Article

Regional Scale Rain-Forest Height Mapping Using Regression-Kriging of Spaceborne and Airborne LiDAR Data: Application on French Guiana

Ibrahim Fayad [1,*, Nicolas Baghdadi [1], Jean-Stéphane Bailly [2], Nicolas Barbier [3], Valéry Gond [4], Bruno Hérault [5], Mahmoud El Hajj [6], Frédéric Fabre [7] and José Perrin [8]**

[1]   IRSTEA, UMR TETIS, 500 rue Jean François Breton, 34093 Montpellier Cedex 5, France;
      nicolas.baghdadi@teledetection.fr
[2]   AgroParisTech, UMR LISAH, 2 Place Pierre Viala, 34060 Montpellier, France; bailly@agroparistech.fr
[3]   IRD, UMP AMAP, 2050 Boulevard de la Lironde, 34000 Montpellier, France; nicolas.barbier@ird.fr
[4]   CIRAD, UPR B&SEF, Campus de Baillarguet, 34398 Montpellier Cedex 5, France; valery.gond@cirad.fr
[5]   CIRAD, UMR EcoFoG (AgroParisTech, Cirad, CNRS, Inra, Université des Antilles, Université de la Guyane),
      Campus Agronomique, BP 709, 97310 Kourou, French Guiana; bruno.herault@ecofog.gf
[6]   NOVELTIS, 153 rue du Lac, 31670 Labège, France; mahmoud.elhajj@noveltis.fr
[7]   Airbus Defense and Space, 31 rue des Cosmonautes Z.I. du Palays, 31402 Toulouse, France;
      frederic.fabre@astrium.eads.net
[8]   BRGM, 3 Avenue Claude Guillemin, Orléans 45060, France; j.perrin@brgm.fr
*   Correspondence: Ibrahim.fayad@teledetection.fr; Tel.: +33-04-67-54-87-24

Abstract: LiDAR data has been successfully used to estimate forest parameters such as canopy heights and biomass. Major limitation of LiDAR systems (airborne and spaceborne) arises from their limited spatial coverage. In this study, we present a technique for canopy height mapping using airborne and spaceborne LiDAR data (from the Geoscience Laser Altimeter System (GLAS)). First, canopy heights extracted from both airborne and spaceborne LiDAR were extrapolated from available environmental data. The estimated canopy height maps using Random Forest (RF) regression from airborne or GLAS calibration datasets showed similar precisions (~6 m). To improve the precision of canopy height estimates, regression-kriging was used. Results indicated an improvement in terms of root mean square error (RMSE, from 6.5 to 4.2 m) using the GLAS dataset, and from 5.8 to 1.8 m using the airborne LiDAR dataset. Finally, in order to investigate the impact of the spatial sampling of future LiDAR missions on canopy height estimates precision, six subsets were derived from the initial airborne LiDAR dataset. Results indicated that using the regression-kriging approach a precision of 1.8 m on the canopy height map was achievable with a flight line spacing of 5 km. This precision decreased to 4.8 m for flight line spacing of 50 km.

Keywords: canopy height mapping; airborne LiDAR; ICESat GLAS; forests; French Guiana

1. Introduction

Global warming and climate change attract significant attention in the quantification of the standing above ground biomass (AGB) over the last few decades, not only to understand its effects on the global carbon cycle but also to mitigate the effects of the global warming via the conservation of carbon sinks. Currently, existing AGB estimation methods from remote sensing data are either limited in terms of their sensitivity to high AGB levels (sensor saturation at certain biomass levels using mainly radar and optical data) or in terms of their spatial coverage (limited horizontal coverage using LiDAR data).

Within such constraints, most research studies focus on allometric relations for linking the characteristics of a forest (tree height, diameter at breast height, and wood density) to its biomass (e.g., [1–4]), either at the tree level, or the plot level (plot aggregate allometries). However, one of the important variables in the allometric relations that can be estimated from remote sensing techniques is tree height. Several allometries relied on only canopy height for biomass estimation (e.g., [1,3]). In addition, studies have shown that the use of canopy height increases the precision of biomass estimation (e.g., [2,5]).

Currently the best technique for canopy height estimation using remote sensing is based on the use of LiDAR technology. However, similar to other technologies such as radar, LiDAR alone does not yield satisfactory results (Table S1). Many studies used airborne or spaceborne LiDAR for the estimation of forest canopy heights (e.g., [6–11]). While canopy height estimation using airborne LiDAR data can be very precise (RMSE better than 2 m, [12]), spaceborne LiDAR has a lower precision on the canopy height estimation ranging between 2 m and 10 m depending on the characteristics of the forest (e.g., [7,10,11,13–19]). In addition, Airborne LiDAR is limited in the horizontal domain (limited spatial coverage for airborne data and limited acquisition density for satellite data), whereas spaceborne LiDAR provides global coverage of waveform data, but with a relatively low point density (about 0.51 points/km^2 over French Guiana for example) and inhomogeneous spatial sampling (sampling lines along satellite tracks). Therefore, merging LiDAR data (spaceborne and/or airborne) with other data sources is essential in order to create forest canopy heights with complete land coverage and a good precision (e.g., [10,20]).

Lefsky *et al.* [20] created a global forest canopy height map estimated from GLAS and 500 m Moderate Resolution Imaging Spectro-radiometer (MODIS) data. Their technique showed canopy height estimates with a root mean square error (RMSE) on the estimation of canopy heights of 5.9 m and a coefficient of correlation (R^2) of 0.67. A more recent study conducted by Simard *et al.* [10] improved on the work of Lefsky *et al.* [20] for global canopy height mapping using other ancillary data such as the annual mean precipitation, seasonal precipitation, annual mean temperature, seasonal temperature, data from a digital elevation model (DEM) and percentage tree cover provided from MODIS. Their global canopy height map validated against in-situ measurements showed moderate canopy height estimation precision with an RMSE of 6.1 m (R^2 of 0.5) on the estimation of canopy heights.

Until now, approaches using interpolation methods of LiDAR data (satellite or airborne data) for mapping canopy heights at a regional scale have not been fully studied, as existing canopy height map products need to be more accurate for them to have more practical value to forest ecologists and forest resource managers. Indeed, the approaches presented in Simard *et al.* [10] and Lefsky *et al.* [20] were unable to produce wall-to-wall canopy height maps explaining more than 60% of the total variation of canopy heights (R^2). The last 40% could be related to forest dynamic endogenous processes (*i.e.*, gap processes). In addition, while the canopy height mapping from Simard *et al.* [10] and Lefsky *et al.* [20] provide somewhat good canopy height estimates at large scales with a medium/low spatial resolution, such precisions are not optimal when estimating forest biomass with allometries that only use canopy heights (e.g., [1,3]). In fact, an RMSE of about 6 m leads to a relative error on the estimation of biomass of about 25%. This precision on the estimation of biomass is more than the recommended relative error of 20% set by the United Nations programme on Reducing Emissions from Deforestation and Forest Degradation (REDD) (e.g., [21,22]). Hence, to satisfy the UN REDD recommendations on the precision of biomass, improved canopy height estimates are required.

Most studies used a single interpolation technique for the production of their canopy height maps. Therefore the first question that rises is: how beneficial would it be to mix different interpolation techniques, with both spatial and non-spatial aspects on the accuracy of forest height mapping? Secondly, the issues regarding the cost of acquiring airborne LiDAR data over a regional scale, and the availability of spaceborne LiDAR data globally with lower spatial density pose a new question. What is the impact of the spatial sampling scheme (especially for spaceborne LiDAR) on the accuracy of the produced forest height maps? In the present study, LiDAR canopy height estimates in combination

to ancillary data were used to create a canopy height map at a 250 m resolution covering the entire French Guiana. The LiDAR datasets used were airborne LiDAR, as well as data from the ICESat/GLAS satellite waveforms. For the ancillary datasets, globally available datasets of vegetation indices, precipitation, topography, geological, and forest landscape types were selected. In addition, the effects of the reference canopy height estimate datasets spatial sampling on the canopy height maps created in this study were analyzed. Finally, all created maps were verified using an independent dataset of airborne LiDAR canopy heights estimates (subset of the LD dataset).

A description of the satellite and airborne LiDAR datasets used in this study is given in Section 2. Section 3 presents the methodology used for the creation of the wall-to-wall canopy height maps using regression-kriging. The results are shown in Section 4. Finally, Sections 5 and 6 present the discussion and conclusions, respectively.

