree species were recorded. The four dominant species (35% of the total number) were *Pentaclethra macroloba* (Fabaceae), *Welfia regia* (Arecaceae), *Iriartea deltiodea* (Arecaceae) and *Socratea exorrhiza* (Arecaceae). During the Sonoma campaign, 1228 trees were measured. Although species information is unavailable, we know that the plots consisted of mostly hardwood tree species (51.7 % of the counts) and softwood tree species (45.6%). The remaining 2.7 % were non-tree species.

Table 1 summarizes the main plot parameters and the forest structural characteristics of the three sites. The two tropical sites have a similar forest structure with a well-developed sub-canopy and a relative small proportion of very large trees, which are generally a lot bigger in Corcovado (see fig. 4).

The Corcovado site has the highest biomass, with a large variability between plots (range: 66.9 – 1892.4 Mg ha^{-1}; see fig. 5). These extreme biomass densities are also a result of the small plot size (0.07ha) which induces high variability between plots and hence increases the biomass range across plots. The tree size distribution in fig. 6 shows that the great difference in biomass between the plots 9 (66.92 Mg ha^{-1}) and 8 (1892 Mg ha^{-1}), which are less than 50 m apart, is mainly due to the presence of 3 large trees in plot 8.

While the mean biomass for the other two field sites are equivalent, the Californian site shows much more variability than the La Selva site, as the field campaign involved different vegetation types (see fig. 5).

Table 1. Plot parameters and forest structural characteristics of the three field sites.

Site	Corcovado	La Selva	Sonoma
Number of plots	17	18	151
Number of measured trees	964	6240	1228
Plot size (ha)	0.07	0.5	variable radius plots
Forest type	Tropical wet forest	Tropical wet forest	Various temperate vegetation types
Trees per ha (DBH ≥ 10 cm)	459	510	-
Height (m)	13.8 ± 8.8 (mean ± sd[*1])	18.1 ± 5.5 (mean ± sd[*1])	-
Hmax (m)	55.1	47.7	78.8
QMSD (cm)	28.7 ± 9.3 (mean ± sd[*1])	25.0 ± 3.0 (mean ± sd[*1])	-
DBHmax (cm)	225	132.5	303.1
Basal area (m² ha^{-1})	53.6 ± 30.3 (mean ± sd[*1])	24.6 ± 3.0 (mean ± sd[*1])	35.5 ± 25.5 (mean ± sd[*1])
Biomass (Mg ha^{-1})	599.3 ± 462.5 (mean ± sd[*1])	241.8 ± 50.6 (mean ± sd[*1])	228.3 ± 172.3 (mean ± sd[*1])
Biomass range of plots (Mg ha^{-1})	66.9 - 1892.4 (min. - max.)	177.5 - 352.0 (min. - max.)	6.25 - 955.6 (min. - max.)
Data source	NGSFC[*2]	Clark & Clark (2000)	NGSFC[*2]

[*1] standard deviation; [*2] NASA Goddard Space Flight Center

3.2 Height modeling

We tested field heights versus LiDAR RH metrics for the Corcovado plots alone, for the La Selva plots alone and for all Costa Rican plots combined.

Huang et al. (2013) showed that RH metrics are highly correlated as they all are computed relative to the ground elevation. By consequence, only single term regression models could be developed. For further analysis, we used the maximum RH value of all footprints contained within or overlapping with the plots, as it generated better model performances than the area weighted average. Regarding the Corcovado plots, none of the tested relationships were significant (see table 2).

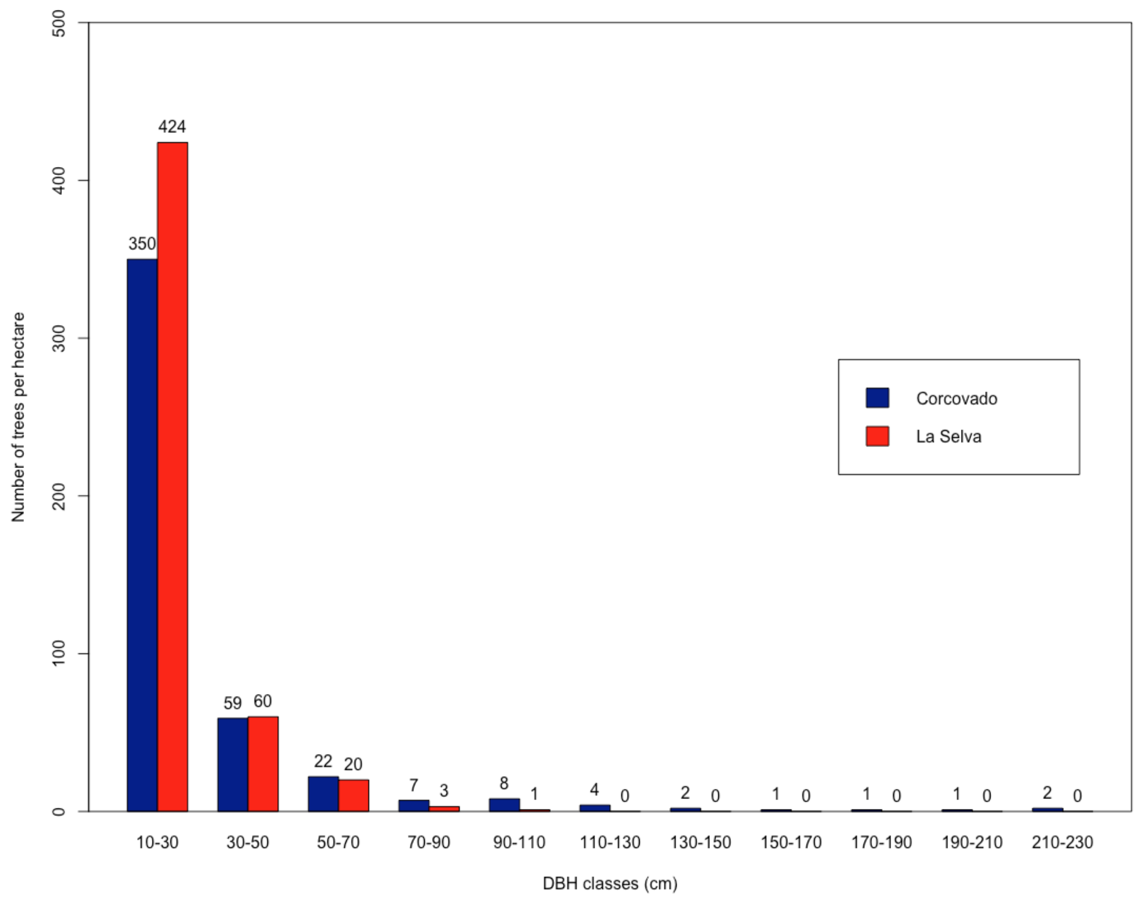

Fig. 4. Size class distribution of tree stems in Corcovado National Park and in La Selva Biological Station, Costa Rica. Stem density in Corcovado is 459 stems/ha and in La Selva is 510 stems/ha (DBH ≥ 10 cm).

