Canopy height estimation on a regional scale : Application to French Guiana

Ibrahim Fayad

THESE DE DOCTORAT

DE L'UNIVERSITE DE MONTPELLIER

Spécialité

Systèmes Automatiques et Microélectroniques (*SYAM*)

(Ecole Doctorale: Information, Structures, Systèmes)

Présentée par

Ibrahim FAYAD

Pour obtenir le grade de

Docteur de l'Université de Montpellier

ESTIMATION DE LA HAUTEUR DES ARBRES A L'ECHELLE REGIONALE : APPLICATION A LA GUYANE FRANÇAISE

Soutenue le 15/06/2015, devant le jury composé de :

Dr. Patrick Chazette
Ingénieur-chercheur, CEA,
Rapporteur

Dr. Xavier Briottet
Directeur de recherche, Onera,
Examinateur

Dr. Nicolas Baghdadi
Directeur de recherche, Irstea,
Directeur de thèse

Dr. Grégoire Vincent
Chargé de recherche, IRD,
Rapporteur

Dr. Mehrez Zribi
Directeur de recherche, CNRS,
Examinateur

Dr. Nicolas Barbier
Chargé de recherche, IRD,
Co-directeur de thèse

ABSTRACT

Remote sensing has facilitated the techniques used for the mapping, modeling and understanding of forest parameters. Remote sensing applications usually use information from either passive optical systems or active radar sensors. These systems have shown satisfactory results for estimating, for example, aboveground biomass in some biomes. However, they presented significant limitations for ecological applications, as the sensitivity from these sensors has been shown to be limited in forests with medium levels of aboveground biomass. On the other hand, LiDAR remote sensing has been shown to be a good technique for the estimation of forest parameters such as canopy heights and above ground biomass. Whilst airborne LiDAR data are in general very dense but only available over small areas due to the cost of their acquisition, spaceborne LiDAR data acquired from the Geoscience Laser Altimeter System (GLAS) have low acquisition density with global geographical cover. It is therefore valuable to analyze the integration relevance of canopy heights estimated from LiDAR sensors with ancillary data (geological, meteorological, slope, vegetation indices etc.) in order to propose a forest canopy height map with good precision and high spatial resolution. In addition, estimating forest canopy heights from large-footprint satellite LiDAR waveforms, is challenging given the complex interaction between LiDAR waveforms, terrain, and vegetation, especially in dense tropical and equatorial forests. Therefore, the research carried out in this thesis aimed at: 1) estimate, and validate canopy heights using raw data from airborne LiDAR and then evaluate the potential of spaceborne LiDAR GLAS data at estimating forest canopy heights. 2) evaluate the fusion potential of LiDAR (using either sapceborne and airborne data) and ancillary data for forest canopy height estimation at very large scales. This research work was carried out over the French Guiana.

The estimation of the canopy heights using the airborne dataset has been carried out using a simple algorithm, which first extracts the canopy top and ground points, and then interpolates the canopy height using the ground point and its surrounding canopy top points. Results indicated an RMSE on the canopy height estimates of 1.6 m. Next, the potential of GLAS for the estimation of canopy heights was assessed using multiple linear (ML) and Random Forest (RF) regressions using waveform metrics and principal component analysis (PCA). Results showed canopy height estimations with similar

precisions using either LiDAR metrics or the principal components (PCs) (RMSE ~ 3.6 m). However, a regression model (ML or RF) based on the PCA of waveform samples is an interesting alternative for canopy height estimation as it does not require the extraction of some metrics from LiDAR waveforms that are in general difficult to derive in dense forests, such as those in French Guiana.

Next, canopy heights extracted from both airborne and spaceborne LiDAR were first used to map canopy heights from available mapped environmental data (geological, meteorological, slope, vegetation indices etc.). Results showed an RMSE on the canopy height estimates of 6.5 m from the GLAS dataset and of 5.8 m from the airborne LiDAR dataset. Then, in order to improve the precision of the canopy height estimates, regression-kriging (kriging of random forest regression residuals) was used. Results indicated a decrease in the RMSE from 6.5 to 4.2 m for the regression-kriging maps from the GLAS dataset, and from 5.8 to 1.8 m for the regression-kriging map from the airborne LiDAR dataset. Finally, in order to study the impact of the spatial sampling of future LiDAR missions on the precision of canopy height estimates, six subsets were derived from the airborne LiDAR dataset with flight line spacing of 5, 10, 20, 30, 40 and 50 km (corresponding to 0.29, 0.11, 0.08, 0.05, 0.04, and 0.03 points/km², respectively).

Results indicated that using the regression-kriging approach, the precision on the canopy height map was 1.8 m with flight line spacing of 5 km and decreased to an RMSE of 4.8 m for the configuration for the 50 km flight line spacing.

RESUME

La télédétection contribue à la cartographie et à la modélisation des paramètres de la forêt. Ce sont les systèmes optiques et radars qui sont le plus généralement utilisés pour extraire des informations utiles à la caractérisation des paramètres forestiers. Ces systèmes ont montré des résultats satisfaisants pour estimer, par exemple, la biomasse dans certains biomes. Cependant, ils présentent des limitations importantes pour des forêts ayant un niveau de biomasse élevé. En revanche, la télédétection LiDAR s'est avérée être une bonne technique pour l'estimation des paramètres forestiers tels que la hauteur de la canopée et la biomasse. Alors que les LiDAR aéroportés acquièrent en général des données avec une forte densité de points mais sur des petites zones en raison du coût de leurs acquisitions, les données LiDAR satellitaires acquises par le système spatial (GLAS) ont une densité d'acquisition faible mais avec une couverture géographique mondiale. Il est donc utile d'analyser la pertinence de l'intégration des hauteurs estimées à partir des capteurs LiDAR et des données auxiliaires (géologiques, météorologiques, pente, indices de végétation, etc.) afin de proposer une carte de la hauteur des arbres avec une bonne précision et une résolution spatiale élevée. En outre, l'estimation de la hauteur des arbres à partir des formes d'onde GLAS avec ses grandes empreintes est difficile compte tenu de l'interaction complexe entre les formes d'onde LiDAR, le terrain et la végétation, en particulier dans les forêts tropicales et équatoriales denses. Par conséquent, la recherche menée dans cette thèse vise à: 1) Estimer et valider la hauteur des arbres en utilisant des données acquises par des LiDAR aéroportés et satellitaire (capteur GLAS). 2) évaluer le potentiel de la fusion des données LiDAR (avec les données aéroportées ou satellitaires) et des données auxiliaires pour l'estimation de la hauteur des arbres à une échelle régionale. Ce travail de recherche a été effectué sur la Guyane française.

L'estimation de la hauteur des arbres en utilisant les données aéroportées a été réalisée en utilisant un algorithme simple, qui extrait d'abord les points haut de la canopée et ceux du sol, puis interpole la hauteur de la canopée en utilisant les points du sol et les points hauts de la canopée. Les résultats ont indiqué une EQM sur les estimations de la hauteur de la canopée de 1,6 m. Ensuite, le potentiel de GLAS pour l'estimation de la hauteur des arbres a été évalué en utilisant des modèles de régression linéaire (ML) ou Random Forest (RF) avec des métriques provenant de la forme d'onde et de l'analyse en composantes principales

(ACP). Les résultats ont montré que les modèles d'estimation des hauteurs des arbres avaient des précisions semblables en utilisant soit les métriques LiDAR ou les composantes principales (PC) (EQM ~ 3,6 m). Toutefois, un modèle de régression (ML ou RF) basé sur les composantes principales obtenues à partir des formes d'onde GLAS est une alternative intéressante pour l'estimation de la hauteur des arbres, car il ne nécessite pas l'extraction de certaines métriques à partir des formes d'onde LiDAR qui sont en général difficiles à dériver dans les forêts denses, telle que la Guyane française.

Finalement, la hauteur des arbres extraite à la fois des données LiDAR aéroporté et GLAS a servi tout d'abord à spatialiser la hauteur des arbres en utilisant les données environnementaux cartographiées disponibles (géologiques, météorologiques, la pente, indices de végétation, etc.). En utilisant la régression RF, la spatialisation de la hauteur des arbres a montré une EQM sur les estimations de la hauteur de la canopée de 6,5 m à partir de GLAS et de 5,8 m à partir du LiDAR aéroporté. Ensuite, afin d'améliorer la précision de la spatialisation de la hauteur de la canopée, la technique régression-krigeage (krigeage des résidus de la régression du Random Forest) a été utilisée. Les résultats de la régression-krigeage indiquent une diminution de l'erreur quadratique moyenne de 6,5 à 4,2 m pour les cartes de la hauteur de la canopée à partir de GLAS, et de 5,8 à 1,8 m pour les cartes de la hauteur de la canopée à partir des données LiDAR aéroporté. Enfin, afin d'étudier l'impact de l'échantillonnage spatial des futures missions LiDAR sur la précision des estimations de la hauteur de la canopée, six sous-ensembles ont été extraits de de la base LiDAR aéroporté. Ces six sous-ensembles de données LiDAR ont respectivement un espacement des lignes de vol de 5, 10, 20, 30, 40 et 50 km (correspondant à une densité de 0,29, 0,11, 0,08, 0,05, 0,04, 0,03 points / km², respectivement).

Finalement, les résultats indiquent qu'en utilisant la technique régression-krigeage, l'erreur quadratique moyenne sur la carte des hauteurs de la canopée était de 1,8 m pour le sous-ensemble ayant des lignes de vol espacés de 5 km, et a augmentée jusqu'à 4,8 m pour le sous-ensemble ayant des lignes de vol espacés de 50 km

- To my late uncle, Samih Fayad for convincing me to pursue this path in life

- To Hanine, I give my deepest expression of appreciation for the encouragement that you gave and the sacrifices you made during all these years

ACKNOWLEDGEMENTS

I would like to acknowledge Nicolas Baghdadi and Nicolas Barbier, my thesis directors, and Jean-Stephane Bailly and Valery Gond for their guidance throughout my years of study.

I would also like to acknowledge Richard Bru, Mahmoud el-Hajj, and Frederic Fabre, members of my thesis progress-committee, whom advice was invaluable for the advancement of the thesis.

I would like to thank Noveltis and Airbus Defense and Space for their financial support. Sincere thanks goes to Richard Bru, president and CEO of Noveltis, for his interest in my work, and his real investment in the training of young researchers. I extend my thanks to Irstea, for funding my work, and giving me the opportunity to work on this thesis.

I would like to acknowledge UMR TETIS and in particular its director Jean-Philippe Tonneau for hosting me during these three years. I would also like to thank the management team at UMR TETIS, especially Françoise Boissier, Coralie Bastide, and Nathalie Jean. A special thank you also goes to Isabelle, and Veronique.

Finally, I would like to thank my fellow PhD students and my friends at UMR TETIS

TABLE OF CONTENTS

INTRODUCTION 1

1.1 General context 1

1.1.1 Global Carbon Cycle 1

1.1.2 Greenhouse gases and climate change 2

1.1.3 Global carbon cycle's carbon sinks 3

1.1.4 Humans and climate change 4

1.1.5 The carbon cycle feedback loop 4

1.1.6 Current issues 5

1.2 Role of forests in the carbon cycle 6

1.2.1 Tropical forests and the carbon stock 6

1.2.2 Link between carbon and forest biomass 8

1.2.3 The importance of quantifying forest biomass 8

1.3 Biomass estimation 8

1.3.1 Biomass estimation with optical and radar data 9

1.3.2 Biomass estimation using allometric relations 10

1.3.3 Plot aggregate allometry for biomass estimation 11

1.4 Forest canopy height in relation to forest biomass 11

1.4.1 Canopy height estimation using radar and optical data 12

1.4.2 Canopy height estimation using LiDAR data 13

1.4.3 Spatial extrapolation of LiDAR canopy height estimates 14

1.5 Forest types in relation to forest biomass 15

1.6 Organization of the dissertation 16

1.6.1 Objectives 16

1.6.2 Dissertation plan 17

CHAPTER 2: STUDY AREA AND DATASETS 19

2.1 Study area 19

2.2 Datasets description 20

2.2.1 Spaceborne LiDAR datasets 20

2.2.2 Airborne LiDAR datasets 22

2.2.2.1 Small footprint low density LiDAR dataset 22
2.2.2.2 Small footprint high density LiDAR dataset 23
2.2.3 Ancillary datasets 24
2.2.3.1 MODerate-resolution Imaging Spectroradiometer (MODIS) data 24
2.2.3.2 SRTM digital elevation model data 26
2.2.3.3 Geological map 28
2.2.3.4 Forest landscape types map 28
2.2.3.5 Average rainfall map 28

CHAPTER 3: CANOPY HEIGHT ESTIMATION IN FRENCH GUIANA WITH LIDAR ICESAT/GLAS DATA USING PRINCIPAL COMPONENT ANALYSIS AND RANDOM FOREST REGRESSIONS 31

3.1 Introduction 31

3.2 Materials and methods 32
3.2.1 Lidar data processing and canopy height estimation 32
3.2.1.1 Processing the LD dataset 32
3.2.1.2 Processing the HD dataset 38
3.2.1.3 Comparison of canopy height estimates from the HD dataset using different estimation methods 39
3.2.1.4 Comparison of canopy height estimates from the LD and HD datasets 41
3.2.1.5 Glas data processing 41
3.2.2 Background on GLAS canopy height estimation 44
3.2.2.1 Direct method 44
3.2.2.2 Multiple regression models using GLAS and DEM metrics 44
3.2.2.3 Proposed techniques for canopy height estimation 48
3.2.2.4 Random forest regressions using principal components 51

3.3 Results 51
3.3.1 Direct method 51
3.3.2 Multiple regression models 51
3.3.2.1 Using GLAS and DEM metrics 51
3.3.2.2 Using principal components 53
3.3.3 Random forest regressions 58
3.3.3.1 Using GLAS and DEM metrics 58
3.3.3.2 Using principal components 60
3.3.4 Model performance in different forest conditions 61
3.3.5 Error on the estimation of biomass 63

3.4 Discussion **64**

3.5 Conclusions **66**

CHAPTER 4: FOREST CANOPY HEIGHT MAPPING OVER FRENCH GUIANA USING SPACE AND AIRBORNE LIDAR DATA **69**

4.1 Introduction **69**

4.2 Materials and methods **70**

4.2.1 Canopy height mapping using regression-kriging 70

4.2.2 Canopy height trend mapping using Random Forest regressions 71

4.2.3 Ordinary krigging of regression residuals 72

4.2.4 Effects of LiDAR sampling density on precision of the mapped canopy heights. 73

4.3 Results **75**

4.3.1 Canopy height mapping using Random Forest regressions 75

4.3.2 Canopy height estimation using regression-kriging 79

4.3.3 Relationship between LiDAR flight lines spacing and the accuracy on the kriged canopy height 84

4.4 Discussion **93**

4.5 Conclusions **96**

CHAPTER 5: COUPLING POTENTIAL OF ICESAT/GLAS AND SRTM FOR THE DISCRIMINATION OF FOREST LANDSCAPE TYPES IN FRENCH GUIANA **99**