2. Study Area and Datasets

2.1. Study Area

French Guiana is located on the northern coast of the South American continent (Figure 1). The study site features almost 96% forested areas of its total area of 83,534 km^2. The forest is mostly old growth rainforest and vegetation types are of natural or anthropogenic origin (forests, swamps, savannas and agricultural crops). French Guiana's terrain is mostly low lying rising occasionally to small hills and low altitude mountains with Altitudes ranging between 0 and 851 m and 67.8% of the slopes are below 5°.
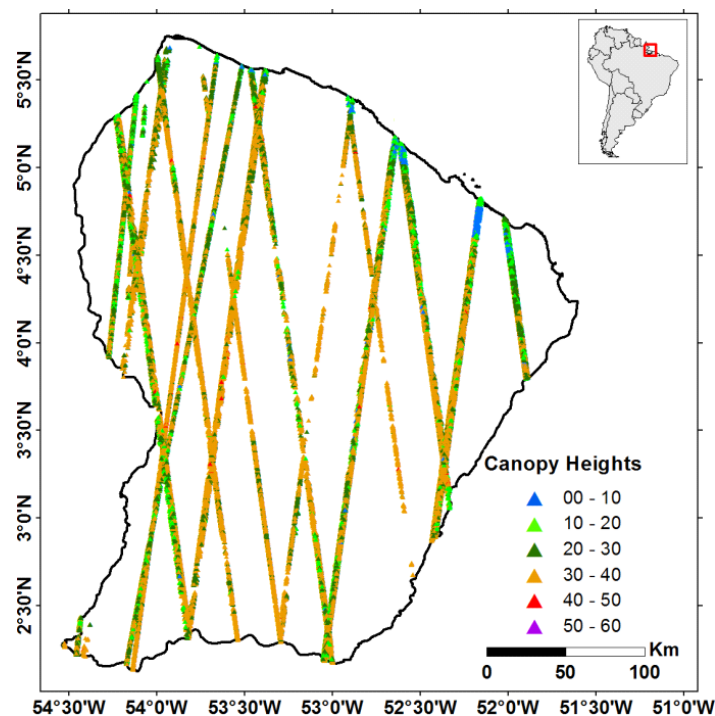
Figure 1. Location of French Guiana and map of canopy heights estimated from the GLAS dataset (in m).

2.2. Datasets Description

2.2.1. Spaceborne LiDAR Dataset

Spaceborne waveform LiDAR data acquired in the near-infrared region at 1064 nm from 2003 till 2009 over French Guiana from the Geoscience Laser Altimeter System (GLAS) on board the Ice, Cloud

and land Elevation Satellite (ICESat) were used in this study. During its operational years (2003–2009), GLAS operated with orbit cycles repeating between every 57 and 197 days for a total of 18 missions. GLAS illuminated surface or footprint has a diameter ranging between 50 and 100 m (average of 70 m), and waveforms were acquired every 175 m along the track. GLAS waveforms were digitized into 544 or 1000 bins with a vertical resolution of 1 ns (15 cm). Hence they were able to measure vertical structures up to 81.6 m and 150 m, respectively. The vertical precision of GLAS was estimated to be between 0 and 3.2 cm on average over flat areas [23].

Over French Guiana, GLAS acquired more than 100,000 waveforms between 2003 and 2009 (Figure 1). However, not all these waveforms are suitable for canopy height estimation, as some of them are contaminated by the atmospheric forward scattering, clouds, *etc.* To remove such footprints, several filters were applied [24]. After data filtering, 47,348 viable footprints remained. Fayad *et al.* [24] recently developed a method for canopy height estimation using GLAS waveforms over French Guiana. This method uses the waveform extent metric calculated from each GLAS waveform and a combination of principal component analysis and Random Forest regressions. Waveform extent (*Wext*) is calculated as the difference between signal end and signal begin. Signal begin and signal end correspond, respectively, to the first and last positions in the waveform crossing a threshold of 4.5 times the standard deviation of the background noise. Their study produced canopy height estimates with a precision of 3.6 m (Figure 1). In addition, the presented technique provides a better alternative to previously developed methods (e.g., [7,11,13,25,26]), as it allows the estimation of canopy heights using only the waveform extent in addition to the waveform samples, instead of relying on metrics derived from the LiDAR waveforms. This is especially useful in dense vegetated areas such French Guiana, where the extraction of metrics is difficult.

2.2.2. Airborne LiDAR Dataset

Small Footprint Low Density LiDAR Dataset

A low point density LiDAR dataset (LD) with small footprint (35 cm) was acquired in 1996 during an airborne geophysical survey that covered 4/5 of French Guiana. Unlike the waveform data acquired by GLAS, this LiDAR dataset provides only the first return pulse. Thus, the data acquired corresponds to the elevation of the first object encountered by the laser beam. Moreover, the database contains laser elevations every 7 m, along track, on flight lines spaced 500 m apart and oriented 30°N, intersected by transverse flight lines spaced 5 km apart and oriented at 120°N. The estimation of canopy heights using the LD dataset showed that canopy heights reached a mean canopy height of 30 m. Only 1% of canopy heights were higher than 50 m in all of French Guiana. In order to estimate canopy heights from the LD dataset, a three-step procedure was implemented [24]. First, LiDAR points acquired over canopy tops were identified (called top-of-canopy points). Next, LiDAR points acquired between two consecutive top-of-canopies were identified (called pseudo-ground points), and then the lowest pseudo-ground point in a distance of 1 km was identified as the ground point. Finally, for each ground point and corresponding top-of-canopy point the canopy height was estimated by calculating the distance from the ground point to the segment joining the two top-of-canopy points. Over French Guiana, canopy height estimates from the LD dataset had a density of 1.19 points/km^2. Low canopy heights (maximum of 20 m) can be observed in the northern parts of French Guiana on the coastal marsh areas.

In this study, the canopy heights estimated from the transverse lines will be used for verification (about 7% of the LD dataset, named LD_val), while the rest of the estimates will be primarily used for model building and calibration (93% of the LD dataset, named LD_cal). Figure 2 shows the canopy height estimates of the entire small footprint low density LiDAR dataset.
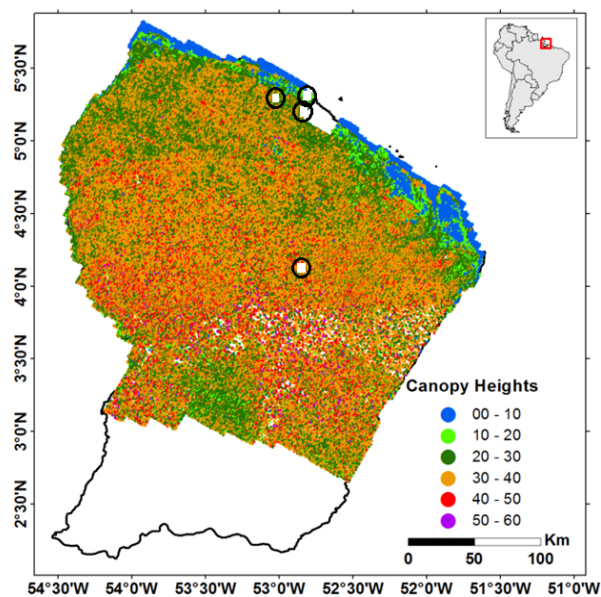
Figure 2. Map of canopy heights calculated from the airborne LiDAR LD dataset for French Guiana. The locations of airborne LiDAR HD datasets are in delineated with circles.

Small Footprint High Density LiDAR Dataset

High point density LiDAR dataset (HD) was acquired between 2004 and 2009 by the Altoa Company with a mean acquisition density of 3.5 points/m^2 (between 0.9 and 5.6 points/m^2). This dataset covers several small reference sites in the North of French Guiana (Figure 2). A comparison of the canopy height estimates of the LD and HD datasets showed a high correlation between LD and HD datasets (R^2 of 93%) with a root mean square error of 1.57 m [24].

2.2.3. Ancillary Datasets

Twelve available mapped environmental predictors were used as in Simard *et al.* [10] in an attempt to explain canopy heights. These variables include geological map, forest landscape type map, SRTM digital elevation model (DEM) derived maps, six vegetation indices maps issued from MODIS optical images, and finally one map issued from rainfall data. The predictors are summarized in Table 1.

Table 1. Description of the mapped environmental predictors used for canopy height mapping.