Regression plots are shown in supplementary material (see figure S2). The La Selva plots, which are situated in forests of lower biomass, showed more encouraging results. Hmean and Hdom were accurately predicted by LiDAR metrics with more than 60% of the variance explained by the models and relative RMSE values ranging between 2 and 4 % (see table 2). However, we only have modeled heights for this site (tree heights were not recorded on field) and as such, these results should be taken with caution. The H-DBH model (Chave et al., 2014), which we used to compute the heights, has indeed a residual variance on its own. Yet, these models were not used for real predictions and were tested exclusively to allow comparison of performances with the other field sites.

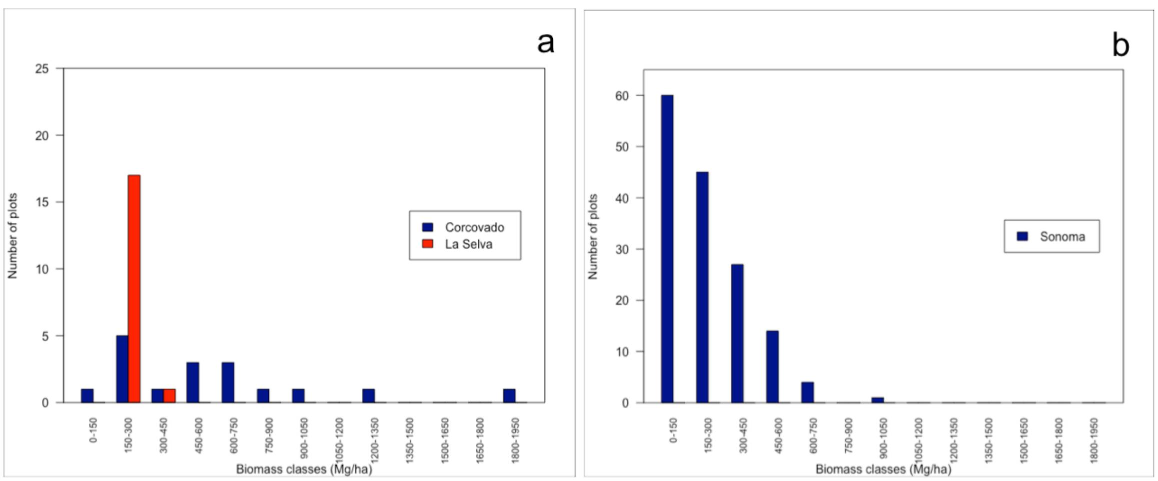

Fig. 5. Biomass variability across field plots in a) Corcovado National Park (blue bars) and La Selva Biological Station (red bars) and in b) Sonoma County.

The combination of all Costa Rican plots enhanced the performance of most models compared to the situation with the Corcovado plots alone. The R^2 values were all however beneath the 0.5.

For the Sonoma plots, as the height was available for only the three tallest trees, we only tested the regression of Hmax as a function of height percentiles (p10, p20, …, p90, p99). P99 generated the most significant regression model with a R^2 of 0.810 and a relative RMSE of 26.9%.

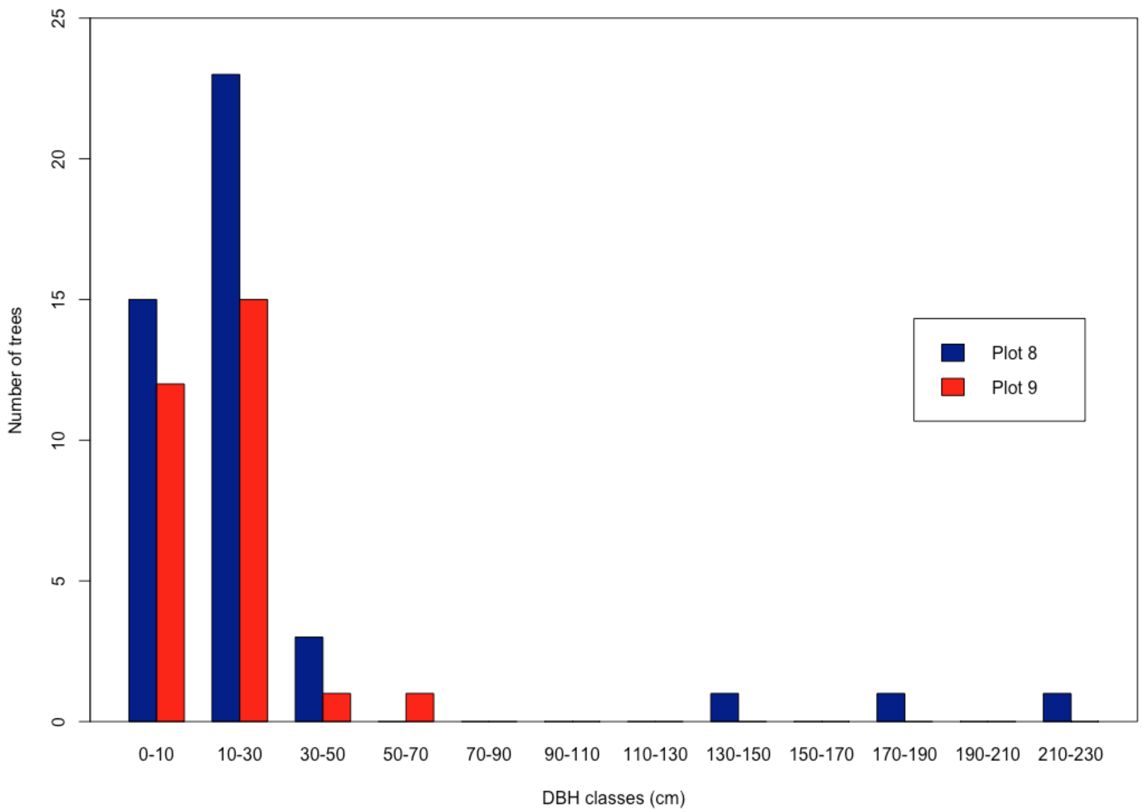

Fig 6. Tree size distribution of the two plots in the Corcovado National Park that show extreme biomass densities. Plot 9 (red bars) has the lowest biomass density (66.92 Mg ha^{-1}), while plot 8 (blue bars) has the highest biomass density (1892 Mg ha^{-1}). Small plot sizes (0.07 ha for Corcovado plots) make the impact of big trees more substantial, which cause high variability between plots and generate exploding biomass densities at the 1-ha scale.

Table 2. Summary of single term regression models between field (heights and biomass) and LiDAR metrics for the Corcovado plots alone, the La Selva plots alone, all Costa Rican plots combined and the Sonoma plots alone.