5.1 Introduction **99**

5.2 Materials and methods **102**

5.2.1 Methodology 102

5.2.2 GLAS waveform processing 107

5.2.3 Canopy height and roughness index estimations 107

5.3 Results and discussion **108**

5.3.1 Global analysis of the differences between the GLAS and SRTM elevations 108

5.3.2 Analysis of the differences between the GLAS and SRTM according to Hc and R 109

5.3.2.1 Differences between the GLAS and SRTM according to Hc 110

5.3.2.2 Differences between the GLAS and SRTM according to R 111

5.3.3 Random Forest classification results 115
5.3.4 Effect of the GLAS acquisition season 119

5.4 Conclusions 123

GENERAL CONCLUSIONS AND PERSPECTIVES 125

6.1 Conclusions 125

6.2 Perspectives 129
6.2.1 Canopy height estimation using GLAS 129
6.2.2 LiDAR canopy height mapping 130
6.2.2.1 Non-spatial canopy height mapping 130
6.2.2.2 Spatial canopy height mapping 130
6.2.2.3 Canopy height map resolution 131
6.2.2.4 Canopy height mapping sampling scheme 131
6.2.3 Above-ground biomass estimation 131

RESUME 135

7.1 Introduction 135

7.2 Description des jeux de données 139
7.2.1 Site d'étude 139
7.2.2 Base de données LiDAR satellitaire 140
7.2.3 Données du radiomètre spectral à moyenne résolution MODIS 140
7.2.4 Données issues du Modèle Numérique de Terrain MNT SRTM 141
7.2.5 Carte géologique 142
7.2.6 Carte des types de paysage forestier 142
7.2.7 Carte de précipitation 143

7.3 Estimation de la hauteur des arbres à partir des données GLAS 143
7.3.1 Contexte de l'estimation de la hauteur des arbres en utilisant GLAS 144
7.3.2 Techniques proposées pour l'estimation de la hauteur des arbres 146

7.4 La spatialisation de la hauteur des arbres LiDAR 147
7.4.1 Contexte sur la technique régression-krigeage 148
7.4.2 La cartographie de la hauteur des arbres en utilisant la régression krigeage 148
7.4.3 Relation entre l'espacement des lignes de vol LiDAR et la précision de la hauteur des arbres krigée 149

7.5 Le potentiel du couplage GLAS et SRTM pour la discrimination des types de paysage forestier 150

7.5.1 Classifications des empreintes GLAS 150

7.5.2 Les effets de la saison sur les acquisitions GLAS 151

7.6 Conclusions et perspectives 152

7.6.1 Conclusions 152

7.6.2 Perspectives 153

7.6.2.1 La spatialisation de la hauteur des arbres à partir du LiDAR 153

7.6.2.2 L'estimation de la biomasse 154

REFERENCES 166

INTRODUCTION

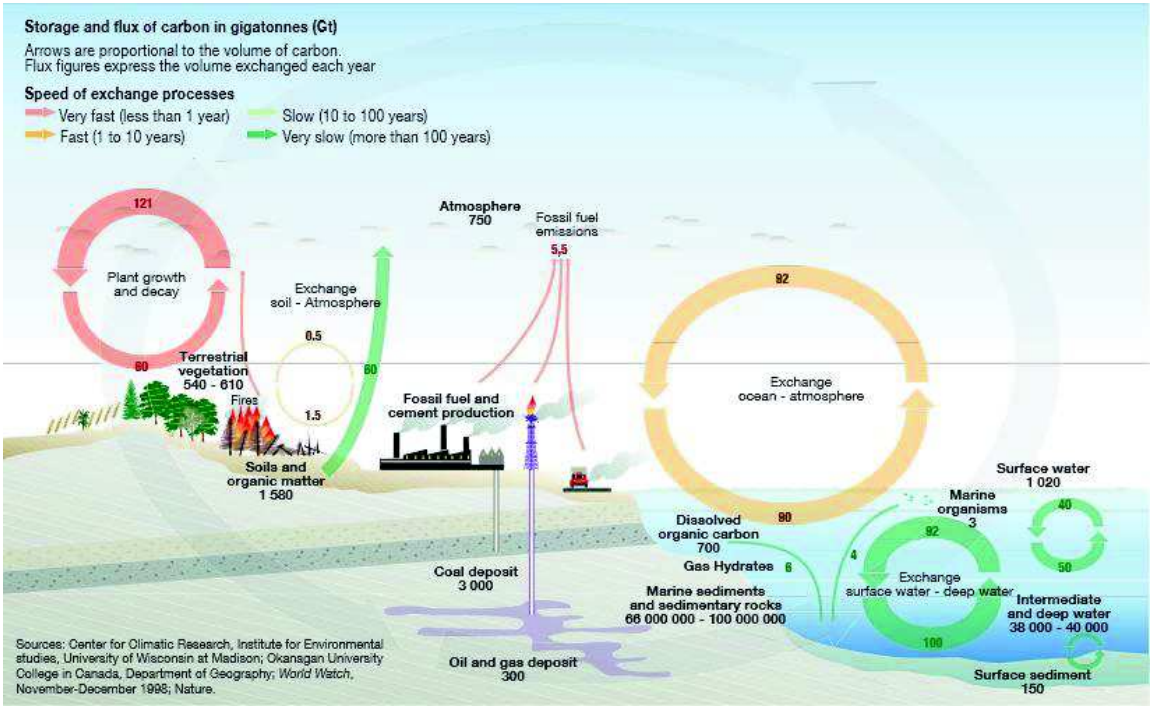
1.1 General context

1.1.1 Global Carbon Cycle

The carbon cycle is the biogeochemical cycle (total exchange of a chemical element) of carbon globally. The earth's carbon cycle is rendered more complex by the existence of large oceanic water masses and especially by the fact that life (and therefore Carbone compounds that are the substrate) has an important place. There are mainly four carbon reservoirs: the hydrosphere, lithosphere, biosphere and the atmosphere. Most of the terrestrial carbon is trapped in compounds that contribute little to the cycle: rocks as carbonates and deep oceans. Therefore most of the cycle is between the atmosphere, the surface layers of soil and oceans, and biosphere (biomass and necromass). Under seas, the carbon is found mostly as carbonate and planktonic biomass. Over land, the carbon is mainly the bogs, meadows and forests. In addition, some soil types play a fairly important role in carbon sequestration or as a carbon sinks. Figure 1.1, shows the global carbon cycle, as well as the exchange of carbon between the different carbon sinks, and the carbon fluxes. The carbon cycle is very important to the biosphere, since life is based on the use of carbon-based compounds: carbon availability is one of the key factors for the development of all living things on earth. Carbon is also a major component of many minerals, and the carbon

dioxide (CO_2) is partly responsible for the greenhouse effect and is the most human-contributed greenhouse gas ([1]).

Figure 1.1. The global carbon cycle with the movement and exchange of carbon between land, atmosphere, and oceans.



1.1.2 Greenhouse gases and climate change

The study of the carbon cycle has recently taken a special relief in the context of the issue of global warming: Two of the greenhouse gases involved: the carbon dioxide (CO2) and methane (CH4), participate in the carbon cycle, as they are the main atmospheric carbon forms. In addition to climate issues, the study of the carbon cycle will allow us to determine the effects on the release of carbon stored in the form of fossil fuels by human activity.

In fact, the global carbon cycle has been greatly altered by human activity in the past decades. Indeed, carbon dioxide resulting from human emissions exceeded natural fluctuations ([1]). The changes in the amount of CO_2 in the atmosphere are altering weather patterns and oceanic chemistry. Studies have shown that even though global temperatures can fluctuate without changes in atmospheric CO_2, the latter cannot change without affecting the atmospheric temperatures. In addition, CO_2 levels are rising higher than ever recorded in the atmosphere ([2]). Therefore it is of high importance to better understand the carbon cycle and its effects on the global climate ([1]).

1.1.3 Global carbon cycle's carbon sinks

The global carbon cycle is divided into four main carbon sinks connected by pathways of exchange ([3]):

- The lithosphere contains carbon in its carbon and carbonated rocks (30 mGt).
- The hydrosphere contains carbon in its dissolved form (38 000 Gt) and in marine organisms (3 Gt).
- The biosphere contains 2,300 Gt of carbon in the form of biomass and necromass and in soils
- The atmosphere contains 700 Gt of carbon as CO2.

The exchanges of carbon between these fours sinks occur as a result of various chemical, physical, geological, and biological processes. The ocean contains the largest active sink of carbon near the surface of the earth ([1]). In addition, carbon exchange between the different compartments is balanced, which makes the carbon levels stable without human influence ([4]).

The lithosphere contains the largest amounts of carbon in the form of carbonated rocks and fossil fuel ([1]); it does not exchange a lot of carbon naturally with the other compartments. This is due to the fossilization rate of living beings or the sedimentation of carbonated rocks which can take several million years. However, the CO_2 emissions in the atmosphere resulting from the use of fossil fuel are the principal flux that concerns this carbon stock.

The hydrosphere and the biosphere are in equilibrium due to the high solubility of the CO_2 in water and the important volume of oceans. In fact, oceanic absorption of CO_2 is one the most important forms of carbon sequestering. This high absorption rate limits the carbon dioxide in the atmosphere caused by human activities. However, this process may make oceanic waters more acidic due to the increase uptake of carbon, as well as limiting the ocean uptake of CO_2 ([1]).

Finally, the biosphere exchanges up to 60 Gt/year of carbon with the atmosphere. This exchange has two sources, while the breathing of animals and plants and fermentation of bacteria releases CO_2 into the atmosphere; the photosynthesis (especially of green plants) fixes the carbon in the biomass. The biosphere plays an important role in the carbon cycle,

as this compartment is directly influenced by human activity. While it is possible to interact with this compartment, on the one hand, deforestation and land use change can diminish carbon stocks ([5]). On the other hand, tree planting and the protection of existing forests increase carbon stocks ([6]).

1.1.4 Humans and climate change

The concentration of atmospheric carbon during the last 100-200 years increased significantly due to human activities (burning of fossil fuel, natural gas, charcoal, etc.). The burning of fossil fuels, which accumulated during millions of years, released huge amounts of CO_2. Another reason for the increase of CO_2 in the atmosphere comes from deforestation and forest fires, especially in tropical regions. This also causes fast release of CO_2 sinks that were also accumulated during a long time (few years to several centuries based on burnt forest age) ([1]).

By determining the contribution of CO_2 to the atmosphere, we can deduce how the carbon cycle influences the global temperature. The rejection of CO_2 of anthropogenic origins is responsible for 70% of the global warming, but in return, the atmospheric concentrations of CO_2, the global temperature as well as the precipitation affect greatly the carbon cycle.

1.1.5 The carbon cycle feedback loop

Feedback in general is the process in which output from a system are “fed back” as inputs as part of a chain of cause-and-effect that forms a loop. For instance, by determining the contribution of CO_2 to the atmosphere, the carbon cycle influences the global temperature. But, in return, the atmospheric concentrations of CO_2, the global temperature as well as the precipitations influence several key elements of the carbon cycle.

At the oceanic levels, there is a complex feedback linked to the solubility of CO_2. This feedback is negatively correlated to the temperature. In the case of global warming, more CO_2 are liberated from oceans into the atmosphere, and therefore contribute to the global warming. This is called a positive feedback. However, the solubility of CO_2 depends on its concentration in the atmosphere, thus limiting the effect of the feedback. The dissolution of CO_2 in the oceans causes water acidification. Temperature changes are therefore

influencing the activity of the plankton, which increases or decreases the oceanic ability to capture CO_2 ([7]; [8]).

In regards to vegetation and thus forests, if the ratio of photosynthesis increases with temperature and CO_2, the ratio of the respiration will also increase with temperature. This effect on photosynthesis is generally positive. An increase in terrestrial vegetation has been observed in response to higher temperatures and CO_2 levels in the atmosphere (IPCC, 2014 [9]). However, for certain vegetation types, it has been observed that the respiration increases more as a function of temperature rather than photosynthesis, this makes these ecosystems more as sources and not sinks of carbon in the long term.

1.1.6 Current issues

Facing these environmental threats, the international community adopted several policies at the national, international and global level. The first United Nations summit concerning the environment took place in 1972 in Stockholm. It was during this summit that the United Nations Environment Program (UNEP) was created in order to debate ecological questions. The countries participating to this summit agreed to meet once each ten years in order to review the state of earth's environment. Following that year, the most notable summits were as follows:

The Montreal protocol of 1987 which prohibited the chlorofluorocarbons gas use (CFC) as it can lead to the destructions of the atmosphere was successful as it allowed the decrease of atmospheric charges of the CFC ([10]). This first success is still limited because of climate change with the massive injection of greenhouse gases, including firstly CO2, which could destabilize the stratosphere, and amplify the loss of the ozone layer in the atmosphere. The changing climates has socio-economic effects and these effects are already being felt, as they lead to the exodus of some populations worldwide, but also break the balance governing ecosystems and jeopardize the biodiversity of our planet. This led to the creation of the UN Framework Convention on Climate Change (UNFCCC), which came into force in 1994 following the Earth Summit in Rio de Janeiro in 1992. In the Rio Janeiro summit in 1992, the participants agreed on the necessity to stabilize atmospheric concentrations of greenhouse gases. The objective was to limit the abrupt changes to ecosystems, in order to have time to adapt. In 1997, 141 nations signed on the protocol of

Kyoto, which engaged the committed nations to reduce by 5.2% their emissions of six greenhouse gases. Recently, the Copenhagen conference which brought together 191 countries, have ratified the UNFCCC. The UNFCCC stressed the importance of forests in regulating climate change and particularly of atmospheric CO2.

Countries in economic development have no commitments in this protocol along with the United States and the main carbon emitters who did not sign. Practically, this agreement allowed the creation of a carbon market. The states which surpass their quota in their carbon emission, can buy carbon credits from other nations that have not surpassed their carbon quotas. These credits allow the nations in need to emit more greenhouse gases. The objective was to motivate the nations to limit their greenhouse gases emissions by giving a monetary value to these emissions. The agreements of Copenhagen, which were signed in 2009, were renegotiations of the agreements of Kyoto. However, no binding commitments were made after the 2012, which marks the end of the Kyoto protocol. However, the 112 participating nations agreed to try and reduce the global temperatures rise by 2^{o}C.