Short Name	Full Name	Source	Resolution
MIN_EVI	Minimum value of EVI time series data		
MEAN_EVI	Mean value of EVI time series data		
MAX_EVI	Maximum value of EVI time series data	MODIS	250 m
PC1	1st principal component of EVI time series data		
PC2	2nd principal component of EVI time series data		
PC3	3rd principal component of EVI time series data		
Slope	Terrain slope in 3 × 3 cells		
Roughness	Terrain roughness in 3 × 3 cells	SRTM	90 m
ln_drain	Log of drainage surface		
GEOL	Geological map (no units, arbitrary shapes)	[27]	Vector
LTs	Forest landscape type (no units, arbitrary shapes)	[28]	1 km (Vector)
Rain	mean value of rainfall	TRMM	8 km

MODerate-Resolution IMAGING Spectroradiometer (MODIS) Data

MODIS sensor mounted on the Terra and Aqua satellites possesses a total of 36 spectral bands, of which seven are designed specifically for land applications with spatial resolutions ranging between 250 m and 1 km. The MODIS dataset used in this study includes ten years (1 January 2003 to

31 December 2012) of the enhanced vegetation index (EVI) time series data (MOD13A1c5). The normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI) are often used in many applications that studies biophysical parameters. However, the relationship between NDVI and EVI varies depending on the characteristics of the study area. EVI data effectively characterize biophysical and biochemical states and processes from vegetated surfaces. Freitas *et al.* [29] and Pascual *et al.* [30] found a strong relationship between canopy heights and vegetation indices. Therefore, it might be interesting to include the EVI in canopy height estimation models. A 10 years period was used to synchronize with the GLAS data (from 2003 to 2009). Using the EVI time series data, six maps were issued: minimum, mean and maximum values of the EVI time series data (MIN_EVI, MEAN_EVI, and MAX_EVI respectively, Figure S1), and the three first principal components issued from the principal component analysis of the EVI time series data (PC1, PC2 and PC3).

SRTM Digital Elevation Model Data

The digital elevation model (DEM) acquired from the shuttle radar topography mission (SRTM) were also used in this study. The three-arcsecond resolution DEM was used, as its resolution is closer to the resolution of most of the used ancillary data (250 m resolution). Bourgine and Baghdadi [31] found that the accuracy of the SRTM DSM elevation data was evaluated at approximately 10 m (standard deviation of error) in French Guiana. The SRTM DEM dataset and its derived maps were considered since local topography and drainage are important for tree anchorage and forest dynamics both directly and through soil types.

From the SRTM DEM data, three maps were derived:(1) A slope map (Slope, Figure S2a) that is calculated using the maximum change in elevation over the distance between each cell of the DEM and its eight neighbors in a 3×3 window (Slope). The slope has been widely used in many studies for the correction of canopy height estimates using either airborne or spaceborne LiDAR (e.g., [7,13,26]). Incorporating the slope in canopy height estimation models increases the accuracy of the canopy height estimates, as it removes any height added by the slope that were misinterpreted as being a part of the canopy height [13]; (2) A local surface roughness map (Roughness, Figure S2b), where the roughness corresponds to the standard deviation of elevation in a 3×3 moving window. Areas with low standard deviation represent areas with more homogeneous canopy heights, while higher standard deviation presents more heterogeneous canopy heights. The roughness index was introduced in the study of Carabajal and Harding [23] and Fayad *et al.* [24] to be correlated with the waveform extent of GLAS waveforms (the waveform extent represents a rough estimate of canopy heights); (3) Finally, a logarithmic drainage area map (ln_drain, Figure S2c), where the drained area measures the surface of the hydraulic basin that flows through a cell using the D8 algorithm. A low value indicates cells located on a ridge upstream, whereas the highest values indicate cells located downstream in valleys.

Geological Map

Geology is an important determinant of soil formation and the conditioning of chemical and physical properties of the soil, which affects tree growth and other forest parameters. A geological substratum map (GEOL) produced by the French Geological Survey [27] was therefore used in this study (Figure S3a). The map was simplified in order to retain only the large five biggest rock formations: recent sediments, volcanic sedimentary rock, granites, gabbros, and gneiss. This simplification was required in order for each geological class to be sampled with a satisfying rate.

Forest Landscape Types Map

A forest landscape types map developed by Gond *et al.* [28] at 1 km resolution was also used (Figure S3b). In this map, 33 remotely sensed landscape types (LTs) using SPOT-VEGETATION images were interpreted. Five classes of the total 33 classes were used in this study, as they occupy more than 78% of the forest in that area (the other 22% represent water bodies and non-forests). The LTs can be summarized as follows:

(1) LT8 represents dense closed-canopy forest with small crowns of the same canopy height and small gaps mixed with regular canopies with well-developed crowns of almost the same canopy height without large gaps interlaced with flooded savannas (10%).

(2) LT9 is a closed canopy forest dominated by well-developed crowns of almost the same canopy height without large gaps.

(3) LT10 is an irregular and disrupted-canopy forest where the trees have very different heights and different crown diameters with large gaps mixed with closed-canopy forest dominated by well-developed crowns at almost the same elevation without large gaps. LT10 is also interlaced with liana forests.

(4) LT11 is similar to LT10 with more liana forest and non-forest land covers.

(5) LT12 is an open forest associated with wetlands and bamboo thickets. The LT dataset was chosen for its correlation with canopy heights. Indeed, in Fayad *et al.* [24], the difference between SRTM and canopy top elevations from ICESat were found to be correlated with different LTs as well as different canopy heights.

Average Rainfall Map

Precipitation data from the NASA tropical rainfall measuring mission (TRMM) which launched in 1997 for the measurement and monitoring of tropical rainfall were used. TRMM data cover the average daily precipitation over the last 10 years (2003–2013) with a resolution of 8 km (Rain, Figure S3c). Recent studies suggest a tight relationship between maximum canopy height in old stands and the ration of annual rainfall [32]. As the ratio of rainfall increases, so did maximum tree height.

3. Canopy Height Estimation Methods

In order to estimate canopy heights on a 250×250 m regular grid and un-sampled by LiDAR data (GLAS or LD datasets) a five step procedure was conducted based on the statistical and spatial relationship between the LiDAR canopy height estimates and the ancillary variable datasets (GEOL, LT, Rain, Slope . . .) using widely used empirical estimation methods: Random Forest regressions, ordinary kriging, and regression-kriging (e.g., [10,33–35]). The wall-to-wall maps will have a 250 m resolution corresponding to the resolution of the majority of the used predictors.

3.1. Canopy Height Trend Mapping Using Random Forest Regressions

The Random Forest regression technique (RF) that was developed by Breiman [36] was employed to estimate canopy heights over a regular grid with a 250 m resolution in French Guiana. The Random Forest (RF) technique is known to be a performant regression method that is becoming widely used by the remote sensing community for, among other, canopy height estimation (e.g., [10,37]), and biomass estimation [33,34,38]. The main advantage of Random Forest is its incorporation of continuous or qualitative predictors without making assumptions about their statistical distribution or covariance structure [36].

To select the predictors that best explained canopy heights, the variable selection using Random Forest algorithm (VSURF) of Guener *et al.* [39] was used. The first step of VSURF is to sort variables based on their initial scores of importance, and remove the variables with very small importance. The variable score of importance is evaluated based on the increase of error in the prediction when removing a certain variable. The second step of VSURF consists of building a nested collection of RF models involving the k first variables (*k* = 1 to m where m is the number of remaining variables after the first step with the highest score of importance), and selecting the variables in the model giving the smallest out-of-bag error. The variable selection process has been done in R using the Variable Selection Using Random Forest (VSURF) package, proposed by Genuer *et al.* [39]. Variable selection is made automatically after providing the variables as well as the dataset. After selecting the best variables, an RF model using these variables was calibrated, and applied on the entire study area in order to create a

wall-to-wall canopy heights trend map. Random Forest is called an ensemble classifier because it uses a tree-based classifier multiple times and aggregates the results. Each tree is grown using a randomized subset of predictors. This procedure is expected to decrease the correlation among the trees, which improves model accuracy [36]. The final prediction decision is based on a voting system of all the predictions from the decision trees that have been created. The RF regressions were implemented in R using the randomforest function from the randomforest package. The randomforest function requires the predicted variable, as well as the predictors (in this case the best selected variables), and the number of trees (default of 500 tress, which was used in this study).

3.2. Canopy Height Mapping Using Regression-Kriging

After using the RF technique for canopy height mapping, the regression-kriging technique (RK) was employed. RK is a spatial prediction technique which combines the regression value of explanatory variables (ancillary datasets) and the kriging of the regression residuals (reference canopy heights—RF estimated canopy heights) [35]. This technique was developed primarily to account for the correlation between environmental variables and the unsatisfactory goodness of fit of the spatial variance model of the dataset [40], thus preventing the stationarity of the studied auto correlated variables, like the RF residuals that are centered at 0. Regression kriging can be summarized as the technique that fits separately the trend and the residuals and sum them [41]. RK can be expressed as follows:

$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0) \tag{1}$$

where $\hat{z}(s_0)$ is the predicted value at an unvisited location s_0, $\hat{m}(s_0)$ the fitted trend (the RF canopy height estimates), and $\hat{e}(s_0)$ the kriged residual.