N°		1	2	3	4	5	6	7
Relationship		Hmean (m) ~ z * RH50 (m) + b	Hmax (m) ~ z * RH100 (m) + b	Hdom (m) ~ z * RH75 (m) + b	LH (m) ~ z * RH75 (m) + b	AGB (Mg/ha) ~ exp (b + z*RH75(1)) (m)	Hmax (m) ~ z* p99 (m) + b	AGB (Mg/ha) ~ (b*p70 (m))^z
Corcovado (N = 17)	AIC	67.2	123.6	112.1	119.8	**244.2**		
	R^2	0.048	0.102	0.098	0.184	**0.406**		
	RMSE (units)	1.46 (m)	7.68 (m)	5.49 (m)	6.88 (m)	**345.7 (Mg/ha)**		
	RMSE (%)	10.6	18.2	18.3	22.2	**57.7**		
	Bias (units)	-0.00 (m)	0.00 (m)	-0.00 (m)	0.00 (m)	**-110.6 (Mg/ha)**		
	p-value	0.399	0.212	0.222	0.086	**7.13e-08*** **		
La Selva (N = 18)	AIC	**31.5**	103.5	**54.0**	**74.8**	**189.5**		
	R^2	**0.636**	0.060	**0.611**	**0.486**	**0.412**		
	RMSE (units)	**0.49 (m)**	3.63 (m)	**0.92 (m)**	**1.64**	**37.72**		
	RMSE (%)	**2.69**	9.22	**3.34**	**6.29**	**15.6**		
	Bias (units)	**0.00 (m)**	-0.00 (m)	**-0.00 (m)**	**0.00 (m)**	**0.09**		
	p-value	**7.40e-05*** **	0.326	**1.27e-04*** **	**1.30e-03*** **	**0.004** **		
All Costa Rican plots (N = 35)	AIC	166.2	**230.7**	**200.6**	**218.5**	**485.9**		
	R^2	0.103	**0.126**	**0.192**	**0.316**	**0.536**		
	RMSE (units)	2.39 (m)	**6.00 (m)**	**3.90 (m)**	**5.04 (m)**	**246.6 (Mg/ha)**		
	RMSE (%)	14.8	**14.7**	**13.6**	**17.7**	**59.4**		
	Bias (units)	0.00 (m)	**0.00 (m)**	**0.00 (m)**	**0.00 (m)**	**-53.03 (Mg/ha)**		
	p-value	0.060	**0.036***	**0.008** **	**4.42e-04*** **	**3.00e-15*** **		
Sonoma (N = 151)	AIC						1017.5	**1859.5**
	R^2						**0.810**	**0.407**
	RMSE (units)						**6.89 (m)**	**132.1 (MG/ha)**
	RMSE (%)						**26.9**	**57.5**
	Bias (units)						**0.00 (m)**	**0.96 (Mg/ha)**
	p-value						**1.56e-55*** **	**6.90e-23*** **

N: number of samples. (1) RH70 for Corcovado. Bolded data are statistically significant models. Mean field-estimated biomass: 599.3 Mg ha^{-1} (Corcovado), 241.8 Mg ha^{-1} (La Selva), 415.4 Mg ha^{-1} (Costa Rica) and 228.3 Mg ha^{-1} (Sonoma). Different LiDAR metrics (RH metrics and RH50/RH100 ratio for the Costa Rican plots, height percentiles for Sonoma plots) were tested but only those which gave the most significant results are shown. * p-value < 0.05 ** < 0.01 *** < 0.001.

3.3 Biomass Density Modeling

To assess if forest structure and biomass follow consistent scaling patterns at plot level, we first examined how field metrics predict plot level biomass. Regression plots are shown in supplementary material (see figure S3).

For the Corcovado plots, the plot level biomass showed a good correlation with field metrics (overall mean R^2 of 0.694), although the relative RMSE were quite high (overall mean of 39.3%), which can be explained by the wide range of biomass values (see table 3). Of all models considering only a height factor (without BA, directly or through LH), the best performance was obtained with Hmax (R^2 of 0.730 and relative RMSE of 38.9 %). LH and BA are known to be good predictors of biomass (e.g. Saatchi *et al.,* 2011; Torres and Lovett, 2013) and the models which integrated these metrics gave indeed better performances (R^2 of 0.890 and 0.885 and relative RMSE of 24.9 and 25.3% respectively).

For the La Selva plots, we only tested the height metrics (Hmean, Hmax, Hdom and LH) against the plot level biomass (see table 3). The relative RMSE were much lower than for Corcovado, which can be explained by the lower biomass variability among plots (see fig. 5). Hdom predicted biomass with the best performance (R^2 of 0.843 and relative RMSE of 8.1 %), even better than LH, while Hmax scored less well (R^2 of 0.468 and relative RMSE of 14.8%). When taking all Costa Rican plots into account, Hmax performed again better than Hdom, although with a quite high RMSE (R^2 of 0.718 and relative RMSE of 46.3%).

For Sonoma, only Hmax could be computed as only the 1 to 3 tallest trees were measured. The model performance was mixed with a R^2 value of 0.396 and a relative RMSE of 57.7%.

Table 3. Summary of single term regression models to predict plot-aggregated biomass from field metrics for the Corcovado plots alone, the La Selva plots alone, all Costa Rican plots combined and the Sonoma plots alone.

N°	Relationship	b	z	AIC	R2	RMSE (Mg/ha)	RMSE (%)	Bias (Mg/ha)
	Corcovado (N = 17)							
8	AGB (Mg/ha) ~ (b*Hmean)^z (m)	0.249**	5.076***	246.4	0.575	292.4	48.8	-0.05
9	**AGB (Mg/ha) ~ (b*Hmax)^z (m)**	**0.119*****	**3.823*****	**228.7**	**0.730**	**233.2**	**38.9**	**-26.25**
10	AGB (Mg/ha) ~ (b*Hdom)^z (m)	0.209***	3.360***	233.0	0.391	350.3	58.4	-56.02
11	**AGB (Mg/ha) ~ (b*LH)^z (m)**	**0.223*****	**3.201*****	**211.0**	**0.890**	**149.1**	**24.9**	**-2.69**
12	**AGB (Mg/ha) ~ (b*(Hmax*BA)^z (m)**	**0.157*****	**1.070*****	**203.2**	**0.885**	**151.9**	**25.3**	**-8.78**
	La Selva (N = 18)							
13	AGB (Mg/ha) ~ (b*Hmean)^z (m)	0.258*	3.531***	186.0	0.508	34.5	14.3	-2.16
14	AGB (Mg/ha) ~ exp (b + z*Hmax) (m)	3.804***	0.042***	185.1	0.468	35.9	14.8	-3.01
15	**AGB (Mg/ha) ~ (b*Hdom)^z (m)**	**0.164*****	**3.633*****	**152.3**	**0.843**	**19.5**	**8.1**	**-0.65**
16	**AGB (Mg/ha) ~ (b*LH)^z (m)**	**0.580***	**2.019*****	**173.2**	**0.765**	**23.9**	**9.9**	**0.02**
	All Costa Rican plots (N = 35)							
17	AGB (Mg/ha) ~ (b*Hmean)^z (m)		no significant relation and R2 under 0					
18	**AGB (Mg/ha) ~ exp (b + z*Hmax) (m)**	**2.090*****	**0.092*****	**450.5**	**0.718**	**192.3**	**46.3**	**-2.23**
19	AGB (Mg/ha) ~ (b*Hdom)^z (m)	0.170***	3.667***	447.0	0.472	263.0	63.3	-41.56
20	**AGB (Mg/ha) ~ (b*LH)^z (m)**	**0.214*****	**3.234**	**399.7**	**0.904**	**112.4**	**27.1**	**-9.28**
	Sonoma (N = 151)							
21	AGB (Mg/ha) ~ (b*Hmax)^z (m)	23.011	0.854***	1837.5	0.396	133.9	57.7	0.38