1.2 Role of forests in the carbon cycle

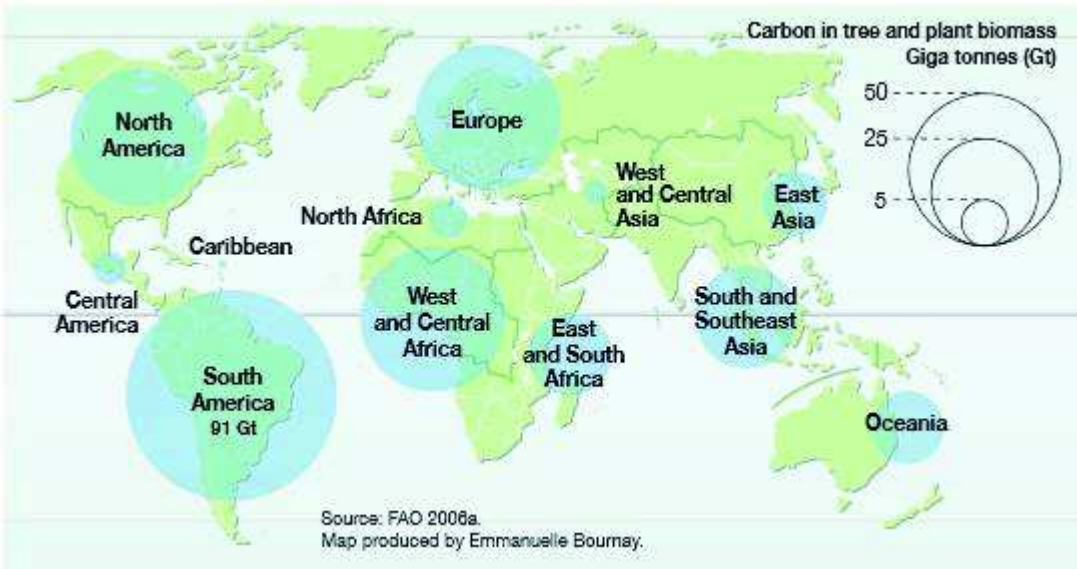
In the framework of the international agreements on the limitation of emission of greenhouse gases and temperature emissions, the case of forests and in tropical forest plays a major role. Carbon stocks in forests comprise above- and below-ground carbon in both living and dead organic matter. Globally, forests and soils are estimated to trap around 2.6 GtC/year. However, there are still many uncertainties about the carbon cycle. Indeed, Food and Agriculture Organization of the United Nations (FAO, 2008 [11]) estimates that the amount of carbon absorbed by the forests can vary between 0.9 and 4.3 GtC/year.

1.2.1 Tropical forests and the carbon stock

Carbon Stocks over land are distributed mostly between forests and northern latitudes (Figure 1.2), but are mostly found in forests, and more precisely in tropical forests. Indeed, studies suggests that tropical forests play a more important role in absorbing carbon with an absorption rate reaching as much as 1 GtC/year or about 40% of the total land based carbon absorption globally. However, tropical forests are principally located in developing

countries (Amazon basin, Congo Basin, South-East Asia). These countries which are currently undergoing an economic and demographic growth, and therefore moving from forested to non-forested areas are causing a significant impact on the accumulation of greenhouse gases in the atmosphere, as has forest degradation caused by over-exploitation of forests for timber and wood fuel and intense grazing that is reducing forest regeneration. Therefore, during the 16th conference of the parties to the agreement of climate change of Cancun (2010), the United Nations program for Reducing Emissions from Deforestation and Forest Degradation (UN-REDD) was adopted. This program aims at protecting forests, preserve and increase forest carbon stocks and sustainable forest management. The REDD initiative and its three main supplementary activities are called REDD+. The basic principle of the REDD+ program is that financial compensation be paid by the developed countries to developing countries that manage to reduce their emissions at the national level. The REDD program is based on the fact that when a forest is damaged and destroyed, CO_2 is released into the atmosphere. If we manage to reduce the rate of deforestation (complete disappearance of forests) or degradation (damaged forest due to exploitation), then it is possible to reduce the amount of released CO_2. However, in order to calculate the magnitude of the reduction in CO_2 emissions, it is necessary to create a baseline or reference base against which to compare actual emissions. Therefore it is necessary to be able to quantify the amounts of carbon contained in forests.

Figure 1.2. Forest carbon stock per region. UNEPP, FAO, UNFF, Forest vital graphics, 2009.



1.2.2 Link between carbon and forest biomass

Studies have stated that more than 40% of global vegetation carbon stocks are located in tropical forests ([6]; [12]). However, forest carbon is not limited to trees and is distributed on average as follows: 45% of carbon is found in the soil, 11% in dead biomass or necromass, and 44% in biomass (both above- and below-ground) (FAO, 2000 [13]). Moreover, the above-ground biomass (AGB) is generally the most studied, as it is the most accessible. AGB is a biological material derived from living organisms, and it most often refers to plants. Biomass is carbon based and is composed of a mixture of organic molecules containing hydrogen, oxygen, and small quantities of several other atoms. The proportion of carbon in AGB varies depending on the forest type, wood composition, or the environment. However, it ranges between 0.43 and 0.55 ([14]; [15]; [16]; [17]; [18]).

1.2.3 The importance of quantifying forest biomass

The interest in studying the AGB comes from the fact that the carbon in the AGB is susceptible to be released into the atmosphere by means of deforestation. In addition, land use change in tropical forests is responsible of 15-20% of global greenhouse emissions globally ([19]; [20]). In contrast, if trees are to be planted, this means more carbon sequestration. However, this natural regeneration of the carbon stock will much likely take several decades ([21]), and a plantation is not, by far, a natural forest. Moreover, even with forest degradation or regeneration, tropical forest can still undergo changes that affect AGB levels. For example, under influence of environmental changes, such as the increase of CO_2 levels in the atmosphere. This increase of CO_2 might increase the photosynthesis of trees and therefore increase the levels of carbon in trees ([6]; [22]). Other environmental changes are caused by tree mortality, which can increase the necromass, and therefore affect the release of carbon in the atmosphere ([23]).

1.3 Biomass estimation

As seen earlier, AGB measurement is an important task for better understanding of the carbon cycle. However, accurate measurements of biomass require weighing of the trees after cutting them. This method yields high biomass measurement accuracy however it is destructive and restrictive. Therefore it is necessary to find other methods for biomass estimation in a non-destructive manner.

1.3.1 Biomass estimation with optical and radar data

Currently, existing AGB estimation methods from remote sensing data are either limited in the vertical domain (sensor saturation at certain biomass levels using mainly radar and optical data) or in the horizontal domain (limited horizontal coverage using LiDAR data). Methods using radar and optical data for the estimation of AGB are successful in forests with low to medium levels of AGB (e.g. [24]; [25]; [26]; [27]; [28]). Indeed, current techniques based on passive optical sensing have shown limited sensitivity to biomass using medium to high resolution imagery when the biomass reaches intermediate levels (150-200 Mg/ha) (e.g. [27]; [28]). This is due to the optical data inability to detect variation in biomass density after complete closure of the canopy top, which can occur from low or intermediate biomass values (depending on forest characteristics). In contrast, the Fourier Transform Textural Ordination (FOTO) using very-high-resolution optical images have been used for non-saturating estimates of tropical forest biomass estimation. As such, this approach may provide higher sensitivity to biomass high levels (>600 t/ha) (e.g. [29]; [30]; [31]).

The synthetic aperture radar (SAR) systems such as PALSAR/ALOS, JERS-1 and SIR-C, as well as airborne SAR such as SETHI and E-SAR were also used as an alternative for biomass estimation. The radar signal saturation threshold with the biomass increases with the increase of the radar wavelength. Indeed, L-band SAR systems (wavelength about 25 cm) are limited to low and intermediate biomass levels, with maximum values reaching 150 t/ha (e.g. [24]; [25]; [26]; [32]; [33]; [34]). This saturation threshold of the radar signal depends on forest characteristics. According to Imhoff *et al.* [35]; the saturation levels are closer to 40 t/ha because the saturation thresholds occur before the regression maxima. In boreal forests, saturation levels were observed up to 150 t/ha. Baghdadi *et al.* [32] observed saturation levels of the ALOS/PALSAR L-band at biomass levels of 50 t/ha when estimating the biomass for Eucalyptus plantations in Brazil. Luckman *et al.* ([36]; [37]) found a saturation point of 60 t/ha in the Central Amazon basin. Le Toan *et al.* [26]; Wu *et al.* [33]; and Dobson *et al.* [34] reported L-band signal saturation levels at 100 t/ha in coniferous forests. In boreal forests, higher saturation levels were observed reaching up to 150 t/ha using PALSAR (Sandberg *et al.* [25]). However, with higher radar wavelengths (P-band for example, wavelength about 70 cm) the use of SAR sensors may allow the

estimation of biomass at higher biomass levels ([38]). Imhoff *et al.* [35] examined AGB levels in broadleaf evergreen forests in Hawaii and coniferous forests in North America and Europe and found saturation levels of 100 Mg/ha for the P-band versus 40 Mg/ha for the L-band. Nizalapur *et al.* [38] found that the sensitivity of radar signal to biomass in a tropical dry deciduous forest increases for approximately 150 t/ha for the L-band to 200 t/ha for the P-band.

Given the limitations of optical (expensive very high resolution images which only cover small areas) and radar (unavailable global coverage for P-band SAR data, signal saturation with lower wavelengths) data for biomass estimation, studies generally use allometric relations for linking the characteristics of a forest (canopy height, diameter at breast height, wood density) to its biomass (e.g. [39]; [40]; [41]).

1.3.2 Biomass estimation using allometric relations

Allometric relationships linking the characteristics of a forest to its biomass were developed by several studies (e.g. [40]; [42]; [39]). The reference model in these studies was developed in the study of Chave *et al.* [42]. In their study they developed a pantropical biomass estimation model at the individual tree level. This model was based on the formula for calculating the mass of a cylinder using stem diameter (D), canopy height (H), and wood density (ρ).

$$\boldsymbol{M = \pi.\left(\frac{D}{2}\right)^{2}.H.\rho} \tag{1.1}$$

This translates to:

$$\boldsymbol{\log(M) = \log(\pi) + 2.\log(\mathrm{D}) - 2.\log(2) + \log(H) + \log(\rho)} \tag{1.2}$$

Using the second formula (2), it is possible to predict a tree mass (M) by adding adjustment coefficients:

$$\boldsymbol{\log(M) = \beta_0 + \beta_1.\log(D) + \beta_2.\log(H) + \beta_3.\log(\rho)} \tag{1.3}$$

This model developed by Chave *et al.* [42] has been shown to produce good biomass estimation results and fits well with data across different tropical forests ([43]).

1.3.3 Plot aggregate allometry for biomass estimation

Asner *et al.* [40] proposed a plot aggregate allometry model for tropical areas drawn from the Chave *et al.* [42] model, but they replaced *in situ* canopy height with top-of-canopy height (TCH), as derived from airborne small-footprint LiDAR measurements, and stem diameter with plot-averaged basal area (BA). BA and wood density were linked with TCH using linear relationships in the form of $BA = aTCH$ and $\rho = bTCH + c$, producing a model for AGB estimation using only TCH. Results showed a RMSE on AGB estimation of 24.7 Mg/ha for the regional models (model coefficients dependent on region) and 26.4 Mg/ha for the generalized model (generalized model coefficients for all regions). Drake *et al.* [39] used a power function to link top-of-canopy height estimated from airborne LiDAR to aboveground biomass ($AGB = aTCH^b$). However, this method is considered plot-aggregate allometry rather than true allometry, as it reflects the whole-plot properties of forest structures in aggregate and not the properties of each particular tree. This method had an RMSE of 42.2 Mg/ha when tested in five tropical forests with different vegetation types. Lefsky *et al.* [44] linked the maximum canopy height (H_{max}) estimated from GLAS data to AGB using the following linear relationship: $AGB = a + bH^2_{max}$. Boudreau *et al.* [45] linked the GLAS waveform extent (difference between signal start and signal end), the slope (θ) between signal start and the first Gaussian canopy peak and the terrain index (TI) metric derived from the SRTM-DEM to AGB. Saatchi *et al.* [46] and Mitchard *et al.* [24] used Lorey's height (basal-area-weighted canopy height) instead of the maximum height for AGB estimation. In the different studies, it was found that Lorey's height is broadly related to canopy height [47]. However, Asner *et al.* [40] found that Lorey's height does not explain any variations in AGB, basal area, or wood density that cannot be explained by canopy height.

1.4 Forest canopy height in relation to forest biomass

One of the most important variables in the allometric relations which can be estimated from remote sensing techniques is the canopy height. Several allometries relied on only the canopy height for biomass estimation ([40]; [42]). In addition, studies have shown that the

use of canopy height increases significantly the precision of biomass estimation at tree level (e.g. [14]; [42]). In Chave *et al.* [42]; the use of tree height reduced relative error on the biomass from 19.5 (model using only DBH and wood density) to 12.8% (model using DBH, wood density and canopy height). In Feldpaush *et al.* [14]; biomass estimation models which used canopy heights, DBH and wood density showed a 50% decrease in the mean relative error in comparison to the models using only DBH and wood density. Other studies such as Asner *et al.* [48]; and Mitchard *et al.* [24] and Lefsky *et al.* [44] found that canopy height are strongly related to forest biomass. In addition to the importance of forest canopy heights in AGB estimation, knowing forest canopy heights is also interesting in itself for answering ecological questions such as on the determinants of plant and forest structure, forest dynamics, edaphic and climatic stress. Forest canopy heights are very important in forest management decisions; as changes in these heights may have direct effects on microclimatic patterns and processes ([49]). Indeed, the micro climate is modified first by local weather conditions, and then by vegetation, due to forest height which amongst the forest structure, controls the quality and quantity, spatial and temporal distribution of light. It also influences local precipitation and air movements. These factors combined will eventually determine to some extent the humidity in the air, temperature and soil moisture. In addition to having less direct effect on the behavior and distribution of various avian species ([50]; [51]; [52]). Moreover, forest height is important for managing resources such as wildlife, hydrologic response, aesthetics, tree growth and yield ([53]), fire hazard, and susceptibility to insects or disease.

1.4.1 Canopy height estimation using radar and optical data

Studies have used radar data to estimate canopy height using PolInSAR (polarimetric interferometric SAR) (e.g. [54]; [55]; [56]) and tomographic techniques (e.g. [57]; [58]). PolInSAR showed promising results for the estimation of canopy heights. In Neumann *et al.* [58]; canopy height estimation using PolInSAR showed an RMSE of 3 m with maximum canopy heights reaching 40 m when compared to reference canopy height estimates. Garestier *et al.* [57] estimated canopy heights using P-band PolInSAR data and found an RMSE on the canopy height estimation of 2 m for 2 to 25 m forest heights. However, it was hindered due to several sources of noise (weather changes, atmospheric heterogeneities, and intrinsic phase noise). SAR tomography is an alternative technique for using radar data in canopy height estimation. This technique is an imaging approach, which

generates a fully 3D representation of the imaged scene using coherent combination of a greater number of images ([59]; [60]). Huang *et al.* [59] used the tomography technique with P-band SAR data for canopy height estimation in a test site in French Guiana. Their results indicated an RMSE on canopy height estimates of 7.7 m. Mercer *et al.* [60] reported a 10% relative error on tree height estimates in comparison to LiDAR canopy height estimates using SAR tomography with L-band SAR data. SAR tomography is more robust against various noise sources in comparison to PolInSAR at the expense of the necessity to require many more flight lines. The BIOMASS Earth Explorer mission selected by ESA (European Space Agency) in the framework of its living planet program with a P-band spaceborne SAR satellite will provide strong opportunities for the estimation of both canopy heights and biomass from SAR images. Furthermore, many studies used medium and high resolution optical imagery such those available from MODIS, Landsat, Quickbird, IKONOS and others in order to extrapolate airborne or spaceborne LiDAR derived canopy height estimates (e.g. [47]; [61]).