3.3. Ordinary Krigging of Regression Residuals

The wall-to-wall map created in the previous section using RF does not take into account the spatial correlation between the canopy heights, as it is a non-spatial method and assumes spatial independence of the predictors and the predicted variables. However, some of the unexplained variance in the RF predictions could be due to the spatial correlation between canopy heights. Thus, a spatial prediction model is required when data are spatially dependent. In this study, we used the ordinary kriging (OK) model, which allows the interpolation of un-sampled data based solely on a linear model of regionalization known as the semivariogram (the semivariogram is a weighing function and is required for the kriging). The semivariogram plots semivariance γ as a function of distance between samples h using the following function:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [e(s_i) - e(s_i + h)]^2 \tag{2}$$

where $\gamma(h)$ is the semivariance as a function of lag distance h, $N(h)$ is the number of pairs of data separated by h, and e, the canopy height estimate residuals at locations s_i and (s_i+h) [42]. Semivariograms have three main parameters: (1) the *nugget* which is the semivariance at a lag distance of zero; (2) the *sill* is the semivariance where there is no spatial correlation; and (3) the *range* is the distance at which the *sill* is reached. After plotting the sample semivariogram that describes the spatial autocorrelation of a given dataset, a mathematical function is fitted to this semivariogram in order to represent the range, the sill and the nugget. Thus, the datasets sample variogram can now be represented using a function. The sample semivariogram was plotted in R using the geoR package. The main parameters of the semivariograms were first roughly estimated, and the estimates corrected by an automated function provided by the geoR package. After model fitting of the sample

semivariogram, ordinary krigging is then used, which estimates values $\hat{e}(s_0)$ at an unvisited location s_0 using the following equation:

$$\hat{e}(s_0) = \sum_{i=1}^{n} \lambda_i e(s_i) \tag{3}$$

where $\hat{e}(s_0)$ is the kriged residual, λ_i are the kriging weights determined by the spatial autocorrelation structure (variogram), and $e(s_i)$ is the residual at location s_i [42]. Ordinary kriging was implemented using ArcMap with the semivariogram parameters from the previous step, as well as the canopy height residuals.

3.4. Effects of LiDAR Sampling Density on Precision of the Mapped Canopy Heights

The purpose of the current analysis was to measure how the accuracy of the canopy height maps is affected by the spatial sampling of LiDAR data. Several subsets of LiDAR data densities were built from the LD_cal dataset in order to study the impact of the spatial sampling of future spaceborne LiDAR systems on the precision of the created canopy height map.

In total, six configurations corresponding to flight-line spacing of 5, 10, 20, 30, 40 and 50 km were considered (with respect to the flight plan of the LD_cal dataset). For each configuration, a subset was extracted from the LD_cal dataset where the flight-line spacing of the subset met the criterion of the configuration. For example, for the configuration with a flight line spacing of 5 km, the first selected flight line is the first available flight line from the LD_cal dataset starting from the West. Next, all flight lines with a distance of a multiple of about 5 km from the first selected flight line were selected. Finally, using each of the LD_cal subsets obtained and named, respectively, LD_5 (0.29 pts/km^2), LD_10 (0.11 pts/km^2), LD_20 (0.08 pts/km^2), LD_30 (0.05 pts/km^2), LD_40 (0.04 pts/km^2), and LD_50 (0.03 pts/km^2), a corresponding canopy height map was created. Canopy height maps were created using the same procedure described in Sections 3.1 and 3.2 which consists of first creating a canopy height map using Random Forest regressions with each one of the LD_cal subsets as reference data and the ancillary variables as predictor variables for the model, and next each canopy height residual from each model were kriged and added to the corresponding canopy height map.

Precision of the resulting kriged canopy height maps using the low-density LiDAR datasets was estimated for each LD_cal subset by comparing these canopy height maps to the verification datasets (HD and LD_val).

4. Results

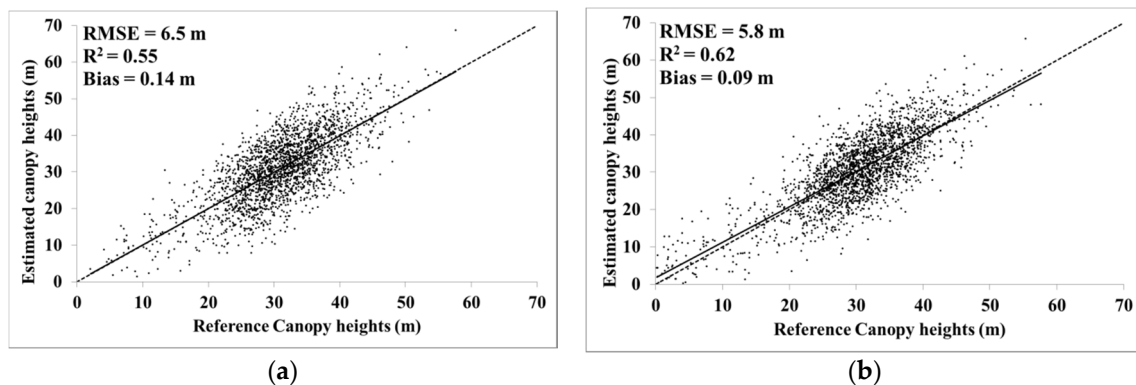### 4.1. Canopy Height Trend Mapping Using Random Forest Regressions

A Random Forest regression model was built for each one of the two calibration datasets (GLAS, LD_cal) with all the twelve predictors (Table 1). The first calibration dataset used in the Random Forest regression contains the canopy height estimates obtained from GLAS waveforms using the PCA and RF based canopy height estimation model [24].

After creating the Random Forest regression models using the twelve predictors, only the predictors that best explained the canopy heights were selected using the variable selector test. Results showed that the best predictors were the same for both calibration datasets. The best predictors according to their importance are respectively: the roughness, the mean value of the EVI time series data, the geology, the mean value of the rainfall, and the slope. Next, each calibration dataset (GLAS, LD_cal) and the best predictors were used to fit a RF model. Then each fitted RF model was used to create wall-to-wall canopy height maps of the entire French Guiana forest (Figure 3). Finally, each wall-to-wall canopy height map was validated against both, the LD_val and the HD datasets. Results showed that the precision of the maps was almost the same when using different calibration datasets (Figure 4 and Table 2). Indeed, when using the RF model with the GLAS dataset, the RMSE for the canopy height estimates was 6.5 m (R^2 of 0.55). The precision of the estimates slightly increased when using the RF model with the LD_cal dataset, with an RMSE on the canopy height estimates of 5.8 m

(R^2 of 0.62). Finally, the bias (mean (verification canopy heights—estimated canopy heights)) for both the GLAS and LD_cal datasets was very low (<0.2 m).
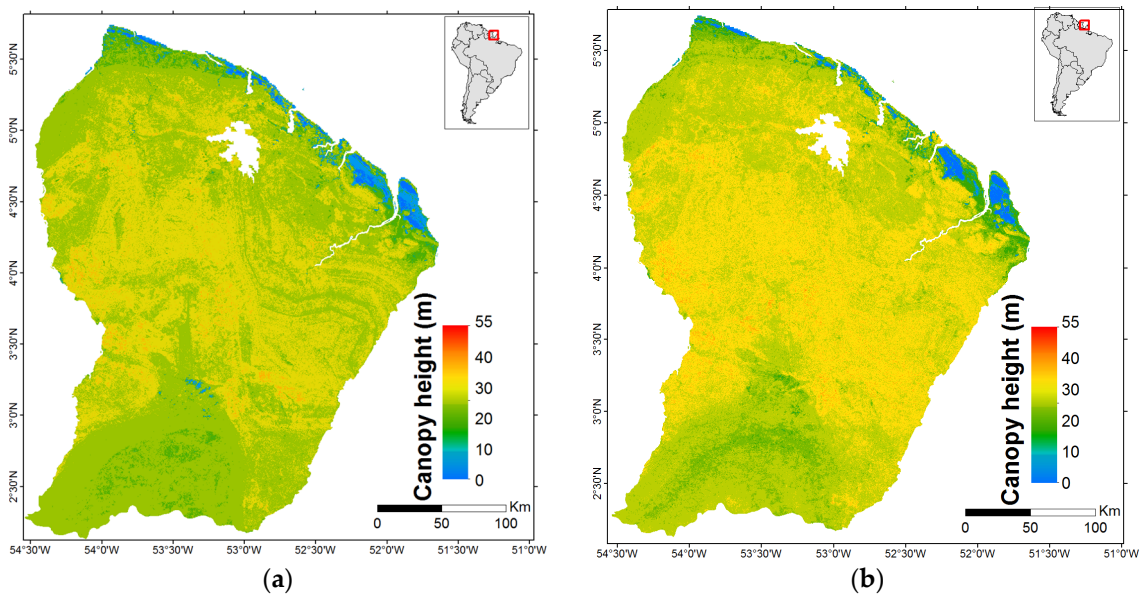
Figure 3. Wall-to-wall map of French Guiana with Random Forest regressions using as reference data the canopy height estimates from: (**a**) GLAS dataset; and (**b**) LD_cal dataset.