N: number of samples. Bolded data are models with best performance. Mean field-estimated biomass: 599.3 Mg ha^{-1} (Corcovado), 241.8 Mg ha^{-1} (La Selva), 415.4 Mg ha^{-1} (Costa Rica) and 228.3 Mg ha^{-1} (Sonoma) * p-value < 0.05 ** p-value < 0.01 *** < 0.001.

We applied power or exponential regression (depending on which performed better) between field biomass and LiDAR metrics (RH metrics and RH50/RH100 ratio for Costa Rican plots, height percentiles for Sonoma plots) to avoid bias associated with back transformation of the data. Regression plots are shown in supplementary material (see figure S2). Unlike for the field heights, some LiDAR RH metrics (e.g. RH70) did significantly predict field biomass when considering the Corcovado plots alone (p-value < 0.05) but with weak model performances (R^2 = 0.406 and relative RMSE = 57.7%; see table 2). Adding the La Selva plots resulted in a R^2 value above 0.5 and a mean bias decreased by half. However, the relative RMSE was almost 60% because of a lower mean biomass for all Costa Rican plots combined compared to the Corcovado plots alone.

Considering only the La Selva plots, biomass was significantly predicted by RH75 with a R^2 value of 0.414 and a relative RMSE slightly above 15%.

For the Sonoma plots, the models between biomass and the height percentiles had moderate performances, the best being the power model between biomass and p70 with an R^2 of 0.502 and a relative RMSE of 59.6% (see table 2).

3.4 Biomass estimation at the swath scale

Some LiDAR relative heights significantly predicted plot level biomass for the Corcovado plots but with weak model performances (see table 2). Considering the field-based models, Hmax performed quite well (see relationship n° 9 in table 3). Given its close relation with RH100 reported in the literature (see section 4.1 below), we used the model that related Hmax to biomass for further analysis.

The model was evaluated using the La Selva plots. Fig. 7a shows the scatter plot of predictions against field-estimated biomass densities, which indicates that the model tends to underestimate the biomass densities, at least for lower values. The relative RMSE is 30.1 %.

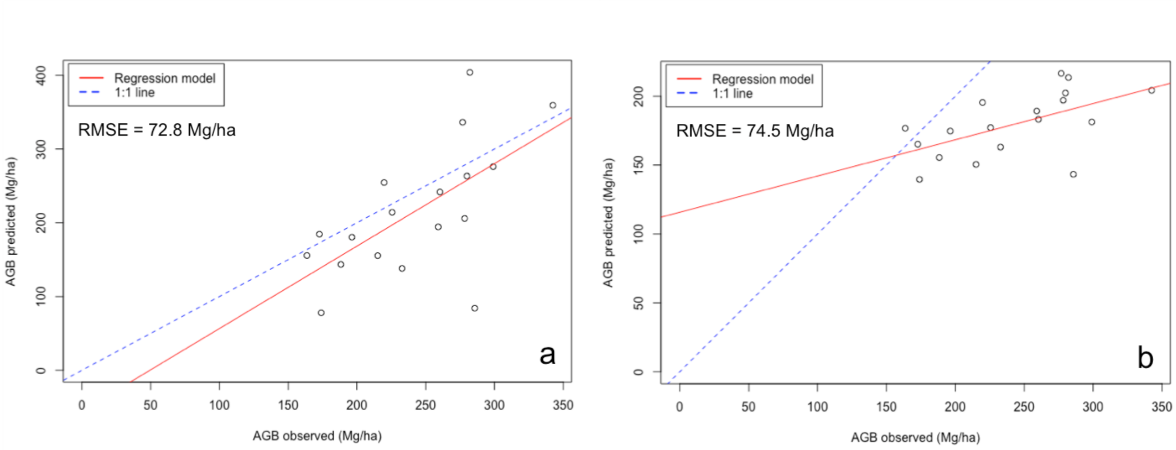

Fig. 7. Estimated values of aboveground biomass (AGB) of the La Selva plots using (a) the selected model based only on the field tree maximum height “Hmax” and (b) the regional model of Taylor *et al.* (2015) based on the LiDAR height “RH100” and other metrics that account for density (BA and WD), compared to plot-estimated AGB, using field measurements and allometric equations. The blue dashed one-to-one line is provided for reference.

The model was applied to each footprint of the laser swaths by substituting Hmax with RH100. The resulting mean forest biomass densities are 281.5 ± 299.4 Mg ha^{-1} (mean ± sd) for Corcovado and 194.8 ± 180.3 Mg ha^{-1} (mean ± sd) for La Selva. Fig. 8 shows the biomass map created by interpolation of the footprint level biomass densities.

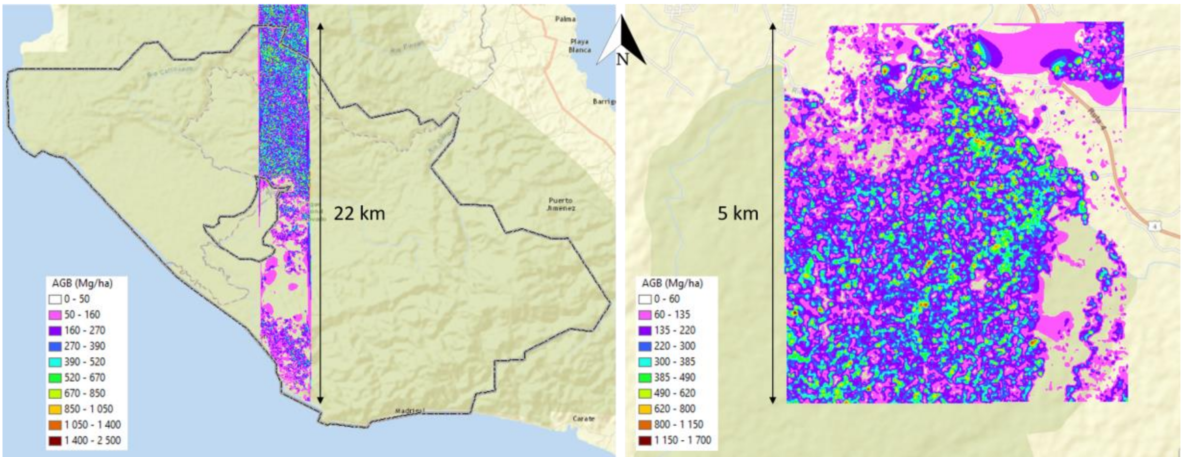

Fig. 8. Biomass maps (25-m pixel resolution) at laser swath level for the Corcovado National Park (on the left) and the La Selva Biological Station (on the right). As the laser swath for Corcovado cover the whole width of the Osa peninsula, we only show the part in the Corcovado National park for visualization purposes. The grey dashed line is the park boundary. Empty spots on the maps result from the footprint filtering process (removal of the footprints that were reflected off clouds or that did not have a ground return).