1.4.2 Canopy height estimation using LiDAR data

To this date, canopy height estimation over large areas is best achieved using LiDAR data (either Airborne or Spaceborne). Lidar (Light Detection and Ranging) is an active remote sensing system well suited to measure specific forest information, including but not limited to: canopy heights, basal area, leaf area index, and canopy cover. LiDAR measures object elevation by sending a laser pulse, and measuring the pulse return time, and thus its distance from the LiDAR system and with the help of an onboard GPS, the system determines the objects elevation from the ground. Currently, Airborne LiDAR, is the most accurate remote sensing system to obtain specific site-level data on forest structure. However, wall-to-wall acquisitions of LiDAR data remain very expensive, therefore the use of spaceborne LiDAR systems, which produce free data globally becomes viable. Several studies have estimated canopy height using airborne or spaceborne LiDAR data (e.g. [24]; [44]; [45]; [46]). At regional and global scales, LiDAR data acquired by the Geoscience Laser Altimeter System (GLAS) have been widely used (e.g. [44]; [47]). Using GLAS data, maximum canopy height within each footprint has been successfully estimated with a precision between 2 and 13 m, depending on forest types and characteristics of the study site (e.g. [44]; [62]; [63]; [64]). Lefsky *et al.* [44] applied linear regressions on waveform metrics and ancillary DEM data for canopy height estimation and obtained site-specific models

with an RMSE between 4.85 and 12.66 m. Hilbert and Schmullius [62] when estimating canopy heights obtained an RMSE of 6.39 m on the canopy height estimation regarding all species and slope classes with a clear negative correlation between accuracy and slope. Lee *et al.* [63] applied a slope correction metric to a GLAS estimation model obtained high correlation between GLAS canopy height estimates and those estimated from a small footprint LiDAR with an RMSE of 2.2 m. Pang *et al.* [6] estimated the crown-area-weighted mean height with airborne LiDAR measurements using linear regression applied to metrics derived from GLAS waveforms. Their results indicated an RMSE of 3.8 m on the estimation of canopy heights in several coniferous forest sites in western North America.

1.4.3 Spatial extrapolation of LiDAR canopy height estimates

Finally, while LiDAR is very precise with canopy height estimates, it is limited in the horizontal domain (limited spatial coverage for airborne data and limited acquisition density for satellite data). Indeed, airborne LiDAR data are very expensive to acquire for very large areas (€135-175/km^2 with 1m point spacing), and while spaceborne LiDAR provides global coverage of waveform data they have a relatively low point density (about 0.51 points/km2 over French Guiana for example). Therefore, it is always necessary to merge LiDAR data (spaceborne or/and airborne) with optical or/and radar data, forest types data, geological data, meteorological data, etc. in order to create forest canopy heights with complete land coverage and a good precision(e.g. [47]; [65]; [66]).

Hudak *et al.* [66] tested one aspatial (linear regression), two spatial (kriging and co-kriging) and two combined spatial and aspatial methods (kriging and cokriging of regression residuals) for mapping canopy heights using airborne LiDAR canopy height estimates and Landsad Enhanced Thematic Mapper (ETM+) using several sampling strategies (250, 500, 1000 and 2000 m) in a 200 km² study site in western Oregon (USA). Their results showed that the regression model maintained vegetation pattern, however it was more biased towards taller and shorter trees (underestimating taller canopy heights while overestimating shorter ones). Using the regression model, the standard deviation on the canopy height residuals (reference canopy heights – estimated canopy heights) was in the order of 10 m regardless of the sampling strategy. The direct kriging or co-kriging of canopy heights were only slightly more precise than the regression model when predicting canopy heights at

locations lower than 200 m from the reference canopy heights. Moreover, the co-kriging method proved to be slightly more precise than the kriging method. Finally, the method which combined the regression and the kriging and co-kriging of the residuals proved to be the best method for mapping canopy heights. This method conserved the pattern of the canopy heights and improved the precision on the canopy height estimates. The standard deviation on the canopy height estimates varied between 5.5 and 10.9 m for respectively a sampling pattern of 250 and 2000 m.

Lefsky *et al.* [47] created a global forest canopy height map using regression analysis of canopy heights estimated from the GLAS data and 500 m Moderate Resolution Imaging Spectroradiometer (MODIS) data. The linear regression model which was used to model MODIS data to the GLAS canopy height estimates in order to map forest canopy heights globally showed canopy height estimates with a root mean square error on the estimation of canopy heights of 5.9 m and a coefficient of correlation (R2) of 0.67.

Finally, a more recent study conducted by Simard *et al.* [65] improved on the work of Lefsky *et al.* [47] for global canopy height mapping by replacing the linear regression model with the Random Forest technique and 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 the percentage tree cover provided from MODIS. Their global canopy height map which was validated against in-situ measurements showed moderate canopy height estimation precision with an RMSE of 6.1 (R^2 of 0.5) on the estimation of canopy heights.

1.5 Forest types in relation to forest biomass

In addition to the role of forest canopy heights in AGB estimation, forest landscape classification also plays a major role in the methods for estimating AGB. Indeed, many studies have found that AGB estimation models are more relevant when including forest types ([24]; [42]; [67]; [68]; [69]). Zheng *et al.* [67] found that the coupling of tree metrics acquired from field measurements and various indices derived from Landsat 7 ETM± substantially improved AGB estimates when separating hardwood from pine forests. Chave *et al.* [42] tested several models for AGB estimation in old growth, dry, moist, wet,

montane and mangrove forests. Their results indicated that one of the most important factors for AGB estimation is forest type. The results also indicated that the best predictive models were forest-type dependent. Ni-Meister *et al.* [68] developed an AGB estimation model that uses a fusion of LiDAR and optical sensors (to provide the vegetation type) in conifer/softwood and deciduous/hardwood forests. Their results indicated that vegetation-type-dependent models provide better AGB estimates in comparison to vegetation-type-independent models. Mitchard *et al.* [24] found a ±25% uncertainty in the estimation of AGB in Lope National Park (Gabon) using LiDAR data and a vegetation structures map extracted from radar images. Finally, Addo-Fordjour [69] developed AGB estimation models for different species of lianas. Their results indicated that forest type has a significant influence on the allometric relationships used in AGB estimation, which led to forest-type-specific equations.

1.6 Organization of the dissertation

1.6.1 Objectives

The main objective of this thesis was to create new methodologies for the mapping of large forested areas, often inaccessible using remote sensing techniques and especially the one that uses LiDAR. LiDAR remote sensing is an attractive and a complementary technique used with other remote sensing techniques for mapping forest biomass notably through the characterization of the height and the vertical structure of the canopy. However, current missions (satellite or airborne), do not allow the acquisition of LiDAR data with sufficient spatial density measurements for accurate mapping of tree height and subsequently the estimation of biomass at a regional scales. The challenge was then to develop methods for spatial estimation of vegetation height from airborne and satellite LiDAR data and other data sources. The goal is to produce a wall-to-wall canopy height map of French Guiana.
From this objective stems different sub-objectives:
- develop a procedure for the estimation of canopy heights from mono-echo airborne LiDAR datasets.
- Evaluate the potential of the LiDAR system ICESat/GLAS to estimate canopy heights in a tropical forest by developing different statistical methods that uses waveforms provided by the ICESat/GLAS system.

- Improve canopy height estimation precision by choosing complementary data issued from different technologies.
- Develop data fusion methodologies using LiDAR canopy height estimates with ancillary data (geological, meteorological phenological ...) in order to propose a forest wall-to-wall canopy height map with good precision and high spatial resolution.
- Analyze the relationship between canopy height estimation precision and the spatial sampling of LiDAR data.
- Evaluate the potential of the ICESat/GLAS data and data from the shuttle radar topography mission (SRTM) for the classification of forest landscape types and forest types.

1.6.2 Dissertation plan

The dissertation contains in total six chapters, including the introduction (chapter 1), chapter 2 which presents the study area and the datasets used, the sub-objectives mentioned in section 1.6.1 are represented in chapters 3, 4, and 5, and, finally the conclusion and perspectives (chapter 6) and the summary of the thesis in French (Chapter 7). In addition, the chapters were written based on scientific articles which were published or submitted at the time of this writing. Each article is introduced later-on in its respective chapter.

Chapter 2 introduces the study area, alongside all the used data in our study. Chapter 3 which is based on a published article in a the peer-reviewed international journal “Remote Sensing” will be dedicated to the introduction of the LiDAR technology (airborne and spaceborne), as well as detailing the methods and procedures used in this thesis in order to estimate canopy heights with either airborne- or spaceborne LiDAR. This chapter starts off with an introduction of the LiDAR datasets used. Next, a detailed description of the methods used in order to estimate forest canopy heights using airborne LiDAR as well as a validation of these estimates is shown. Following that, an introduction of spaceborne LiDAR is presented, as well as the processing of waveforms provided by the ICESat/GLAS and the extraction of the most useful metrics used for canopy height estimation. The remaining of the chapter will be dedicated to the presentation of canopy height estimation methods using ICESat/GLAS with a validation of each method.

Chapter 4 focuses on using the LiDAR based canopy height estimates obtained in chapter 3 as well as ancillary data in order to produce a validated wall-to-wall canopy height map of the entire French Guiana. In addition, the effect of spatial sampling of the LiDAR datasets on the canopy height estimation precision was also studied in this chapter. The contents of this chapter constitute an article that has already been submitted to the remote sensing (MDPI) journal.

Chapter 5 is also based on a published article in the peer-reviewed “International Journal of Applied Earth Observation and Geoinformation”. In this chapter, first, comparisons between elevations extracted from ICESat/GLAS waveforms and elevations from SRTM data were used in order to classify French Guiana’s forest landscape classes. Next, several metrics were extracted from GLAS waveforms in order to classify forest types. Finally, chapter 6 presents the main conclusions and the perspectives of this thesis

CHAPTER 2:

STUDY AREA AND DATASETS

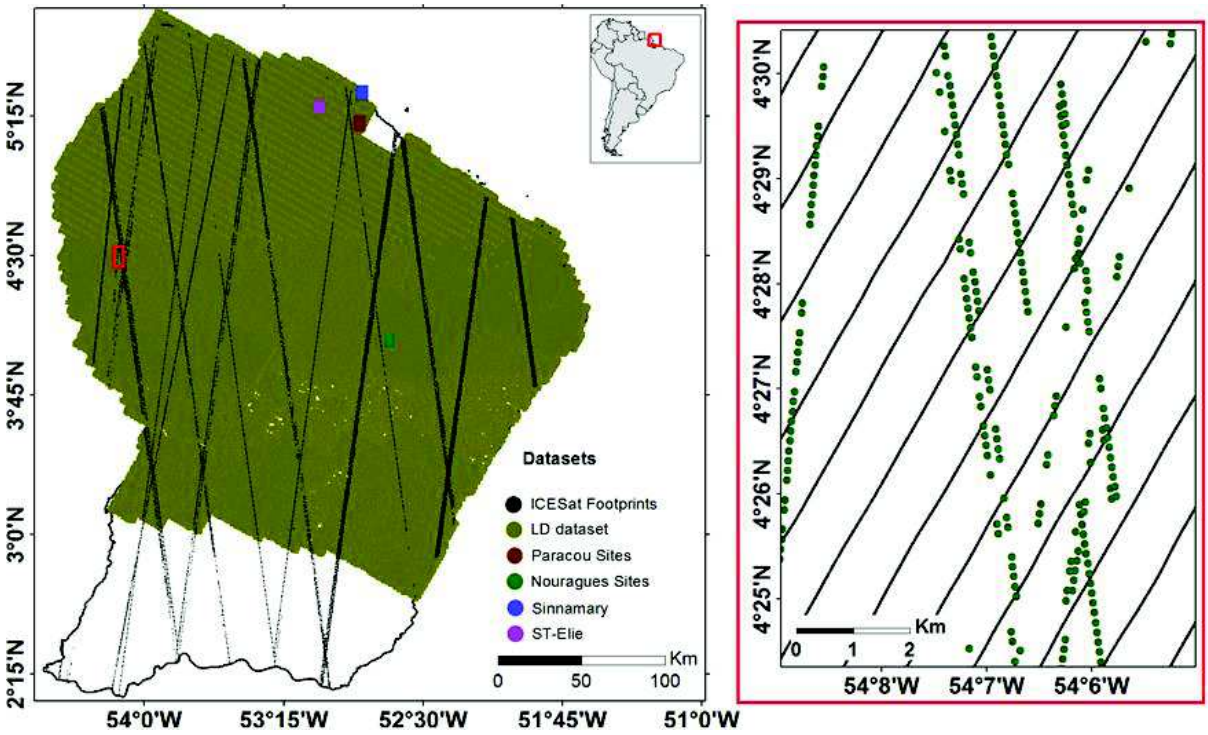
In this study, different datasets were used over our study area for the estimation and mapping of canopy heights. These datasets are comprised of LiDAR data and data from different auxiliary datasets. In order to use the different datasets, filtering, and processing is required. In this chapter, the study is first presented. Next, all the datasets used in this study are presented, as well as any required filtering and processing.

2.1 Study area

French Guiana is situated on the northern coast of the South American continent, bordering the Atlantic Ocean as well as Brazil and Suriname (Figure 2.1). The study site's area is approximately 83,534 km^2, and forest occupies approximately 80,820 km^2 or approximately 96.75% of its total size. The terrain is mostly low-lying, rising occasionally to small hills and mountains, with an altitude ranging from 0 to 851 m. In addition, 67.8% of its slopes are lower than five degrees, 24.0% are between five and ten degrees and 8.2% are higher than ten degrees (derived from the SRTM elevations). Dense tropical forests predominate outside the coastal plain and cover more than four-fifths of the land area. Other vegetation types also exist, such as savannas and agricultural crops. French Guiana

has an equatorial climate with two main seasons, the dry season, from August to December, and the rainy or wet season, from December to June.