Figure 4. Comparison between the reference canopy heights of the verification datasets and the canopy height trend estimates using Random Forest: (**a**) GLAS dataset; and (**b**) LD_cal dataset.

Table 2. Comparison between the canopy heights of the verification datasets (LD_val and HD) and the canopy height estimates using regression kriging.

	Using RF Only			Using Regression Kriging		
Dataset	Bias (m)	RMSE (m)	R^2	Bias (m)	RMSE (m)	R^2
GLAS	0.14	6.5	0.55	0.09	4.2	0.75
LD_cal	0.15	5.8	0.62	0.12	1.8	0.94
LD_5	0.06	5.7	0.65	0.12	1.8	0.94
LD_20	0.09	6.0	0.63	0.14	3.3	0.75
LD_30	0.14	6.2	0.60	0.05	3.9	0.75
LD_40	0.11	6.1	0.62	0.09	3.9	0.74
LD_50	0.07	6.2	0.60	0.13	4.8	0.66

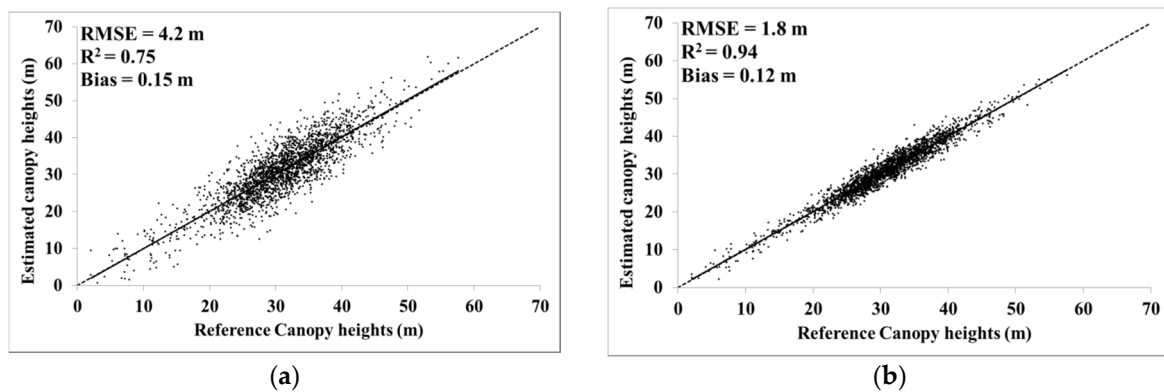### 4.2. Canopy Height Estimation Using Regression-Kriging

After creating the wall-to-wall maps using the Random Forest regression models, canopy height residuals (reference canopy height—estimated canopy height) were kriged from each model in an attempt to increase the precision of the canopy height estimates using the Random Forest regressions for both the GLAS and LD_cal datasets. For each canopy height residual map issued from the GLAS dataset or the LD_cal dataset, a semivariogram was generated. Results showed that semivariograms issued from each of the two height residual map presented similar patterns that could be best-fitted using an exponential model:

$$\gamma(h) = S^2 + \sigma^2 \left[1 - exp\left(\frac{-h}{a} \right) \right] \tag{4}$$

where S^2 is the nugget, σ^2 the sill, and a the range of the semivariogram (γ). For the different canopy height residual datasets, the fitted semivariograms presented similar nuggets (between 15 and 18 m^2), sills (between 28 and 32 m^2), and ranges (between 4421 and 4823 m). Next, the fitted semivariograms were used in the kriging of the canopy height residuals for each of the GLAS and LD_cal datasets. In total, two residual maps were obtained. Then, each residual-kriged map was added to the wall-to-wall map corresponding to that model (Figure 5). These maps were then validated using the verification datasets (LD_val and HD) (Figure 6, Table 2). Results showed that using the regression-kriging technique increased the estimation precision of these maps. Indeed, for the canopy heights map obtained using the GLAS dataset, the RMSE on the canopy height estimation decreased from 6.5 m with Random Forest regression to 3.6 m (R^2 of 0.76) with regression-kriging. For the canopy heights map obtained using the LD_cal dataset, the RMSE on the canopy height estimation decreased from 5.8 to 1.8 m (R^2 of 0.95) with regression-kriging. Moreover, the bias for the two datasets was very low (<0.2 m). These results show that the maps derived from the LD_cal datasets and using regression-kriging clearly captured finer local variations when estimating canopy heights. Finally, the canopy height estimates uncertainty from both maps appears to be correlated with the location of the reference dataset measurements (Figure 7). For the GLAS dataset, the standard deviation of canopy height estimates uncertainty ranged between 4 and 7 m (Figure 7a). In addition, standard deviation values appear to be lower near the location of the GLAS canopy height estimates, and increases with increasing distance until they reach 7 m. Similar results appear for the LD_cal dataset (Figure 7b), with lower standard deviations in areas with denser LiDAR acquisitions (*i.e.*, north of French Guiana) and higher standard deviations with sparser LiDAR acquisitions (*i.e.*, center of French Guiana). However, due to the generally denser dataset in comparison to the GLAS dataset, standard deviation of canopy height estimates uncertainty ranged between 1 and 4 m (Figure 7b).
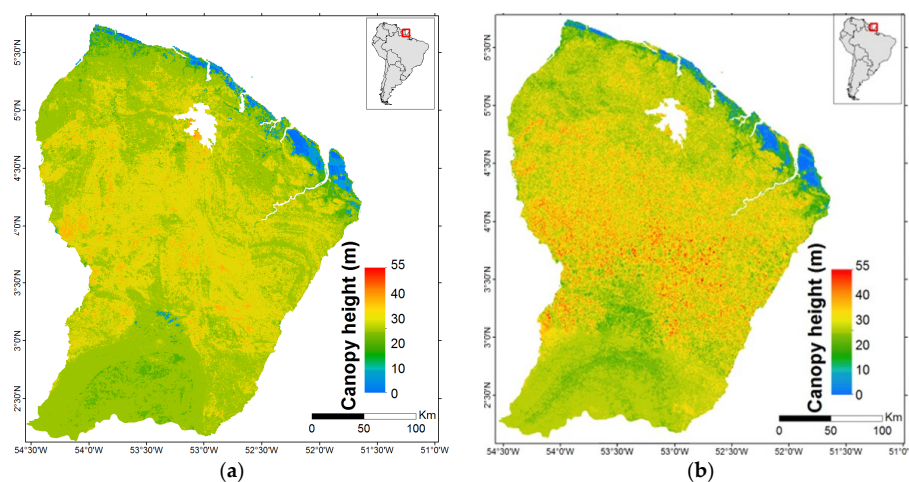
Figure 5. Wall-to-wall map of French Guiana with regression-kriging using as reference data canopy height estimates from: (**a**) GLAS dataset; and (**b**) LD dataset.