3.5 Model comparison

Our model for biomass prediction includes only Hmax, which accounts for the vertical distribution of biomass, but ignores its spatial distribution within plots (Duncanson *et al.*, 2015). The model developed by Taylor *et al.* (2015) includes terms, that account for stand and tree density (BA and WD), by means of relationships with TCH (equivalent to RH100). When we used this model to predict biomass for the La Selva plots, the results were not better than those found with our single term model (relative RMSE of 30.8%; see fig. 7b). The model underestimates rather strongly the biomass densities and the trend worsens with increasing biomass values. This can be explained by the bad correlation, which is observed between TCH (RH100) and BA for the La Selva plots (R^2 = 0.292 and relative RMSE of 10.0 %; see fig. S4 in supplementary material), whereas the modeling approach of Asner and Mascaro (2014) assumes a linear relationship between both metrics.

4.0 Discussion

4.1 Relationship between field heights and LiDAR RH metrics

LiDAR metrics have proven capable of predicting canopy height by using both single term linear regression models (Means *et al.*, 1999; Anderson *et al.*, 2006; Park *et al.*, 2014) and regression models relying on multiple LiDAR derived variables (Hyde *et al.*, 2006; Duncanson, Niemann and Wulder, 2010). The simpler single term models often use RH100 to predict field maximum height (Hmax). For example, Anderson *et al.* (2006) found strong agreement between RH100 from LVIS and Hmax (R^2 = 0.80; RMSE = 3.49 m; relative RMSE = 13.2 %) in a forest of central New Hampshire (USA). Park *et al.* (2014) showed also for various North American forests, that LVIS RH100 can satisfactorily provide a proxy for forest canopy heights (R^2 = 0.78; RMSE = 4.99 m; relative RMSE = 11.2 %). Our findings in Sonoma support these results as we obtained a good model performance between Hmax and the LiDAR 99th height percentile, which is the discrete return equivalent of the full waveform RH100 metric (R^2 = 0.810; relative RMSE = 26.9 %; see relationship n° 6 in table 2).

The fact that no good relationships were found between field heights and LiDAR RH metrics for the Corcovado plots is mostly due to the low spatial overlap between the field plots and the LiDAR footprints. Even when the plots and footprints perfectly overlap, geolocation errors occur, especially because dense forest canopies block or scatter the GPS signal, which makes it difficult to locate field plots with planimetric accuracies better than 1-5 m (Frazer *et al.*, 2011). The geolocation of the LiDAR footprints is much more accurate as, for LVIS, the horizontal accuracy is reported to be around 0.1 m (Blair, Rabine and Hofton, 1999). A simulation study by Frazer *et al.* (2011) analyzed the impact of GPS errors from 1 to 6 m on goodness-of-fit statistics (R^2 and RMSE) for plots from 300 to 1300m^2. For plots of 15-m radius like the Corcovado plots, an increase in the GPS error from 3 to 6 m resulted in a decline of 1.4 % in median R^2 and an increase of 9.1 % in median RMSE. These are acceptable results, although the range of values associated with each of these two fit statistics also increased markedly with increasing GPS error (see fig. 8 in Frazer *et al.* 2011). For the Corcovado plots, the GPS precision is estimated at 1.4 ± 0.7 m (mean ± sd), which is not bad for a very dense forest. However, the spatial overlap between the field plots and the LiDAR footprints is low because of the small plot size and the low footprint density (on average 3 footprints per plot; see table S1 in supplementary material). The spacing between footprints is often greater than the footprint diameter, which limits overlap between shots. In addition, there were few overlapping flight lines and a lot of laser shots were eliminated, either because they were reflected off clouds or because of the absence of ground return. This results in a quite large distance between the footprint and plot centroids, with the minimum distance averaged over the plots being 11.3 ± 3.5 m (mean ± sd). For plot sizes of about the size of only one footprint, this distance should be restricted to a maximum of 3 m (pers. comm. Steven Hancock), as otherwise the edge effects become too strong. Hmax is especially sensitive to edge effects as it corresponds to a single tree and can easily be missed in case of limited overlap between footprints and field plots.

By comparison, in the La Selva site, the minimum distance between plot and footprint centroids is 5.0 ± 2.5 m (mean ± sd), although with plot sizes of 0.5 ha and a higher footprint density, the impact of edge effects and positional errors is dampened (Frazer *et al.*, 2011).

4.2 Prediction of plot level biomass

For our three field sites, our results indicate that single term models based on LiDAR height metrics fail to provide accurate predictions of plot level biomass even though the relationships are significant. This cannot solely be attributed to the low co-registration of the field plots and LiDAR footprints as we obtained similar model performances for all three sites. Adding terms accounting for density (BA and WD), did not improve plot level biomass predictions for La Selva (see fig. 7b). However, this was likely caused by a bad correlation between RH100 and BA (see fig. S4), while a good correlation between both metrics is a prerequisite for using the modeling approach of Asner and Mascaro (2014).

On the other hand, our Corcovado data indicate that the field maximum height (Hmax) is a good predictor of plot level biomass. The fact that Hmax did not predict biomass equally well for the La Selva plots is probably related to the much larger plot size. Hmax corresponds to a single tree and, with 250 trees on average per plot, this metric does not relate well to plot-aggregated biomass. Not surprisingly, Hdom, which is averaged on 50 trees, is a more suitable candidate. Hmax performs also less well for the Sonoma plots but as it is a temperate site with variable radius plots, allometric relations at plot level are logically very different.

As Hmax is a good predictor of plot level biomass and given its close relation with RH100, it seems reasonable to apply model n° 9 (see table 3) to all footprints of the swaths by substituting Hmax with RH100. We obtained mean biomass densities at swath level of 281.5 Mg ha^{-1} for Corcovado and 194.8 Mg ha^{-1} for La Selva. A study by Malhi *et al.* (2006), based on data from 227 plots of 0.8 - 22.5 ha, found that the regional mean biomass in South American forests varied between 200 and 350 Mg ha^{-1}. Our mean biomass value for the Corcovado National Park falls inside that range, while the value for the La Selva Biological Station is very close to the lower limit. Although Taylor *et al.* (2015) found a lower biomass density for the Osa Peninsula (mean of 150 - 200 Mg ha^{-1}, depending on the soil type), we have shown that their model tends to underestimate plot level biomass (see fig. 7b).