Figure 2.1. LiDAR datasets acquired for French Guiana (the right image corresponds to the red rectangle in the left image). The small rectangles represemt the location of the HD LiDAR dataset



2.2 Datasets description

2.2.1 Spaceborne LiDAR datasets

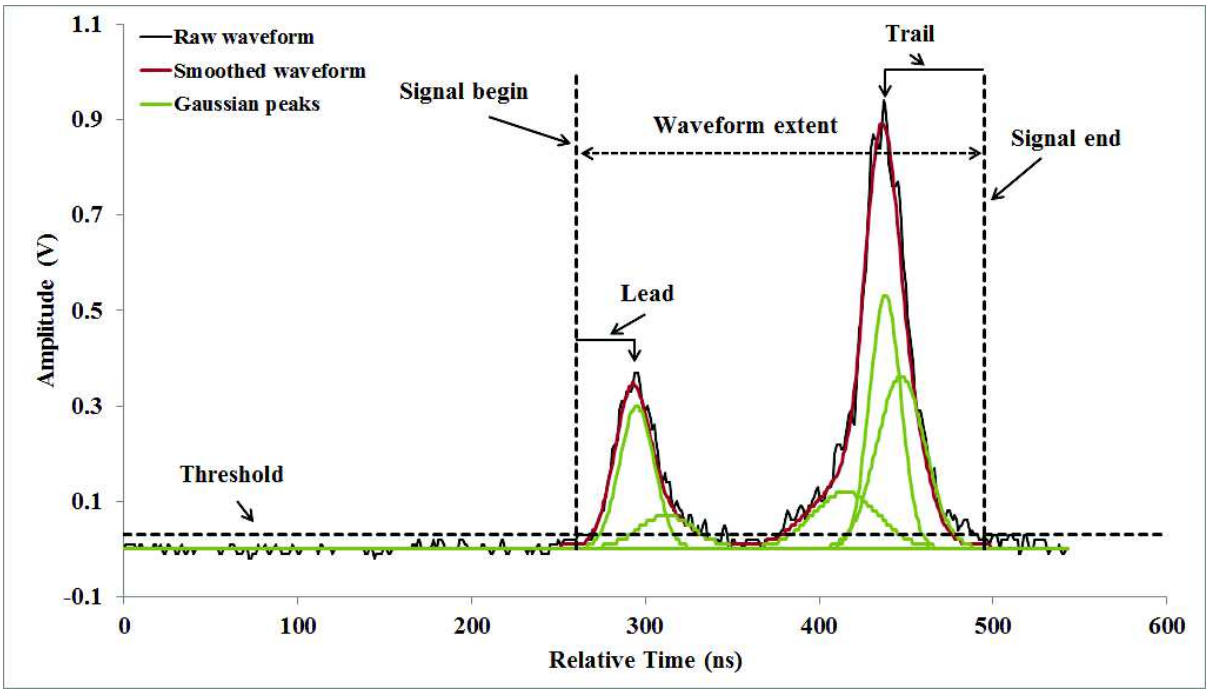
LiDAR data were acquired from the GLAS on board the Ice, Cloud, and Land Elevation Satellite (ICESat) between 2003 and 2009. The GLAS laser footprints have a nearly circular shape of approximately 80 m in diameter and a footprint spacing of approximately 170 m along their track. The data were acquired during 18 missions using three on-board lasers with orbit cycles repeating between 57 and 197 days. Over French Guiana, GLAS data acquisition time coincides with the wet (GLAS acquisition in Feb-March and May-June) and dry (GLAS acquisition in October-November) seasons.

The horizontal geolocation error of the ground footprints is less than 5 m, on average, for all ICESat missions (http://nsidc.org/data/icesat/laser_op_periods.html). Several studies

(e.g. [59]; [70]) have estimated the vertical accuracy of the GLAS to be between 0 and 3.2 cm over flat surfaces, on average.

From the 15 data products available from the ICESat GLAS, the GLA01 and GLA14 data products were used in this study. The GLA01 comprises the full waveform data, and GLA14 comprises the global land surface altimetry data. Over flat terrain, the waveforms acquired over vegetated areas are bimodal distributions, with the first peak representing reflections from the canopy top and the last peak representing the ground (Figure 2.2). To exclude unreliable GLAS data (i.e., data affected by atmospheric conditions, clouds, etc.), several filters were applied. (1) Signals with high noise were removed when the signal to noise ratio was higher than 15 (e.g. [70]; [71]; and [63]). This filter removed 36.4% of the data. (2) The GLAS waveforms with delays from either saturation or atmospheric forward scattering were removed (14.1% of the data). Only cloudless waveforms were kept using the cloud detection flag (FRir_qaFlag = 15). This filter removed 32.4% of the data. Saturated signals were identified using the GLAS flag (SatNdx > 0). (3) The waveforms with a centroid elevation significantly higher or lower than the corresponding SRTM elevation were removed (|SRTM - GLAS| > 100 m) ([72]). This filter removed 2% of the data. (4) The GLAS footprints with SRTM values higher than the GLAS canopy top elevation and lower than the GLAS ground elevation were also removed, which accounted for 33.4% of the data. Both the FRir_qaFlag and SatNdx flags were found in the GLA14 product. From the original database of 101312 footprints, 12238 footprints that satisfied the 4 filters conditions were kept (Figure 2.1). Finally, the GLAS data referenced to the TOPEX/Poseidon were converted to WGS84 by subtracting 70 cm from the elevation values. The conversion between the two ellipsoids also depends on latitude; however, as this change is smaller than the horizontal accuracy of the GLAS, it was omitted.

Figure 2.2 a typical GLAS waveform acquired over a vegetated area on a flat terrain. Airborne LiDAR datasets



2.2.2 Airborne LiDAR datasets

2.2.2.1 Small footprint low density LiDAR dataset

A LiDAR dataset was acquired in 1996 during an airborne geophysical survey that covered 4/5 of French Guiana (northern part, Figure 2.1). Because laser data were acquired for assessing the quality of the survey, and particularly for flight ground clearance, a low sampling frequency was used, and only the first pulse was considered [73]. The data correspond to the elevation of the first obstacle encountered by the laser. The sampling frequency was 10 Hz with a 905-nm wavelength laser and a footprint size of 35 cm (laser beam divergence of approximately 3 mrad). The laser measurements are therefore considered point data. The database contains laser elevations every 7 m on flight lines spaced 500 m apart and oriented at 30°N, intersected by transverse flight lines spaced 5 km apart and oriented at 120°N. The mean density of this database is approximately 285.2 points/km^2. Bourgine *et al.* [74] evaluated the quality of this low-density LiDAR dataset (LD), and the accuracy of the terrain elevation was estimated to be approximately ±2 m.

2.2.2.2 Small footprint high density LiDAR dataset

LiDAR datasets with high points density (HD) acquired during several airborne surveys in 2004, 2007, 2008 and 2009 as part of the Guyafor project by a private contractor, Altoa (http://www.altoa.fr/), operating a helicopter-borne LiDAR were used in this study (Table 2.1). These data were previously made available to the ESA Tropisar project. The Biomass project at Jet Propulsion Laboratory (JPL) used this dataset for the evaluation of forest structure estimation from radar data. The elevations were recorded using two LiDAR systems: Riegl LMS-Q140i-60 in 2004, 2007 and 2008 and the newer LMS-280i system in 2009. The elevation data were acquired for several small study sites in French Guiana (Figure 2.1). The mean acquisition density of the HD datasets is 3.5 points/m^2 (between 3 and 7 points/m^2). The laser wavelength was 905 nm with a mean footprint size of 45 cm for the first system and 10 cm for the second, and the precision of the elevation was smaller than 0.1 m. Moreover, the HD, unlike the LD, is a last-return laser elevation measurement, as using the last return increased the percentage of ground returns [75].

Table 2.1. Description of the HD datasets used in this study.

Site	Acquisition Date	Location	Area (km^2)	PointDensity (points/m^2)
Paracou_2004	2004	5°15.9′N 52°55.9′W	5.35	4.0
Sinnamary	2004	5°24.7′N 52°56′W	6.52	7.0
St-Elie	2009	5°18.2′N 53°3.3′W	4.50	5.3
Nouragues07A	2007	4°5.3′N 52°40.7′W	7.24	3.2
Nouragues07B	2007	4°2.4′N 52°40.6′W	2.42	3.8
Nouragues08A	2008	4°5.1′N 52°41.2′W	1.96	4.5
Nouragues08B	2008	4°3.8′N 52°40.9′W	7.82	3.8
Nouragues08C	2008	4°2.5′N 52°40′W	2.89	4.2
Nouragues08D	2008	4°2.5′N 52°41.0′W	1.08	3.5
Paracou_2009	2009	5°16.1′N 52°55.8′W	12.08	6.0

2.2.3 Ancillary datasets

In this study, twelve environmental and geographical variable maps were used. These variables were chosen for their supposed influence on forest characteristics. In addition, these variables are accessible from available free maps (Table 2.2.) The environmental variables will be used later in regression models in order to estimate canopy heights over the entire French Guiana. These variables are: geological map, forest landscape type map, three variable maps computed from SRTM digital elevation model (at 90 m resolution), six variable maps derived from vegetation indices issued from MODIS optical images, and finally one variable map issued from rainfall data.

Table 2.2. Description of the variable maps used for canopy height mapping.

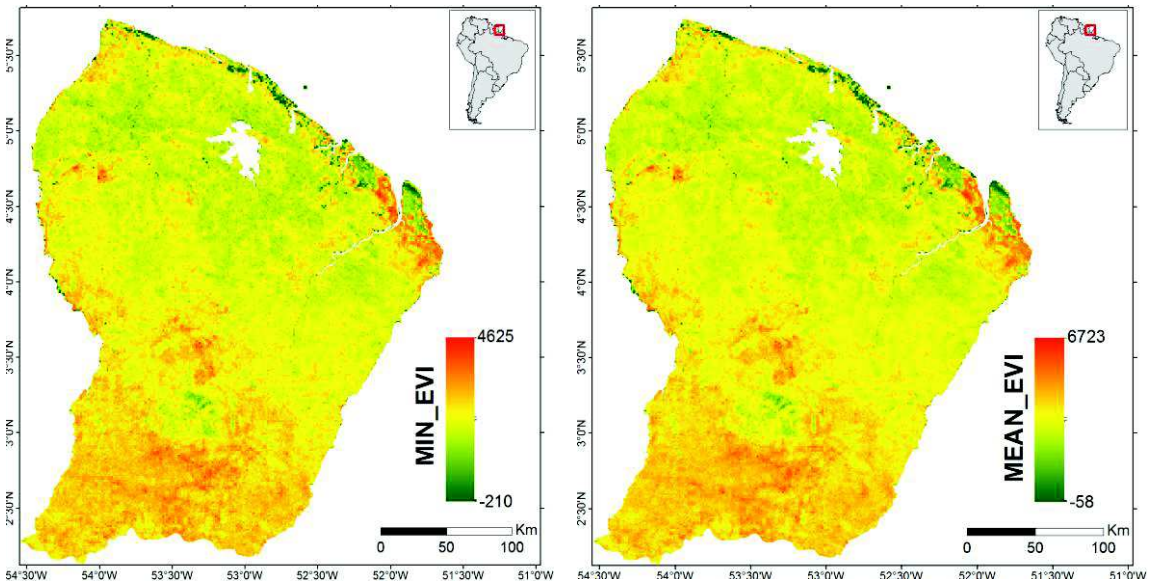
Short name	Full name	Source	Resolution
MIN_EVI	Minimum value of EVI time series data	MODIS	250 m
MEAN_EVI	Mean value of EVI time series data		
MAX_EVI	Maximum value of EVI time series data		
PC1	1st principal component of EVI time series data		
PC2	2nd principal component of EVI time series data		
PC3	3rd principal component of EVI time series data		
Slope	Terrain slope	SRTM	90 m
Roughness	Terrain roughness		
ln_drain	Log of drainage surface		
GEOL	Geological map	Delor *et al.* [76]	Vector
LTs	Forest landscape type	Gond *et al.* [77]	1 km (Vector)
Rain	mean value of rainfall	TRMM	8 km

2.2.3.1 MODerate-resolution Imaging Spectroradiometer (MODIS) data

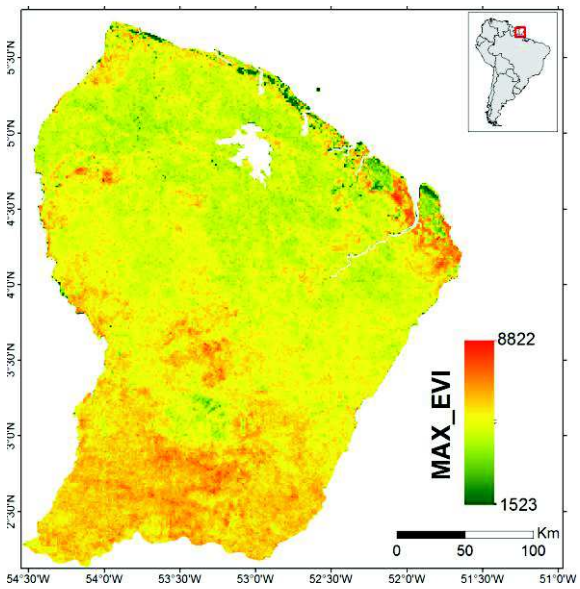
MODIS sensor mounted on the Terra and Aqua satellites possesses a total of 36 spectral bands of which seven designed specifically for land applications with spatial resolutions that range from 250 m to 1 km. The MODIS dataset used in this study includes ten years

(January 1st, 2003 to December 31st, 2012) of the enhanced vegetation index (EVI) time series data. EVI data effectively characterize biophysical and biochemical states and processes from vegetated surfaces. The 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 2.3), and the three first principal components issued from the principal component analysis of the EVI time series data (PC1, PC2 and PC3).

Figure 2.3. Minimum (a), mean (b), and maximum (c) values of the EVI time series data.



(a) (b)



(c)

2.2.3.2 SRTM digital elevation model data

The Shuttle Radar Topography Mission (SRTM) acquired a digital elevation model of the earth's surface on a nearly global scale (50°S to 60°N). The vertical accuracy of the SRTM 90 m DEM is 16 m with a 20 m horizontal accuracy (http://www2.jpl.nasa.gov/srtm/datafinaldescriptions.html). In this study, the SRTM 90 m DEM currently available for French Guiana was used. Over French Guiana, Bourgine and Baghdadi [73] found that the accuracy of the SRTM DEM was approximately 10 m (standard deviation of error). The SRTM data are available as orthometric heights, with WGS84 as the horizontal datum and the Earth Gravitational Model (EGM96) geoid as the vertical datum. To compare the ICESat/GLAS and SRTM elevations, the SRTM geoidal heights were converted to ellipsoidal heights by adding the EGM96 geoidal undulations. The geoidal undulations are available on a 0.1x0.1-degree grid interpolated onto the pixel coordinates. The SRTM dataset was interpolated onto each ICESat/GLAS footprint using bilinear interpolation.