Figure 6. Comparison between the reference canopy heights of the verification datasets and the canopy height estimates using Random Forest regressions and residual-kriging: (**a**) GLAS dataset; and (**b**) LD_cal dataset.
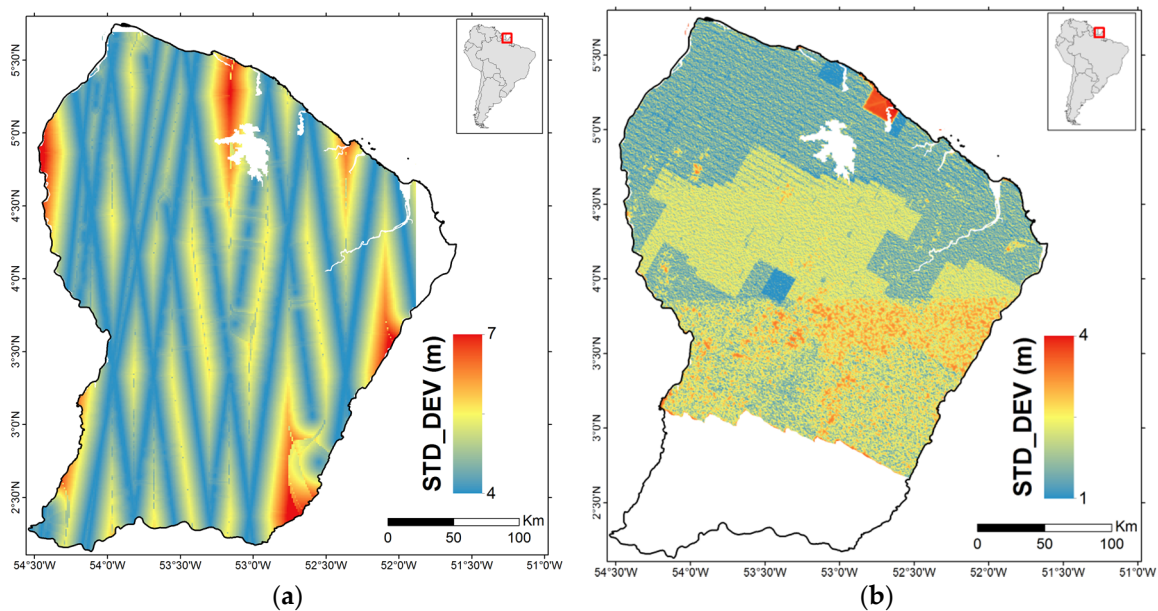
Figure 7. Wall-to-wall standard deviation map (STD_DEV) of the canopy height estimates uncertainty for: (**a**) GLAS dataset; and (**b**) LD_cal dataset.

4.3. Relationship between LiDAR Flight Lines Spacing and the Precision on the Kriged Canopy Height

The analysis performed in the previous section showed a significant improvement of canopy height estimation precision when adding the kriged residuals. This improvement was observed for the two calibration datasets. In the case of the LD_cal dataset, the improvement was the highest from 5.8 m without the height residual kriging to 1.8 m with the residual kriging. This is due to the high density of canopy height estimates in this dataset. Indeed, for the LD_cal dataset, canopy height estimates are distributed over flight lines with an average distance of about 500 m, while canopy height estimates for the GLAS dataset are distributed over flight lines with an average distance of 20 km.

Therefore, in this section the precision of the kriged canopy height maps obtained using different LiDAR densities was assessed. The purpose was to analyze the impact of the LiDAR flight line spacing from the LD_cal dataset on the precision of the kriged canopy height map. In order to create canopy height maps using the LD_cal subsets, first the best predictor variables to be used in the Random Forest regressions were selected using the procedure mentioned in Section 3.1. Results indicated that for the LD_cal subsets, the predictor variables that best explained canopy heights were the same as those for

the GLAS and LD_cal datasets. The predictor variables were namely the roughness, the mean value of the EVI time series data, the geology, the mean annual rainfall, and the terrain slope. Results showed that the precision of the produced canopy height maps using RF regressions with the LD_cal subsets without kriging of the residuals was in the same order as the canopy height maps obtained with the two calibration datasets (GLAS and LD_cal) (Table 2). For these subsets, the RMSE on the canopy height estimates ranged between 5.7 and 6.2 m (R^2 between 0.60 and 0.65). In order to add the kriged height residuals to the canopy height maps, the semivariograms of the canopy height residuals for each LD_cal subset were fitted. Similar sill, range and nuggets were obtained as those from the canopy height residuals from the GLAS and LD_cal datasets. When adding the kriged residuals corresponding to each of the LD_cal subsets (Figure 8), the precision on the canopy height estimate maps increased as expected (Figure S4, Table 2). This increase in the precision on canopy height estimation was found to be negatively correlated with LiDAR flight lines spacing of the LD subsets. For the LD_5 and LD_10 subsets, the precision on the canopy height estimates was similar to the results obtained with the LD_cal dataset (RMSE = 1.8 m, R^2 = 0.94). However, for the LD_20, LD_30, LD_40, and LD_50 subsets, the precision on the canopy height estimates decreased from RMSE = 3.3 m for LD_20 to RMSE = 4.8 m for LD_50.
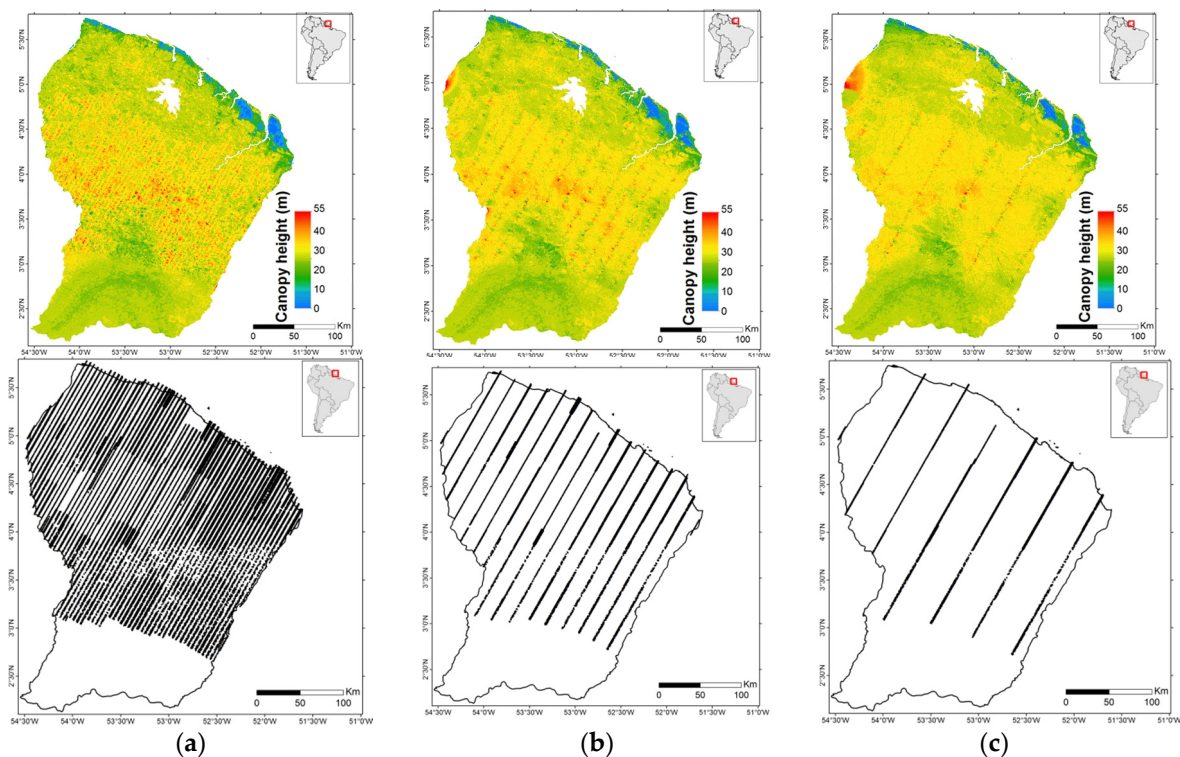
Figure 8. Examples of wall-to-wall maps of French Guiana with regression-kriging using as reference data the canopy height estimates from: (**a**) LD_5; (**b**) LD_20; and (**c**) LD_50.

5. Discussion

For the first time, airborne and spaceborne LiDAR data were combined to predict and map (potential) forest heights at the scale of a tropical country. Our approach is based on the merging of LiDAR canopy height estimates (airborne and spaceborne) with ancillary environmental and geographical data and using regression-kriging. Because of this approach, both calibration and verification are way more robust than in previous studies. Error and bias were also lower than previous studies. Indeed, using Random Forest regressions instead of linear models produced canopy height estimates with no bias regardless of the LiDAR dataset density used, nor the type of LiDAR data (airborne or spaceborne). High bias estimates are due to the nature of the linear regression models,

which are high-bias/low-variance models, and therefore the problem facing these types of models is reducing the bias especially with larger datasets. This problem is non-existent in RF model due to its opposite nature (low-bias/high-variance).