Our interpolated biomass maps (see fig. 8) show broad biomass ranges (0 – 2500 Mg ha^{-1} for Corcovado and 0 – 1700 Mg ha^{-1} for La Selva), although the highest biomass levels only occur very locally, usually over only one or a few pixels. These extreme biomass densities are a result of the small plot size and in reality, few forests support biomass densities this high except over very small areas (Zolkos, Goetz and Dubayah, 2013). Some studies (Réjou-Méchain *et al.*, 2015; Kim *et al.*,

2016) showed that, in the same study site, a smaller plot size results in a higher standard deviation (sd) of plot biomass and thus in increased biomass ranges. For example, in a recent study, which used discrete return LiDAR to estimate biomass of a lowland rainforest in Brunei Darussalam, the biomass range in 20 x 20-m and 30 x 30-m plots were 77.4 - 904.6 Mg ha^{-1} and 154.1 – 585.9 Mg ha^{-1} respectively, while the average remained equivalent (313.8 Mg ha^{-1} for 20-m plots vs. 302.7 Mg ha^{-1} for 30-m plots; Kim *et al.*, 2016). As plot size decreases, large trees have a greater impact and cause high variability between plots, which generates exploding biomass densities at the 1-ha scale (see fig. 6). Increasing the size of the field plots will therefore reduce this artefact. Indeed, when Drake *et al.* (2002) used 1998 LVIS data for the permanent 0.5 ha field plots of the Carbono project (see section 2.1), they predicted biomass values for La Selva with a much narrower range (0-300 Mg ha^{-1}).

Furthermore, through its effect on the biomass variability across plots (see fig. 5), plot size has a major impact on model errors (Mascaro *et al.*, 2011; Zolkos, Goetz and Dubayah, 2013). Although the R^2 value of the models using LiDAR metrics to predict biomass is around 0.4 for both tropical sites, the RMSE is much higher for Corcovado which has a smaller plot size (see equation n°5 in table 2). Zolkos, Goetz and Dubayah (2013) showed that RMSE values decrease with increasing plot size and estimated that a minimum plot size of approximately 0.2 ha is required to achieve biomass prediction accuracies of 20 %. On the other hand, the study by Kim *et al.* (2016) showed that relative errors below 20% are achievable with much smaller plot sizes. Using four LiDAR metrics to predict biomass, they found that plots with sizes of 30m x 30m (0.09 ha) allowed for much more accurate predictions than plots with sizes of 20m x 20m (0.04 ha) with a relative RMSE value of 11.6% and 34.3% respectively (Kim *et al.*, 2016). However, considering our results, we argue that in forests with large trees as in Corcovado or in Sonoma, the plot size should be at least 0.2 ha, to limit biomass variability between plots and to avoid exploding biomass densities at higher scales.

Finally, plot size also affects the uncertainty of the field-based biomass estimates at plot level computed by the pantropical allometric model of Chave *et al.* (2014). The model developers argue that tree-level uncertainty in biomass estimation from their model is about 50% of the mean but that at plot level, uncertainty drops to ca. 5-10% for a 1-ha plot. When we apply equation 8 in Chave *et al.* (2014) to the Corcovado plots, mean plot-level uncertainty scores about 23 %. In contrast, the uncertainty for the La Selva plots is around 8 %. These results show that the allometric model has a considerable bias when applied to small plot sizes.

4.3 Recommendations for future studies

Our findings suggest that small field plots are not suitable for biomass estimation in tropical forests. Even if the spatial overlap is perfect between field plots and LiDAR footprints, a small plot size will introduce high variability in biomass densities between plots and will lead to large model errors (RMSE). To minimize this effect, we prescribe using plots with a minimum size of 0.2 ha. If smaller plots are chosen (same size as the footprint for example), it is imperative that the plot centroids perfectly match with those of the footprints.

Also, geolocation of the measured trees would enable one to generate models at footprint level instead of solely at plot level.

Finally, further studies should investigate whether introducing other LiDAR metrics improves model performances, e.g. canopy cover (Hyde *et al.*, 2006) for biomass prediction, waveform extent (Lefsky *et al.*, 2007; Duncanson, Niemann and Wulder, 2010; Lefsky, 2010) or energy quartiles, i.e. the proportion of energy in four equal elevation divisions of the waveform (Duncanson, Niemann and Wulder, 2010), for canopy height prediction.

5.0 Conclusion

Even though LiDAR is thought not to saturate at higher biomass, areas of dense canopy cover and of very high biomass as found in Corcovado, may sometimes present a challenge for LiDAR.

In conclusion, we address the three objectives as set out in the introductory section.

We first calculated plot level biomass for which we found that, while Hmax is a good predictor in case of small plot size, Hdom and even Hmean seem to outperform Hmax as plot size increases. This finding is important to consider when planning future field campaigns, as not all field designs (e.g. Sonoma in our study) allow calculation of Hdom or Hmean.

Upon assessing how LiDAR performs in high biomass tropical forests, we concluded that perfect spatial overlap is very important in case of small plot size. We suggest to consider a plot size of at least 0.2-ha, as smaller plot sizes introduce high biomass variability between plots and lead to considerable model errors. Larger plot sizes will also diminish the plot-level uncertainty on field-based biomass estimates.

Finally, our swath level biomass values are within the range of values reported by other studies in the Neotropics. It should be noted that the associated uncertainty is about 30% and our results should therefore be taken with caution.

6.0 Acknowledgements

I gratefully acknowledge Lola Fatoyinbo for enabling me to conduct this research and for welcoming me at Nasa Goddard Space Flight Center. I cordially thank Philippe Lejeune for his review and wise suggestions. I thank for field data collection in Corcovado National Park: Lola Fatoyinbo, Amanda Armstrong, Seung Kuk Lee, Paul Montesano, Naiara Pinto and Guoqing Sun. I also thank members of the Carbono project for the field data collected in La Selva Biological Station and Laura Duncanson for providing the Sonoma data. I kindly acknowledge Michelle Hofton and Brian Blair for processing the LVIS data. I show gratitude to the department of Geographical Sciences at the University of Maryland, which is under the supervision of Ralph Dubayah, and especially to Steve Hancock for patiently answering my numerous queries. I am finally grateful to Watna Horemans for introducing me to Lola Fatoyinbo and to Daphnis De Pooter for proofreading this article several times.