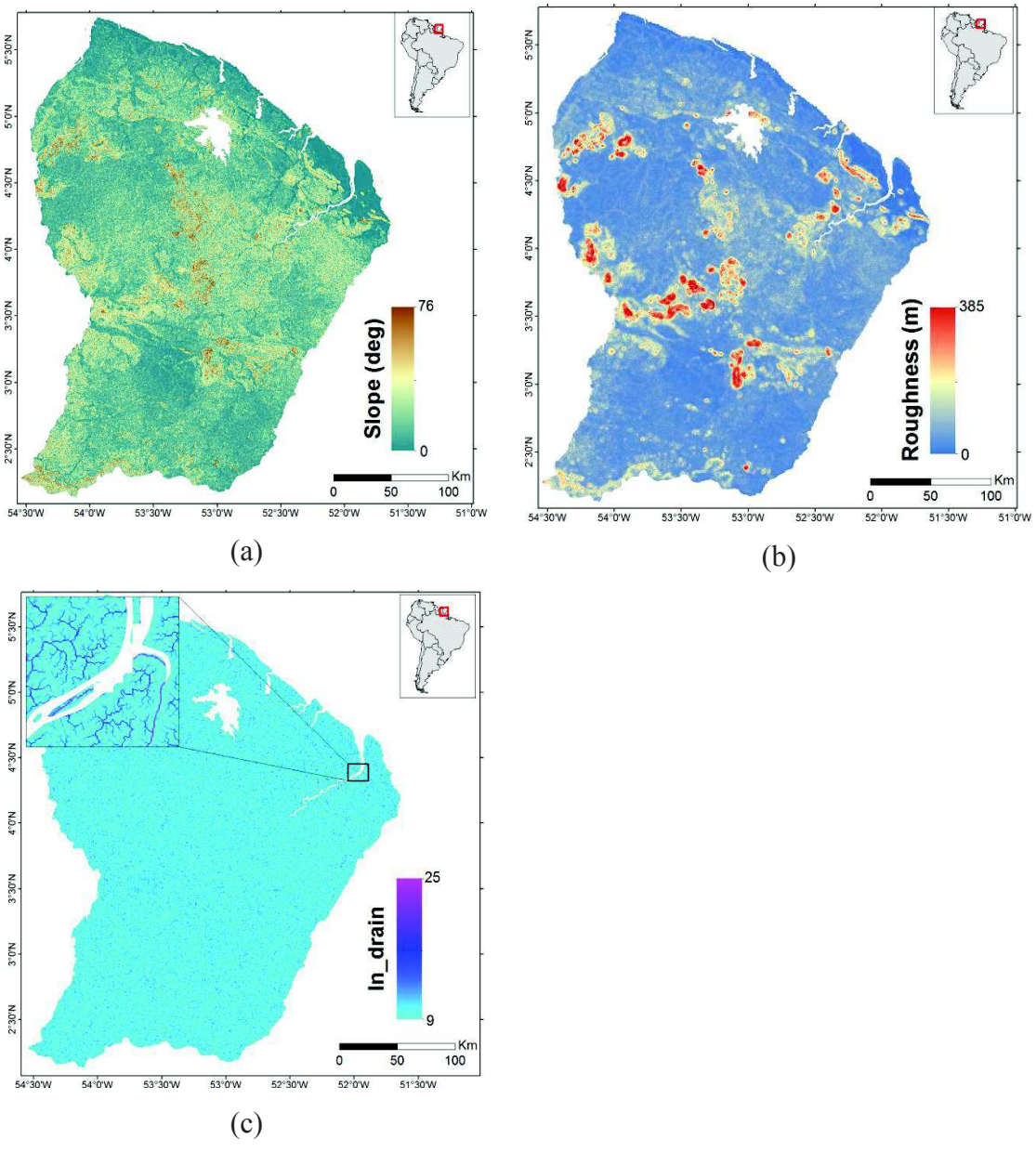
From the SRTM DEM data, three derivative maps were created (Figure 2.4):

(1) A slope map (Slope), which is calculated using the maximum change in elevation over the distance between each cell of the DEM and its eight neighbors in a 3x3 window (Slope) (Figure 2.4a);

(2) A surface roughness map (Roughness), where the roughness is the standard deviation of a 3x3 moving window. Areas with low standard deviation represent areas of low roughness, while higher standard deviation presents higher surface roughness (Figure 2.4b);

(3) Finally a drainage surface map (ln_drain), where the drained area measures the surface of the hydraulic basin that flows through a cell. A low value indicates cells located at the border between two hydraulic basins, whereas the highest values indicate cells located downstream (Figure 2.4c).

Figure 2.4. SRTM DEM derived maps: slope map "in degrees" (a), surface roughness map "in m" (b), and drainage surface map (c).



2.2.3.3 Geological map

A geological substratum map (GEOL) produced by the French Geological Survey ([76]) was used in this study (Figure 2.5a). 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 geology class to be sampled with a satisfying accuracy.

2.2.3.4 Forest landscape types map

A forest landscape types map developed by Gond *et al.* [77] at 1 km resolution was also used (Figure 2.5b). In this map, 33 remotely sensed landscape types (LTs) using VEGETATION/SPOT 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 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.

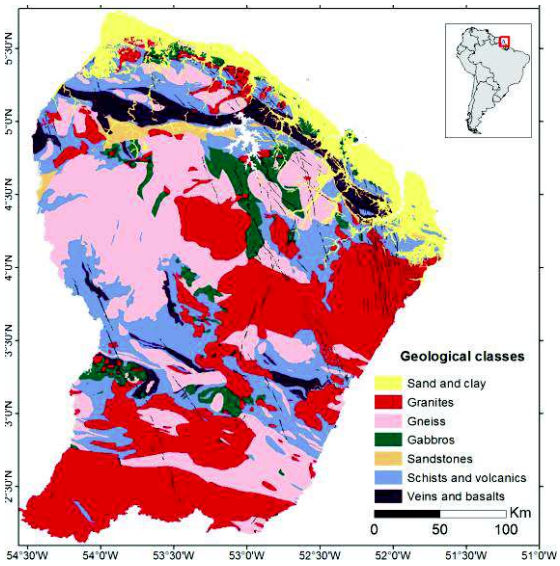
(4) LT11 is similar to LT10 with more liana forest and non-forest land covers.

(5) LT12 is an open forest associated with wetlands and bamboo thickets.

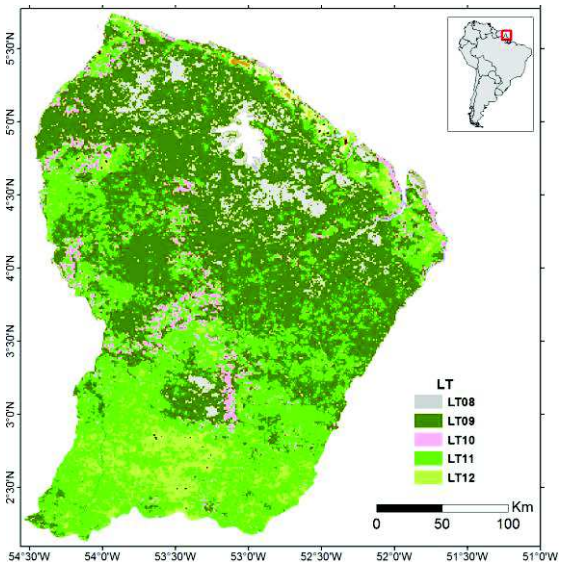
2.2.3.5 Average rainfall map

In addition, 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 used in this study are is the average daily precipitation over the last 10 years (2003-2013) with a resolution of ~25 km (Rain) (Figure 2.5c).

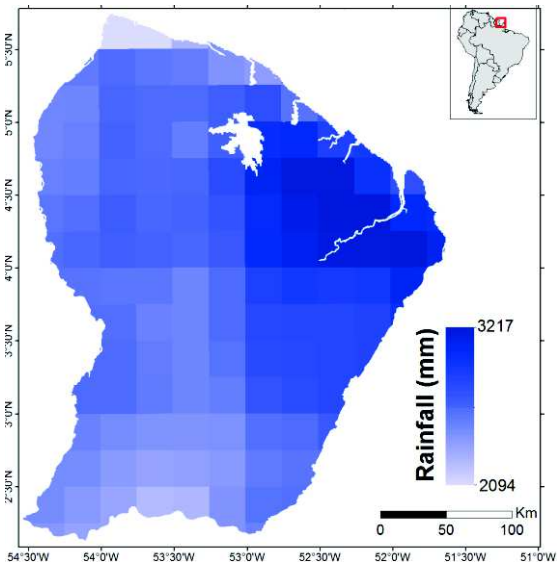
Figure 2.5. Geological map (a), Forest landscape types map (b), and Average rainfall map (c).



(a)



(b)



(c)

CHAPTER 3:

CANOPY HEIGHT ESTIMATION IN FRENCH GUIANA WITH LIDAR ICESAT/GLAS DATA USING PRINCIPAL COMPONENT ANALYSIS AND RANDOM FOREST REGRESSIONS

3.1 Introduction

Canopy height estimation models based on full waveform data can be divided into two categories: the direct method and statistical models. The direct method enables canopy height estimation in low relief areas using the difference in elevation between signal start and the ground. However, over sloping areas, the direct method overestimates canopy heights because of the additional height introduced by the slope. To remove the effects of the slope, statistical models using GLAS and DEM metrics have been developed. Nevertheless, while the metrics developed in previous studies were very successful in increasing the precision of the canopy height estimation models (e.g. [44]; [71]; [78]), they presented their own shortcomings. Indeed, in order to use these metrics for better canopy height estimation, the exact position of the top-of-canopy and ground peaks is often required. Over dense vegetated areas such tropical forests, extracting the top-of-canopy and ground peaks is especially difficult using an automated process, as the LiDAR waveform does not often present distinctive peaks [24]. The extraction of these metrics

manually is always possible, but becomes time consuming and inefficient when dealing with a large number of GLAS waveforms.

The goals of this chapter are to test several commonly used canopy height estimation models that utilize metrics derived from GLAS waveforms and SRTM-DEM and to test two techniques, new in the field of forest applied LiDAR: principal component analysis (PCA) and Random Forest. The purpose of using the PCA approach is to eliminate the need for metrics extracted from GLAS in canopy height estimation models, as the extraction of these metrics is error-prone, especially in dense forests, such as those in French Guiana. For the Random Forest regressions, the same metrics derived from GLAS footprints will first be used. Then, the principal components from the PCA of the GLAS waveform will be tested. The results of each model will be validated against canopy height estimates obtained from an airborne LiDAR dataset.

Section 3.2 is the presentation of methods for forest height estimation using airborne and satellite LiDAR. The results are shown in in Section 3.3. Finally, Sections 3.4 present the discussion and section 3.5 represents the conclusions. The chapter is based on the published article: **Fayad. I *et al.* (2014) Canopy Height Estimation in French Guiana with LiDAR ICESat/GLAS Data Using Principal Component Analysis and Random Forest Regressions. Remote Sensing 6:11883-11914**.

3.2 Materials and methods

3.2.1 Lidar data processing and canopy height estimation

3.2.1.1 Processing the LD dataset

To estimate canopy heights using the airborne low-density LiDAR data, several steps were required. First, the dataset was filtered to remove erroneous elevation measurements. Next, the canopy-top and the ground points were extracted to estimate the canopy heights. The process for canopy height estimation is summarized in the following sections.

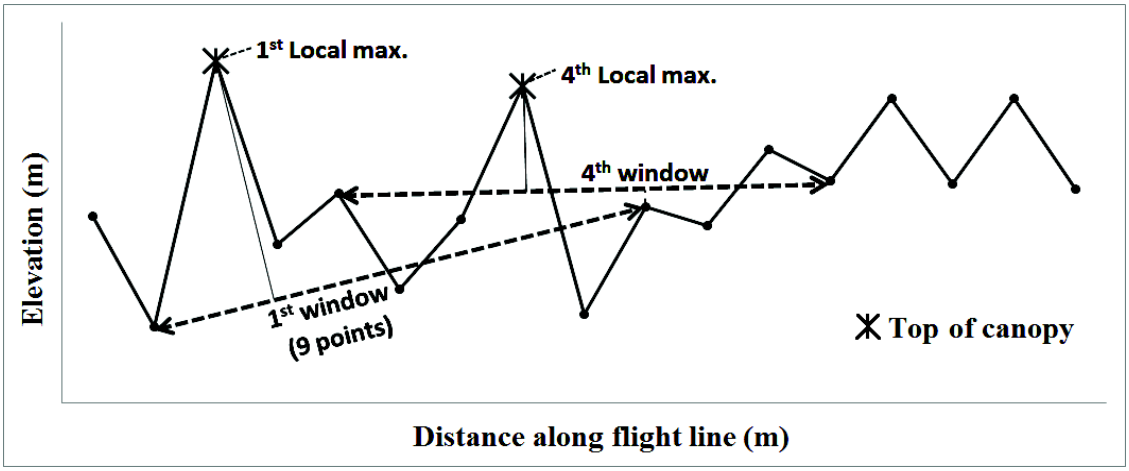
– ***Data filtering***

Airborne LD LiDAR data showed local-scale fluctuations according to whether the point corresponded to a treetop, a branch at intermediate level, or even a stream or the ground. The analysis of LiDAR data showed important differences due to measurement errors in LiDAR elevations (Z) between two neighboring points (a distance of 7 m in the LD dataset). Elevation differences up to 150 m were observed. LiDAR points with a difference in Z greater than 60 m were discarded (less than 3% of the total dataset was removed). This threshold of 60 m was chosen considering the extreme case in which one laser point represents the top of a tree and its neighboring point reaches the ground, giving an approximate maximum canopy height of 60 m.