Moreover, our canopy height estimates using Random Forest show precisions slightly higher in comparison to a recent study [10]. The study of Simard *et al.* [10], which estimated canopy heights globally and obtained a precision on the canopy height estimates of 6.1 m (RMSE) using a RF regression with GLAS data and some common predictor variables used in this study. Comparing their global canopy height map with our verification dataset (LD_val and HD) showed a slightly higher RMSE of 7.3 m (Figure 9). Our slightly better results can be attributed to: (1) using variables better correlated with canopy heights; and (2) our canopy height estimates used for model calibration, especially those obtained from airborne LiDAR are more precise in comparison to their spaceborne counterparts. In addition, the canopy height maps estimated in this study were also compared to the global canopy height map produced in the study of Lefsky *et al.* [20]. The comparison showed very poor correlations with an RMSE of 12.4 m (R^2 insignificant). This is mainly due to the canopy heights obtained in the study of Lefsky *et al.* [20] representing Lorey's height while the canopy heights in our study represent maximum canopy height. Lorey's heights are generally expected to be lower than maximum canopy heights [10].
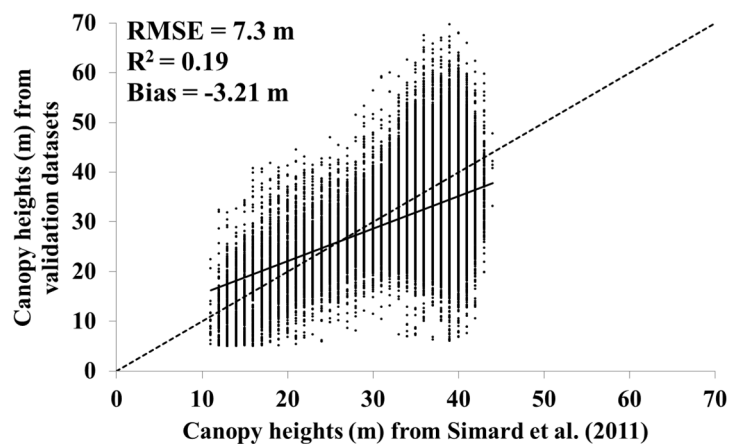
Figure 9. Comparison between the canopy heights of our verification datasets (LD_val and HD) and the canopy height estimates from the study of [10].

However, since neither our approach nor the approaches used in Simard *et al.* [10] and Lefsly *et al.* [20] were able to produce wall-to-wall canopy height maps explaining more than 60% of the total variation of canopy heights (R^2) (Table 2), and in order to improve the precision of the canopy height product, canopy height estimation residuals (reference canopy heights—estimated canopy heights by RF) were kriged and used. This approach proved very efficient, although highly sensitive to the spatial sampling of the reference LiDAR dataset (flight line spacing). Indeed, for the French Guiana, the semivariograms indicated that the autocorrelation in the canopy height residuals did not go beyond 5 km, beyond this distance their contribution to the precision of the final canopy height maps started to decrease. In contrast, kriging only the LiDAR canopy heights without using the predictor variables with RF did not yield satisfactory results. For instance, by kriging directly the LD_cal canopy heights, we obtained a RMSE on the canopy height estimates of 5.1 m in comparison to the verification datasets against an RMSE of 5.8 for the RF technique with the LD_cal and 1.8 m for the regression-kriging technique. For the kriged GLAS canopy heights, the precision on the estimated canopy heights was 7.3 m in comparison to the verification datasets. The low precision of the kriged canopy heights from the GLAS dataset is due to the fact that the distance between the available canopy height estimates (~20 km) is higher than the range of their spatial autocorrelation (5 km), so a high smoothing occurred. This also explains the difference between the kriged canopy height estimates

and the estimates from the verification datasets (bias of -4 m). To analyze the contribution of the regression-kriging technique on the canopy height precision, the kriging of the height residuals were replaced by the mean value of the height residuals in a 5 km radius. Results showed that for the LD_cal using the mean of the residuals, the R^2 decreases from 0.94 to 0.85 and the RMSE increases to 2.4 m in comparison to the kriging method (RMSE = 1.8 m).

Finally, the 250 m resolution of the canopy height map was chosen for different reasons. First, it was set initially because half of the used predictors had a resolution of 250 m. Three out of twelve predictors had a resolution of 90 m, and the rest had a resolution of more than 250 m. After the variable importance test was carried, the predictor with the highest importance was identified as the MEAN_EVI (250 m resolution). The importance of the predictors with a 90 m resolution was far behind. Therefore, a 250 m product was deemed best, as using a lower resolution product will produce smoother canopy height maps with less local canopy height variations, while maps with higher resolution (90 m) will not necessarily capture finer local canopy height variations than the 250 m product.

Given the low error level obtained from our canopy height wall-to-wall map, our approach can be used to create valuable forest height maps that can be fed in biomass estimation efforts, either at the tree, plot or landscape level (e.g., for calibrating local H-DBH relationships or for inverting AGB directly from H). In addition, the unique combination of data available here allowed testing different models and sampling strategies (sensitivity study). This will help future mapping efforts over other regions were airborne datasets might not be available, as well as for dimensioning future LiDAR spaceborne missions.

6. Conclusions

An approach for canopy heights mapping over the tropical forest of French Guiana was presented. It is based on the merging of LiDAR canopy height estimates (airborne and spaceborne) with ancillary data. To create the canopy height maps, the predictor variables (ancillary variables) that best explained the canopy heights were firstly chosen. The best predictor variables happened to be the same for all the canopy height datasets: the roughness, the mean value of the EVI time series data, the geology, the mean value of the annual rainfall, and the terrain slope. Random Forest (RF) regressions, which were used to fit the best predictors to the LiDAR canopy height datasets showed moderate canopy height estimation precision when using either airborne or spaceborne LiDAR (RMSE better than 6 m).

To improve the precision of the obtained canopy height maps, regression-kriging (RK) was used. The height residuals (reference canopy heights—estimated canopy heights by RF) obtained from each reference LiDAR dataset were kriged and added to the canopy height estimates obtained from RF regressions. An improvement in the precision of the canopy height maps was observed, however, this improvement is positively correlated to the point density of the calibration datasets used. For the GLAS dataset, which has a flight line spacing of 30 km, the RMSE on the canopy height estimates was improved to 4.2 m; however, for LD_cal dataset, which has a 0.5 km flight line spacing, the RMSE on canopy height estimates was improved to 1.8 m. Further investigation shows that for LiDAR datasets with flight lines spacing below the range of the spatial autocorrelation of the height residuals (5 km lines spacing), the precision of the canopy height estimates was at its highest (RMSE of 1.8 m), and decreases with the increase of the flight line spacing until it reached 4.8 m for a flight line spacing of 50 km. In comparison, studies such as [10], which used only the Random Forest regressions, were only able to estimate canopy height at a precision of 6.1 m with a lower resolution maps (1 km).

The procedure presented in this paper, which uses the regression-kriging, has strong prospects for application to other tropical forests.

roughness map "in m" (b); and drainage surface map (c). Figure S3: Geological map (a); Forest landscape types map (b); and Average rainfall map (c). Figure S4: Comparison between reference canopy heights of the validation datasets (LD_val and HD) and canopy height estimates using Random Forest regressions and residual-kriging for (a) LD_5; (b) LD_20; (c) LD_50.

Author Contributions: Fayad I., Baghdadi N. and Bailly J.S. had the original idea for the study, with all co-authors carrying out the design. Fayad I. performed the experiments, and the results were interpreted by all co-authors. Fayad I. and Baghdadi N. drafted the manuscript, which was then improved and revised by all authors. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1.   Drake, J.B.; Robert, G.K.; Ralph, O.D.; David, B.C.; Richard, C.; Bryan, B.J.; Michelle, H. Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: Factors affecting the generality of relationships. *Glob. Ecol. Biogeogr.* **2003**, *12*, 147–159. [CrossRef]

2.   Chave, J.; Andalo, C.; Brown, S.; Cairns, M.A.; Chambers, J.Q.; Eamus, D.; Fölster, H. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* **2005**, *145*, 87–99. [CrossRef] [PubMed]

3.   Asner, G.P.; Mascaro, J. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. *Remote Sens. Environ.* **2014**, *140*, 614–624. [CrossRef]

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

5.   Feldpaush, T.R.; Lloyd, J.; Lewis, S.L.; Brienen, R.J.W.; Gloor, M.; Mendoza, A.M.; Lopez-Gonzalez, G.; Banin, L.; Salim, K.A.; Affum-Baffoe, K.; *et al*. Tree height integrated into pantropical forest biomass estimates. *Biogeosciences* **2012**, *9*, 3381–3403. [CrossRef]