7.0 References

Anderson, J., Martin, M.E., Smith, M-L., Dubayah, R.O., Hofton, M.A., Hyde, P., Peterson, B.E., Blair, J.B., Knox, R.G. (2006) 'The use of waveform lidar to measure northern temperate mixed conifer and deciduous forest structure in New Hampshire', *Remote Sensing of Environment*, 105(3), pp. 248–261. doi: 10.1016/j.rse.2006.07.001.

Ankersen, T.T., Regan, K.E. and Mack, S.A. (2006) 'Towards a bioregional approach to tropical forest conservation: Costa Rica's Greater Osa Bioregion', *Futures*, 38(4), pp. 406–431. doi: 10.1016/j.futures.2005.07.017.

Asner, G.P. and Mascaro, J. (2014) 'Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric', *Remote Sensing of Environment*. Elsevier Inc., 140, pp. 614–624. doi: 10.1016/j.rse.2013.09.023.

Blair, J., Rabine, D. and Hofton, M. (1999) 'The Laser Vegetation Imaging Sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography', *ISPRS Journal of Photogrammetry and Remote Sensing* 54 1999., pp. 115–122. PII: S0924- 2716Ž99.00002-7.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., & Vieilledent, G. (2014) 'Improved allometric models to estimate the aboveground biomass of tropical trees', *Global Change Biology*, 20(10), pp. 3177–3190. doi: 10.1111/gcb.12629.

Clark, D.B., and Clark, D.A. (2000) 'Landscape-scale variation in forest structure and biomass in a tropical rain forest', *Forest Ecology and Management*, 137, pp. 185–198. Available at: dbclark@sloth.ots.ac.cr.

Cornejo, X., Mori, S. A., Aguilar, R., Stevens, H. and Douwes, F. (2012) 'Phytogeography of the trees of the Osa Peninsula, Costa Rica', *Brittonia*, 64(1), pp. 76–101. doi: 10.1007/s12228-011-9194-0.

Curtis, R. P. and Marshall, D. D. (2000) 'Why quadratic mean diameter?', *Western Journal of Applied Forestry*, 15(360), pp. 137–139.

Dashora, A., Lohani, B. and Deb, K. (2013) 'Two-step procedure of optimisation for flight planning problem for airborne LiDAR data acquisition', *International Journal of Mathematical Modelling and Numerical Optimisation*, 4(4), pp. 323–350. doi: 10.1504/IJMMNO.2013.059194.

Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F., & Prince, S. (2002) 'Estimation of tropical forest structural characteristics, using large-footprint lidar', *Remote Sensing of Environment*, 79(2–3), pp. 305–319. doi: 10.1016/S0034-4257(01)00281-4.

Drake, J. B., Knox, R.G., Dubayah, R.O., Clark, D.B., Condit, R., Blair, J.B., & Hofton, M. (2003) 'Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships', *Global Ecology & Biogeography*, 12(2), pp. 147–159.

Dubayah, R.O., Sheldon, S.L., Clark, D.B., Hofton, M.A., Blair, J.B., Hurtt, G.C., and Chazdon, R.L. (2010) 'Estimation of tropical forest height and biomass dynamics using lidar remote sensing at la Selva, Costa Rica', *Journal of Geophysical Research: Biogeosciences*, 115(2), pp. 1–17. doi: 10.1029/2009JG000933.

Dubayah, R.O. (2015). Crowd-Sourced Calibration: The GEDI Strategy for Empirical Biomass Estimation Using Spaceborne Lidar. American Geophysical Union, Fall Meeting 2015, abstract #B51I-02.

Duncanson, L. I., Niemann, K. O. and Wulder, M. A. (2010) 'Estimating forest canopy height and terrain relief from GLAS waveform metrics', *Remote Sensing of Environment*. Elsevier Inc., 114(1), pp. 138–154. doi: 10.1016/j.rse.2009.08.018.

Duncanson, L. I., Dubayah, R.O. and Enquist, B.J. (2015) 'The importance of spatial detail: Assessing the utility of individual crown information and scaling approaches for lidar-based biomass density estimation', *Remote Sensing of Environment*. Elsevier Inc., 168, pp. 102–112. doi: 10.1016/j.rse.2015.06.021.

Duncanson, L.I., Huang, W., Johnson, K., Swatantran, A., McRoberts, R. and Dubayah, R.O. *In Revision*. Implications of Allometric Model Selection for County-Level Biomass Estimates. *Scientific Reports.*

Fatoyinbo, T. E. and Simard, M. (2013) 'Height and biomass of mangroves in Africa from ICESat/GLAS and SRTM', *International Journal of Remote Sensing*, 34(2), pp. 668–681. doi: 10.1080/01431161.2012.712224.

Frazer, G. W., Magnussen, S., Wulder, M.A. and Niemann, K.O. (2011) 'Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass', *Remote Sensing of Environment*. Elsevier Inc., 115(2), pp. 636–649. doi: 10.1016/j.rse.2010.10.008.

Goetz, S. and Dubayah, R. (2011) 'Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change', *Carbon Management*, 2(April), pp. 231–244. doi: 10.4155/cmt.11.18.

Hancock, S., Anderson, K., Disney, M. and Gastona, K.J. (2017) ‘Measurement of fine-spatial-resolution 3D vegetation structure with airborne waveform lidar: Calibration and validation with voxelised terrestrial lidar’, *Remote Sensing of Environment*. The Authors, 188, pp. 37–50. doi: 10.1016/j.rse.2016.10.041.

Hofton, M. A., Minster, J. B. and Blair, J. B. (2000) ‘Decomposition of laser altimeter waveforms’, *IEEE Transactions on Geoscience and Remote Sensing*, 38(4 II), pp. 1989–1996. doi: 10.1109/36.851780.

Holdridge, L. R. (1967) ‘Life zone ecology’, p. 206. doi: Via 10.1046/j.1365-2699.1999.00329.x.

Huang, W., Sun, G., Dubayah R.O., Cook, B., Montesano, P., Ni, W. and Zhang, Z. (2013) ‘Mapping biomass change after forest disturbance: Applying LiDAR footprint-derived models at key map scales’, *Remote Sensing of Environment*. Elsevier Inc., 134, pp. 319–332. doi: 10.1016/j.rse.2013.03.017.

Hyde, P., Dubayah, R.O., Walker, W., Blair, J.B., Hofton, M. and Hunsaker, C. (2006) ‘Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy’, *Remote Sensing of Environment*, 102(1–2), pp. 63–73. doi: 10.1016/j.rse.2006.01.021.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., and Birdsey, R.A. (2003) 'National scale biomass estimates for United States tree species', *Forest Science*, 49(1), pp. 12–32.

Kim, E., Lee, W.-K., Yoon, M., Lee, J.-Y., Son, Y. and Salim, K.A. (2016) 'Estimation of Voxel-Based Above-Ground Biomass Using Airborne LiDAR Data in an Intact Tropical Rain Forest, Brunei', *Forests*, 7(12), p. 259. doi: 10.3390/f7110259.