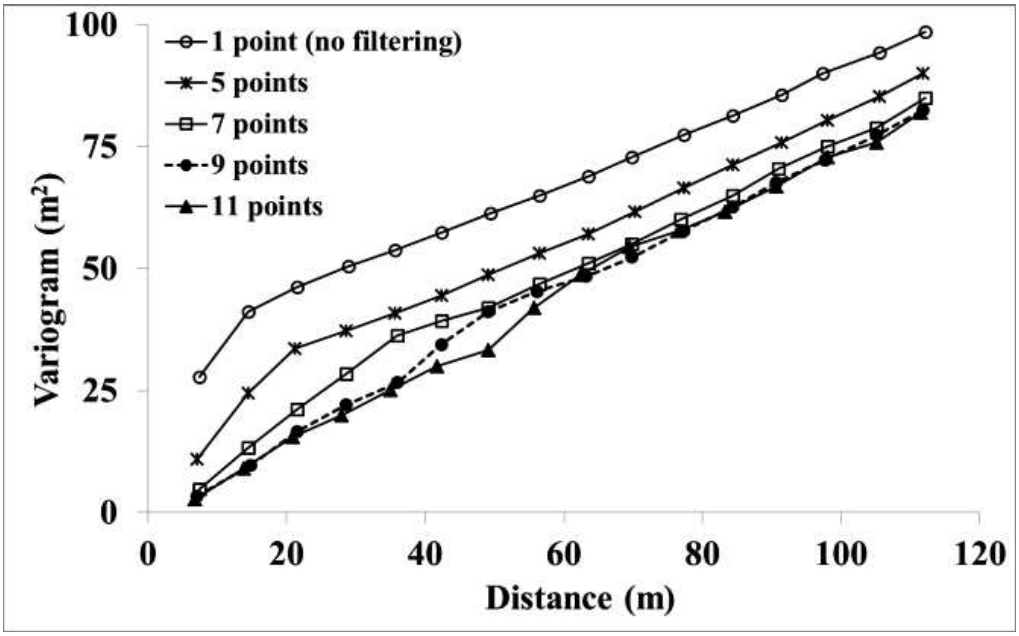
– ***Canopy top identification***

Next, airborne LiDAR data were filtered to select the points that most likely corresponded to canopy tops. This was achieved by selecting the local maximum in a sliding window of n points (n being odd numbers). In each window, the local maximum was selected as the point with the maximum amplitude with respect to the line segment joining the boundaries of each window (Figure 3.1a). The window size was selected so that the variogram of LiDAR elevations (Z) no longer displayed an apparent nugget effect. Figure 3.1b shows that the nugget effect disappears when windows are larger than seven points and that a window of nine points (*i.e.*, 56 m) gives a nearly linear variogram. Windows of a larger size did not improve the results and tended to decrease the number of available points. With a nine-point window, more than a quarter of the filtered LD LiDAR points were conserved (a point every 42 m, on average, along the flight lines), making a total of 3,289,076 top-of-canopy points available over French Guiana (49.21 pts/km^2).

Figure 3.1. (a) Points selected as top of canopy (local maximum); (b) Variogram of airborne LiDAR elevations from the LD dataset with local maximum points as a function of the size of the filtering window; (c) Canopy height calculation; (d) Ground points selected from a 1000-m window.

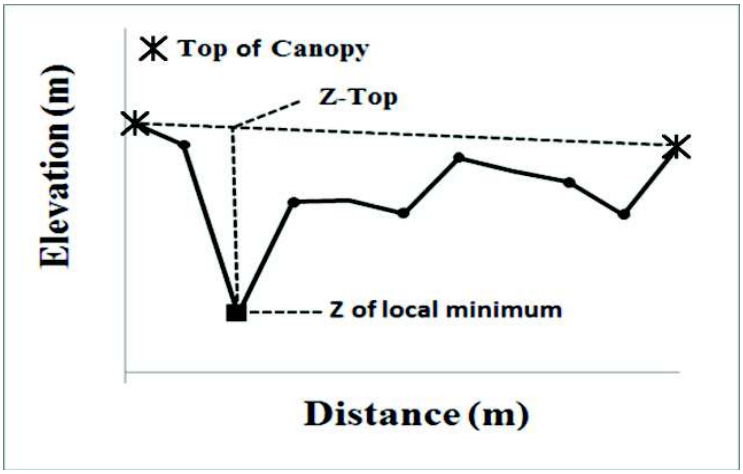


(a)

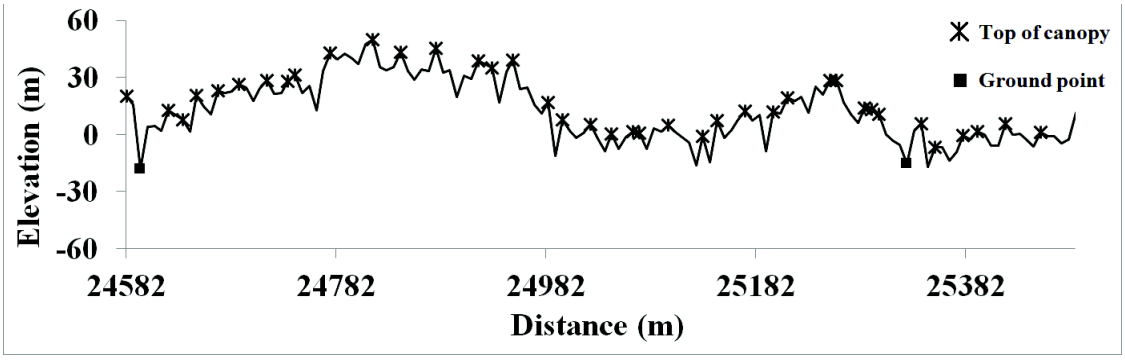


(b)

Figure 3.1 (cont.)



(c)



(d)

– ***Identification of ground points***

Few LiDAR returns reach the ground in tropical forests. Vincent *et al.* [75] estimated that, in last-return mode, only 1% of all laser returns are ground measurements. Bourgine *et al.* [74] estimated the ground returns in the LD dataset to be several hundred meters apart. To select the ground points from the LD dataset, the following procedure was attempted (e.g. [74]; [79]):

(1)Between two successive points identified as top of canopy, identify the local minimum, i.e., the point that gives the maximum canopy height (Figure 3.1c). For all points situated between the two top-of-canopy points, the canopy height is calculated as the difference between the elevation of each point (Z) and the top-of-canopy elevation (ZTOP). ZTOP is obtained using a linear interpolation between the elevations of two canopy tops.

(2)Among the local minimum points selected in the previous step, retain the lowest one inside a non-overlapping moving window (point corresponding to the greatest canopy height) (Figure 3.1d). With the use of a small window size, the selected ground points are

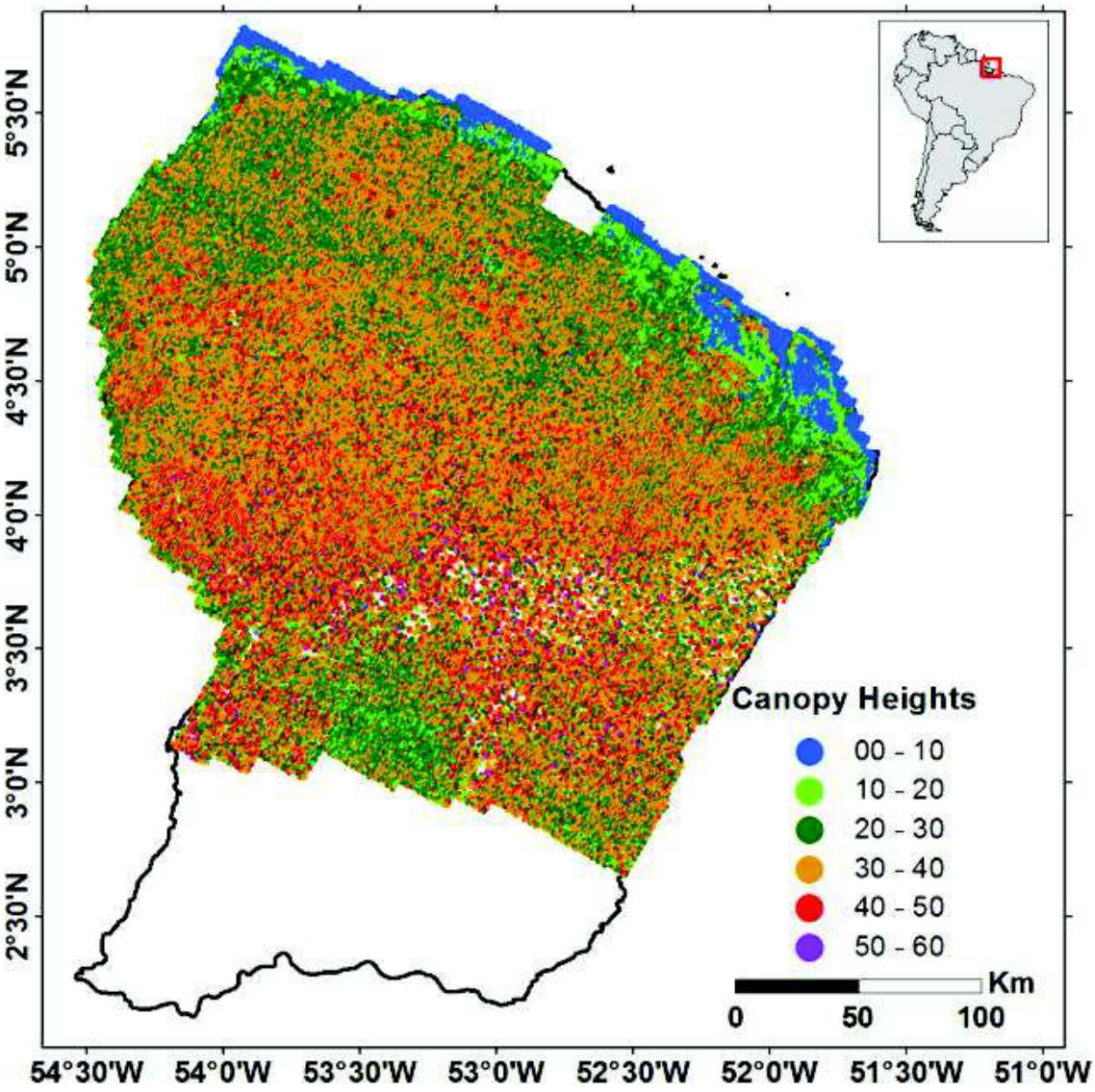
often located above the ground, leading to an underestimation of the canopy height. For too-large windows, too many ground points are eliminated, leading to an excessive smoothing of the estimated canopy height during the subsequent interpolation. Bourgine *et al.* [74] demonstrated that the best window size for this LD dataset is 1000 m. The number of ground points available for French Guiana is 105438 (1.59 pts/km^2).

– ***Canopy height estimation***

Canopy heights were calculated for the LD dataset using points identified as top of canopy and ground (Figure 3.1d). The estimation of the canopy height was performed at the level of the 105,438 ground points using linear interpolation between the elevations of the top-of-canopy points (spaced 42 m apart, on average). Canopy height estimation cannot be conducted at the canopy-top level by interpolating the ground points because the distance between ground points (1000 m, on average) is too great to assume a linear trend between the elevations of ground points.

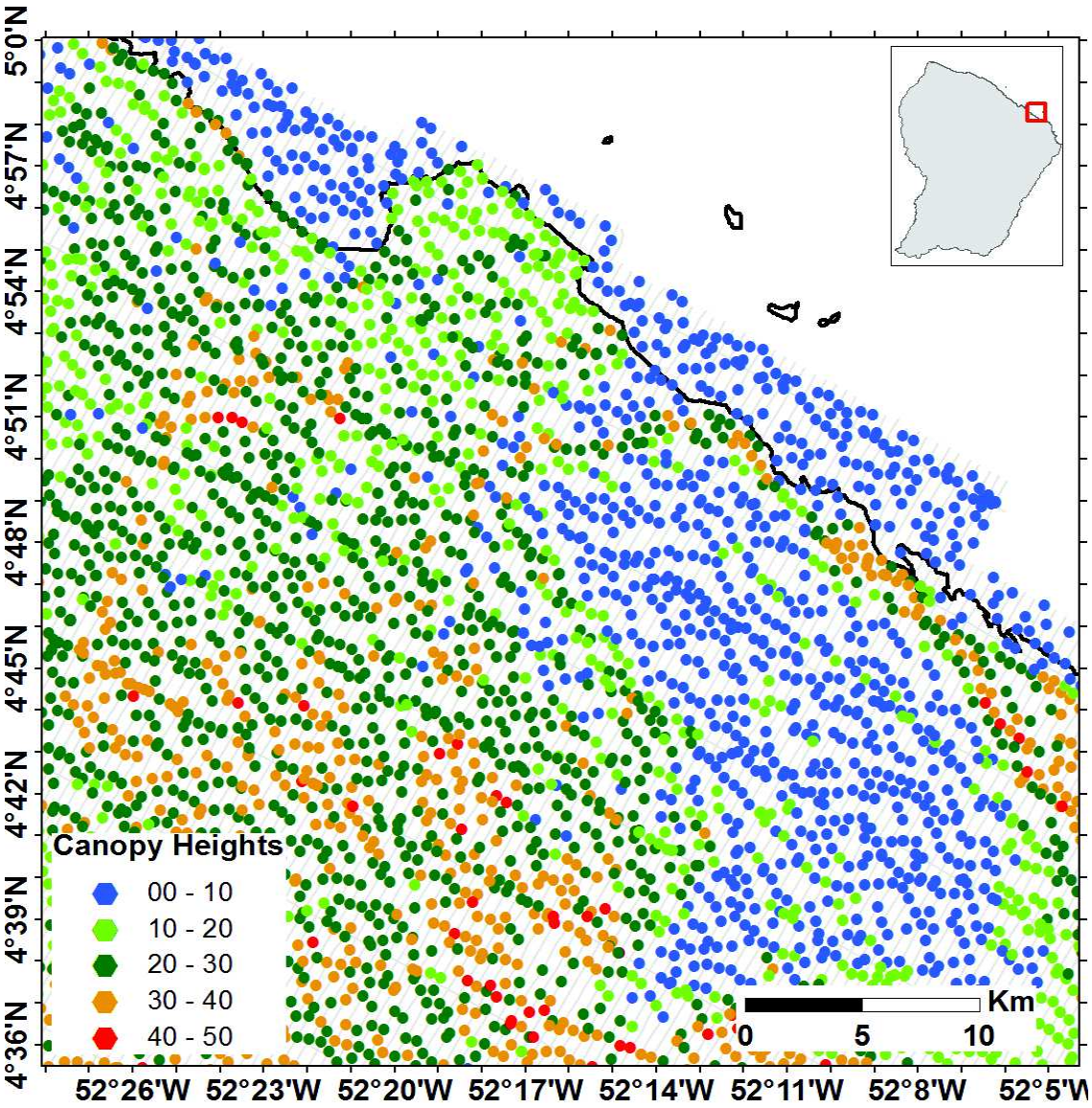
The estimation of canopy height using the LD dataset showed that canopy heights reached a maximum of 69 m with a mean height of approximately 30.4 m. The lower canopy heights (maximum of 20 m) were observed in the coastal marsh areas, situated in the northeastern part of French Guiana (Figure 3.2).

Figure 3.2. Map of canopy heights calculated from the airborne LiDAR dataset LD for (a) French Guiana and (b) a portion of the coastal marsh. Only 1% of canopy heights were higher than 50 m in all of French Guiana.



(a)

Figure 3.2 (cont.)



(b)

3.2.1.2 Processing the HD dataset

The estimation of canopy height from the airborne high-density (HD) dataset used a similar procedure. However, as the density of points is higher (on average 3.5 pts/m^2) than that of the LD dataset (on average 285.2 pts/km^2), several changes were made to account for the difference between the two datasets:

(1) The procedure described in Section 3.2.1.1 requires flight lines for top-of-canopy and ground point extraction. From the HD dataset, a grid of 1 m × 1 m was created over the

study sites. Then, two datasets were created: the first contained the point with the highest elevation in each square of the grid, and the second contained the lowest elevations.