6.   Mitchard, E.T.A.; Saatchi, S.S.; White, L.J.T.; Abernethy, K.A.; Jeffery, K.J.; Lewis, S.L.; Collins, M. Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: Overcoming problems of high biomass and persistent cloud. *Biogeosciences* **2012**, *9*, 179–191. [CrossRef]

7.   Lefsky, M.A.; David, H.J.; Michael, K.; Warren, B.C.; Claudia, C.C.; Fernando Del Bom, H.E.-S.; Maria, O.; Raimundo, D.O. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophys. Res. Lett.* **2005**, *32*, L22S02. [CrossRef]

8.   Boudreau, J.; Ross, F.N.; Hank, A.M.; André, G.; Luc, B.; Daniel, S.K. Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sens. Environ.* **2008**, *112*, 3876–3890. [CrossRef]

9.   Saatchi, S.S.; Nancy, L.H.; Sandra, B.; Michael, L.; Edward, T.A.M.; William, S.; Brian, R.Z. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9899–9904. [CrossRef] [PubMed]

10. Simard, P.M.; Naiara, F.J.B.; Alessandro, B. Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res.* **2011**, *116*, G04021. [CrossRef]

11. Baghdadi, N.; le Maire, G.; Fayad, I.; Bailly, J.S.; Nouvellon, Y.; Lemos, C.; Hakamada, R. Testing different methods of forest height and aboveground biomass estimations from ICESat/GLAS data in Eucalyptus plantations in Brazil. *IEEE (JSTARS)* **2014**, *7*, 290–299. [CrossRef]

12. Næsset, E. Predicting forest stand charactersitics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* **2002**, *80*, 88–99. [CrossRef]

13. Lee, S.; Ni-Meister, W.; Yang, W.; Chen, Q. Physically based vertical vegetation structure retrieval from ICESat data: Validation using LVIS in White Mountain National Forest, New Hampshire, USA. *Remote Sens. Environ.* **2011**, *115*, 2776–2785. [CrossRef]

14. Hilbert, C.; Schmullius, C. Influence of surface topography on ICESat/GLAS forest height estimation and waveform shape. *Remote Sens.* **2012**, *4*, 2210–2235. [CrossRef]

15.	Pang, Y.; Michael, L.; Hans-Erik, A.; Ellen, M.M.; Kirk, S. Validation of the ICEsat vegetation product using crown-area-weighted mean height derived using crown delineation with discrete return lidar data. *Can. J. Remote Sens.* **2008**, *34*, S471–S484. [CrossRef]

16.	Enßle, F.; Heinzel, J.; Koch, B. Accuracy of vegetation height and terrain elevation derived from ICESat/GLAS in forested areas. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *31*, 37–44. [CrossRef]

17.	Gwenzi, D.; Lefsky, M.A. Modeling canopy height in a savanna ecosystem using spaceborne lidar waveforms. *Remote Sens. Environ.* **2014**, *154*, 338–344. [CrossRef]

18.	Montesano, P.M.; Sun, G.; Dubayah, R.; Ranson, K.J. The Uncertainty of Plot-Scale Forest Height Estimates from Complementary Spaceborne Observations in the Taiga-Tundra Ecotone. *Remote Sens.* **2014**, *6*, 10070–10088. [CrossRef]

19.	Peterson, B.; Nelson, K.J. Mapping Forest Height in Alaska Using GLAS, Landsat Composites, and Airborne LiDAR. *Remote Sens.* **2014**, *6*, 12409–12426. [CrossRef]

20.	Lefsky, M.A. A global forest canopy height map from the moderate resolution imaging spectroradiometer and the geoscience laser altimeter system. *Geophys. Res. Lett.* **2010**, *37*, L15401. [CrossRef]

21.	Houghton, R.A.; Hall, F.; Goetz, S.J. Importance of biomass in the global carbon cycle. *J. Geophys. Res.* **2009**, *114*, G00E0. [CrossRef]

22.	Hall, F.G.; Bergen, K.; Blair, J.B.; Dubayah, R.; Houghton, R.; Hurtt, G.; Kellndorfer, J.; Lefsky, M.; Ranson, J.; Saatchi, S.; *et al.* Characterizing 3D vegetation structure from space: Mission requirements. *Remote Sens. Environ.* **2011**, *115*, 2753–2775. [CrossRef]

23.	Carabajal, C.C.; Harding, D.J. SRTM C-band and ICESat laser altimetry elevation comparisons as a function of tree cover and relief. *Photogram. Eng. Remote Sens.* **2006**, *72*, 287–298. [CrossRef]

24.	Fayad, N.; Baghdadi, J.-S.; Bailly, N.; Barbier, V.; Gond, M.; Hajj, E.L.; Fabre, F.; Bourgine, B. Canopy Height Estimation in French Guiana with LiDAR ICESat/GLAS Data Using Principal Component Analysis and Random Forest Regressions. *Remote Sens.* **2014**, *6*, 11883–11914. [CrossRef]

25.	Harding, D.J.; Carabajal, C.C. ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure. *Geophys. Res. Lett.* **2005**, *32*, L21S10. [CrossRef]

26.	Chen, Q. Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry. *Remote Sens. Environ.* **2010**, *114*, 1610–1627. [CrossRef]

27.	Delor, C.; Lahondère, D.; Egal, E.; Marteau, P. *Carte Géologique de la France à 1/500 000. Département de la Guyan*, 2nd ed.; BRGM, Service Géologique National: Orléans, France, 2001.

28.	Gond, V.; Freycon, V.; Molino, J.-F.; Brunaux, O.; Ingrassia, F.; Joubert, P.; Pekel, J.-F.; Prévost, M.-F. Broad-scale spatial pattern of forest landscape types in the Guiana Shield. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 357–367. [CrossRef]

29.	Freitas, S.R.; Mello, M.C.; Cruz, C.B. Relationships between forest structure and vegetationindices in Atlantic Rainforest. *For. Ecol. Manag.* **2005**, *218*, 353–362. [CrossRef]

30.	Pascual, C.; Garcia-Abril, A.; Cohen, W.; Martin-Fernandez, S. Relationship between LiDAR-derived forest canopy height and Landsat images. *Int. J. Remote Sens.* **2010**, *31*, 1261–1280. [CrossRef]

31.	Bourgine, B.; Baghdadi, N. Assessment of C-band SRTM DEM in a dense equatorial forest zone. *C. R. Geosci.* **2005**, *337*, 1225–1234. [CrossRef]

32.	Givnish, T.; Wong, S.; Stuart-Williams, H.; Holloway-Phillips, M.; Farquhar, G. Determinants of maximum tree height in Eucalyptus species along a rainfall gradient in Victoria, Australia. *Ecology* **2014**, *95*, 2991–3007. [CrossRef]

33.	Baccini, A.; Laporte, N.; Goetz, S.J.; Sun, M.; Dong, H. A first map of tropical Africa's above-ground biomass derived from satellite imagery. *Environ. Res. Lett.* **2008**, *3*, 049001. [CrossRef]

34.	Powell, S.L.; Cohen, W.B.; Healy, S.P.; Kennedy, R.E.; Moisen, G.G.; Pierce, K.B.; Ohmann, J.L. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sens. Environ.* **2010**, *114*, 1053–1068. [CrossRef]

35.	Hengl, T.; Heuvelink, G.B.; Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **2004**, *120*, 75–93. [CrossRef]

36.	Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

37.	Ota, T.; Ahmed, O.; Franklin, S.; Wulder, M.; Kajisa, T.; Mizoue, N.; Yoshida, S.; Takao, G.; Hirata, Y.; Furuya, N.; *et al.* Estimation of Airborne Lidar-Derived Tropical Forest Canopy. *Remote Sens.* **2014**, *6*, 10750–10772. [CrossRef]

38. Baghdadi, N.; le Maire, G.; Osé, K.; Nouvellon, Y.; Mehrez, Z.; Lemos, C.; Hakamada, R. Evaluation of ALOS/PALSAR L-band data for the estimation of Eucalyptus plantations aboveground biomass in Brazil. *IEEE JSTARS* **2014**, *8*, 3802–3811. [CrossRef]

39. Genuer, R.; Poggi, J.-M.; Tuleau-Malot, C. Variable selection using Random Forests. *Pattern Recognit. Lett.* **2010**, *31*, 2225–2236. [CrossRef]

40. Sun, W.; Minasny, B.; McBratney, A. Analysis and prediction of soil properties using local regression-kriging. *Geoderma* **2012**, *171*, 16–23. [CrossRef]

41. Odeh, I.O.A.; McBratney, A.B.; Chittleborough, D.J. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. *Geoderma* **1995**, *67*, 215–226. [CrossRef]

42. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, NY, USA, 1997.