Lefsky, M.A., Keller, M., Pang, Y., de Camargo, P.B. and Hunter, M.O. (2007) 'Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms', *Journal of Applied Remote Sensing*, 1(1), p. 13537. doi: 10.1117/1.2795724.

Lefsky, M.A. (2010) 'A global forest canopy height map from the moderate resolution imaging spectroradiometer and the geoscience laser altimeter system', *Geophysical Research Letters*, 37(15), pp. 1–5. doi: 10.1029/2010GL043622.

Le Toan, T., Quegan, S. , Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H. and Ulander, L. (2011) 'The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle', *Remote Sensing of Environment*. Elsevier Inc., 115(11), pp. 2850–2860. doi: 10.1016/j.rse.2011.03.020.

Magruder, L., Neuenschwander, A.L. and Marmillion, S.P. (2010) 'Lidar waveform stacking techniques for faint ground return extraction', *Journal of Applied Remote Sensing*, 4(1), p. 43501. doi: 10.1117/1.3299657.

Malhi, Y., Wood, D., Baker, T.R., Wright, J., Phillips, O.L., Cochrane, T., Meir, P., Chave, J., Almeida, S., Arroyo, L., Higuchi, N., Killeen, T.J., Laurance, S.G., Laurance, W.F., Lewis, S.L., Monteagudo, A., Neill, D.A., Vargas, P.N., Pitman, N.C.A., Quesada, C.A., Salomão, R., Silva, J.N.M., Lezama, A.T., Terborgh, J., Martínez, R.V., and Vinceti, B. (2006) 'The regional variation of aboveground live biomass in old-growth Amazonian forests', *Global Change Biology*, 12(7), pp. 1107–1138. doi: 10.1111/j.1365-2486.2006.01120.x.

Mallet, C. and Bretar, F. (2009) 'Full-waveform topographic lidar: State-of-the-art', *ISPRS Journal of Photogrammetry and Remote Sensing*. International Society for Photogrammetry and Remote Sensing, Inc. ISPRS, 64(1), pp. 1–16. doi: 10.1016/j.isprsjprs.2008.09.007.

Martin, A. R. and Thomas, S. C. (2011) 'A reassessment of carbon content in tropical trees', *PLoS ONE*, 6(8). doi: 10.1371/journal.pone.0023533.

Mascaro, J., Detto, M., Asner, G.P. and Muller-Landau, H.C. (2011) 'Evaluating uncertainty in mapping forest carbon with airborne LiDAR', *Remote Sensing of Environment*. Elsevier Inc., 115(12), pp. 3770–3774. doi: 10.1016/j.rse.2011.07.019.

Means, J.E., Acker, S.A., Harding, D.J., Blair, J.B., Lefsky, M.A., Cohen, W.B., Harmon,M.E. and McKee, W.A. (1999) ‘Use of Large-Footprimt Scanning Airborne Lidar to Estimate Forest Stand Characteristics in the Western Cascades of Oregeon’, *Remote Sensing Enivornment*, 308(67), pp. 298–308. doi: 10.1016/S0034-4257(98)00091-1.

Mitchard, E.T.A., Saatchi, S.S., White, L.J.T. , Abernethy, K.A., Jeffery, K.J., Lewis, S. L., Collins, M., Lefsky, M.A., Leal, M.E., Woodhouse, I.H. and Meir, P. (2012) ‘Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: Overcoming problems of high biomass and persistent cloud’, *Biogeosciences*, 9(1), pp. 179–191. doi: 10.5194/bg-9-179-2012.

Mountrakis, G. and Li, Y. (2017) ‘A linearly approximated iterative Gaussian decomposition method for waveform LiDAR processing’, *ISPRS Journal of Photogrammetry and Remote Sensing*. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS), 129, pp. 200–211. doi: 10.1016/j.isprsjprs.2017.05.009.

Park, T., Kennedy, R.E., Choi, S., Wu, J., Lefsky, M.A., Bi, J., Mantooth, J.A., Myneni, R.B. and Knyazikhin, Y. (2014) ‘Application of physically-based slope correction for maximum forest canopy height estimation using waveform lidar across different footprint sizes and locations: Tests on LVIS and GLAS’, *Remote Sensing*, 6(7), pp. 6566–6586. doi: 10.3390/rs6076566.

Pirotti, F. (2011) 'Analysis of full-waveform LiDAR data for forestry applications: A review of investigations and methods', *IForest*, 4(JUNE), pp. 100–106. doi: 10.3832/ifor0562-004.

Qi, W. and Dubayah, R. O. (2016) 'Combining Tandem-X InSAR and simulated GEDI lidar observations for forest structure mapping', *Remote Sensing of Environment*. Elsevier Inc., 187(2016), pp. 253–266. doi: 10.1016/j.rse.2016.10.018.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Réjou-Méchain, M., Tymen, B., Blanc, L., Fauset, S., Feldpausch, T.R., Monteagudo, A., Phillips, O.L., Richard, H. and Chave, J. (2015) 'Using repeated small-footprint LiDAR acquisitions to infer spatial and temporal variations of a high-biomass Neotropical forest', *Remote Sensing of Environment*. Elsevier Inc., 169, pp. 93–101. doi: 10.1016/j.rse.2015.08.001.

Réjou-Méchain, M., Tanguy, A., Piponiot, C., Chave, J. & Hérault, B. (2016). "BIOMASS" package: Estimating Aboveground Biomass and Its Uncertainty in Tropical Forests. Version 1.1. Date: 2017-01-03.

Rondeux, J. (1993).'La mesure des arbres et des peuplements forestiers', Les Presses Agronomiques De Gembloux, a.s.b.l., D/1999/1665/7.

Saatchi, S.S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L. and Hagen, S. (2011) 'Benchmark map of forest carbon stocks in tropical regions across three continents', *Proc. Natl. Acad. Sci. USA*, 108, pp. 9899–9904. doi: 10.1073/pnas.1019576108.

Taylor, P., Asner, G., Dahlin, K., Anderson, C., Knapp, D., Martin, R., Mascaro, J., Chazdon R., Cole, R., Wanek, W., Hofhansl, F., Malavassi, E., Vilchez-Alvarado, B. and Townsend, A. (2015) 'Landscape-scale controls on aboveground forest carbon stocks on the Osa Peninsula, Costa Rica', *PLoS ONE*, 10(6), pp. 1–18. doi: 10.1371/journal.pone.0126748.

Torres, A.B. and Lovett, J.C. (2013) 'Using basal area to estimate aboveground carbon stocks in forests: La Primavera Biosphere's Reserve, Mexico', *Forestry*, 86(2), pp. 267–281. doi: 10.1093/forestry/cps084.

Zolkos, S.G., Goetz, S.J. and Dubayah, R.O. (2013) 'A meta-analysis of terrestrial aboveground biomass estimation using lidar remote sensing', *Remote Sensing of Environment*. Elsevier Inc., 128, pp. 289–298. doi: 10.1016/j.rse.2012.10.017.