(2) Using the grid of the highest elevations, the procedure developed in Section 3.2.1.1 for canopy top extraction was applied to extract canopy-top points along the East-West and North-South directions. The window size for the canopy top extraction differed between datasets according to their point density (between 20 and 50 m).

(3) Using the lowest elevations grid, the ground point's extraction procedure detailed in Section 3.2.1.1 was performed along the horizontal and vertical lines of the grid. However, unlike with the LD dataset, the window sizes used in the selection of ground points were much smaller (between 70 and 120 m, according to the HD dataset). The window sizes of the HD dataset were also determined using an analysis of variograms.

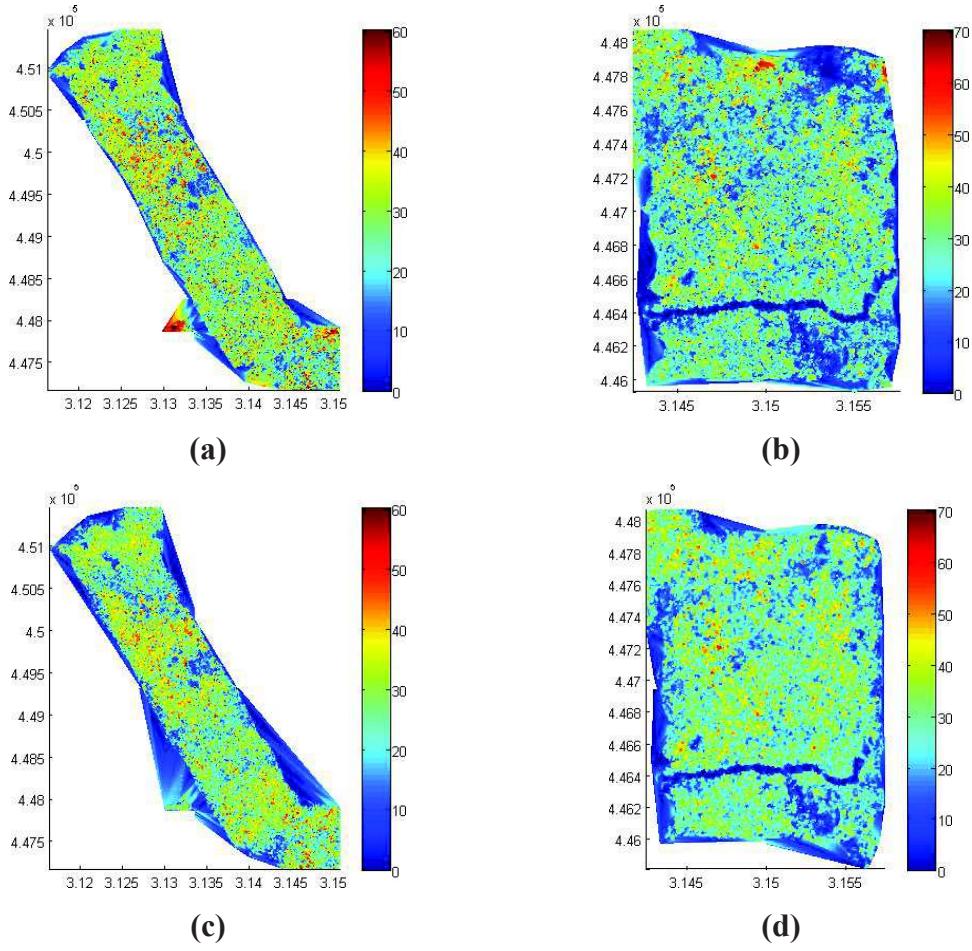
(4) Finally, as the distances between ground points and between canopy-top points were small, the estimation of canopy height was calculated at each canopy-top and ground point. However, unlike the LD dataset, the canopy heights were not estimated using linear interpolation but rather using bilinear interpolation. First, Delaunay triangulations were computed separately for the canopy-top and the ground points. Next, the triangle containing each ground point in the lat/lon plane of the top-of-canopy mesh was identified, and the ground point was projected on this triangle. Finally, the canopy height was calculated as the difference between the elevation of the projected ground point on the top-of-canopy mesh and the elevation of the actual ground point. A similar procedure was carried out for canopy height estimation at each canopy-top point using the projection of canopy-top points on the Delaunay triangles of the ground points' mesh.

3.2.1.3 Comparison of canopy height estimates from the HD dataset using different estimation methods

In this section the HD canopy height estimates obtained by our method were compared to the HD canopy height estimates obtained by the method used in Vincent *et al.* [80]. This comparison was conducted in order to: (1) Analyze the pertinence of the method used in our study to estimate the canopy heights; (2) assess the quality of these estimates. The

comparison was conducted on two sites in Nouragues, French Guiana. The results presented in Figure 3.3 indicate a good correlation between the canopy heights estimates from each method. The mean heights obtained from the two methods were similar with a mean difference about 1 m. This difference is due to the different methods used for the estimation of canopy heights. In this study, from the HD dataset, two datasets were created, one containing the top of canopy points and one containing the ground points. The canopy heights were estimated at the level of each ground and top of canopy point by projecting each ground point onto the top of canopy points and vice versa. However, for Vincent *et al.* [80], after identifying the canopy tops and ground points, the ground points were interpolated in order to create an evenly distributed digital elevation model. The same was done to the canopy tops to create a canopy surface model. Finally, the canopy heights are estimated by subtracting the canopy surface model and the digital elevation model.

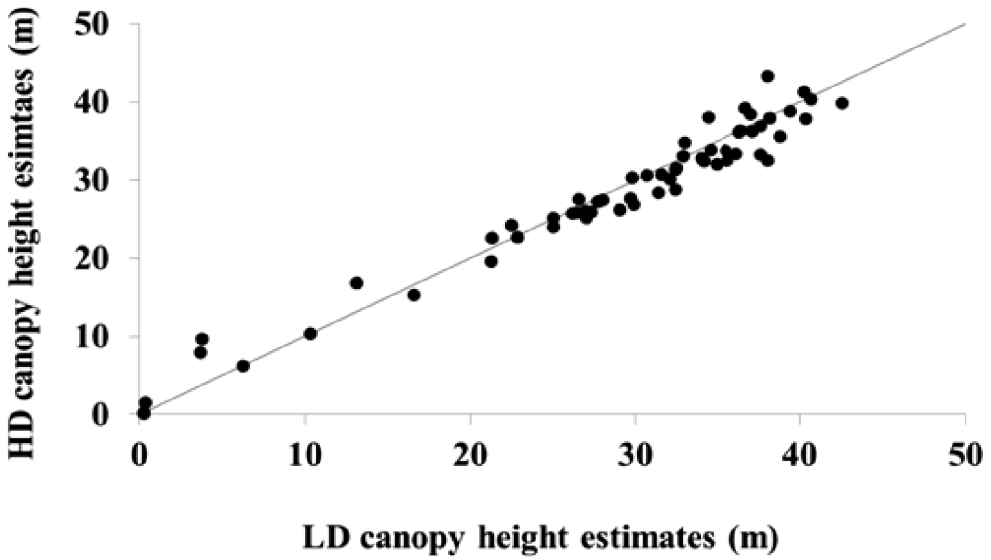
Figure 3.3. Comparison between canopy height estimates from the LD dataset over two sites in Nouragues (French Guiana) using the algorithm proposed by Vincent *et al.* [80] (a and b) and our algorithm (c and d)



3.2.1.4 Comparison of canopy height estimates from the LD and HD datasets

The canopy height estimates from the HD dataset are considered near-terrain measurements because of their small footprint size and high density. Unfortunately, the HD dataset does not intersect with the GLAS footprints. To use the LD dataset as reference data for GLAS's canopy height estimation models, the accuracy of the canopy heights of the LD dataset was assessed against the estimates from the HD dataset. For each LD canopy height estimate, the nearest point from the HD dataset, at a maximum distance of 10 m, was chosen. The results of the comparison between canopy heights from the LD and HD datasets showed a mean difference of 0.22 m, an RMSE of 1.57 m, and an R^2 of 93% (Figure 3.4).

Figure 3.4. Comparison between canopy height estimates from the LD and HD datasets.



3.2.1.5 Glas data processing

– ***GLAS waveform metrics extraction***

Several canopy height estimation models from GLAS waveforms have been developed in recent years (e.g. [44]; [47]; [72]; [81]; [71]; [78]). They depend on several parameters extracted from waveforms (primarily signal start and end, waveform extent, and leading and trailing edges) and on ancillary data such as DEMs (slope or terrain index).

Signal start and end are defined as the first and last locations where the waveform intensity exceeds a certain threshold level (n.σb, σb is the standard deviation of the background noise) above the mean background noise (μb) (Figure 2.2) [44]. Both μb and σb are found in the GLA14 product. The difference between the signal end and signal start is called the

waveform extent. However, there are no consistent optimal thresholds that can be used for every study area. Different thresholds have been used in different studies, including 3σb [82]; 3.5σb [81]; 4σb [44] and 4.5σb [78]. The difficulty in identifying the noise threshold could be explained by the difficulty in consistently identifying signal start and signal end.

The Gaussian peaks resulting from the decomposition of the GLAS waveform represent canopy features such as canopy top, canopy trunks, ground or a mix of these elements. The last Gaussian peak does not necessarily represent the ground return. Moreover, there is no general rule to determine the ground peak (e.g. [81]; [71]; [82]; [83]). Duong *et al.* [83] and Sun *et al.* [82] identified the ground as the last peak. Rosette *et al.* [81] and Chen [71] found that the elevation of the stronger of the last two Gaussian peaks has a better correspondence to the ground. In this study, the stronger of the last two Gaussian peaks was selected as the ground return.

The leading edge is defined as the difference between signal start and the first bin that is at half the maximum intensity (Figure 2.2). The trailing edge corresponds to the difference between signal end and the last bin that is at half the maximum intensity [78] (Figure 2.2). However, some LiDAR waveforms have a large difference in the intensity between the canopy and the ground peaks. If the ground peak return is significantly lower than the canopy peak, an overestimation of the trailing edge could be observed using Lefsky's metrics. Conversely, with a low intensity return from the canopy peak and a high intensity return from the ground peak, an overestimation of the leading edge could be observed using Lefsky's metrics [44]. Hence, Hilbert and Schmullius [62] proposed modified leading edge and trailing edge definitions. The modified leading edge is defined as the elevation difference between signal start and the canopy peak's center, and the modified trailing edge is the difference between signal end and the ground peak's center (Figure 2.2). These modified metrics better represent the characteristics of the canopy top and the ground surface. This study used the modified leading and trailing edges.

- ***Principal Component Analysis of GLAS waveforms***

PCA (and other types of dimensionality reduction) is a technique used to emphasize variation and bring out strong patterns in a dataset. It's often used to make data easy to explore and visualize. Looking at all the samples from a GLAS return signal, it is very difficult to see any relationships or natural groupings among the data. The actual samples are more than 450 per GLAS return signal, but using PCA, it is easier to find the dimensions that are the most informative about the ways in which those measurements differ. The

identified principal components are then analyzed in order to find how they relate (if they do) to properties of the canopies (e.g. canopy heights).

PCA of LiDAR waveforms has been conducted in a handful of studies. Allouis *et al.* [84] used PCA to estimate the water depth in shallow water using airborne LiDAR waveforms. Principal components were then used to perform a regression model between the principal components and water depth. The model relying on PCA for water depth estimation provided the lowest mean error and had the lowest detectable water depth in comparison to other models (mathematical approximation, Heuristic methods, statistical approaches, and convolution methods). However, to convert waveform samples into principal components, further processing of the GLAS waveforms was required. First, the parts of the waveforms useful for canopy height estimation corresponding to the waveform extent were extracted. Next, because not all the waveforms have the same waveform extent, the waveform with the largest extent was identified, and waveforms with shorter waveform extents were padded with the remaining waveform samples after the signal end to give them the same length as the largest waveform extent (same sample count). Note that the first sample of the extracted waveform now corresponds to signal start. In this study, the largest waveform extent had 470 samples. Next, the extracted waveform samples were converted into principal components (PCs), and the number of PCs to be used in the regression model for dominant canopy height estimation (Hmax) was calculated. The number of PCs used in the regression model has a major impact on the performance of the model, as choosing too many PCs will include noise from the sampling fluctuations in the analysis and by choosing too few, relevant information will be lost. A vast literature has developed methods to choose the statistically significant PCs. In this study, the number of statistically significant PCs was determined using a statistical process based on the study by Karlis *et al.* [85]. The PCs with eigenvalues higher than a certain threshold were selected. The threshold (λ) was defined as follows:

$$\boldsymbol{\lambda = 1 + 2\sqrt{\frac{p-1}{n-1}}} \tag{3.1}$$

where *p* is the number of variables (PCs) and *n* is the number of observations (waveforms). For our dataset composed of 470 variables and 474 observations, the threshold (λ) was determined at 2.99. Thus, the first 13 PCs were selected.

3.2.2 Background on GLAS canopy height estimation

3.2.2.1 Direct method

The estimation of the canopy height using the direct method is simply the difference between the waveform signal start (canopy top) (H_b) and the ground peak (H_g):

$$H_{max} = H_b - H_g \tag{3.2}$$

The direct method estimates the canopy height with good precision over flat areas. An average difference between GLAS and airborne LiDAR data lower than 3 m was observed in several studies (e.g. [72]; [86]).

3.2.2.2 Multiple regression models using GLAS and DEM metrics

Over sloping areas, both the ground and vegetation peaks are broader and lower in intensity (e.g. [64]; [63]). The peak identified as the ground peak will no longer represent only the ground but a mix of ground and terrain objects (e.g. [71]; [70]). In fact, over sloped terrain, waveform extent will increase with the terrain slope and the footprint size [70]. This increase will lead to an earlier detection of the signal start and this will lead to an overestimation of the canopy height [87].

To correct for the effect of terrain slope on the GLAS signal, several studies have developed models to better estimate canopy heights. Lefsky *et al.* [78]; Pang *et al.* [64]; Duncanson *et al.* [87]; and Chen [71] developed models based on parameters derived from the waveforms themselves (waveform extent "Wext", leading edge "Lead" and trailing edge "Trail"). Lefsky *et al.* [44] and Rosette *et al.* [81] developed models based on the waveform extent and terrain index. The terrain index as defined by Lefsky *et al.* [44] is the difference between the maximum and minimum elevations in an m x m sampling window applied to a DEM at the GLAS footprint location. The window size depends on the resolution of the DEM. A 3x3 window has been deemed best for a 90-m-resolution DEM [44].

The first model was developed by Lefsky *et al.* [44] for the estimation of the tallest canopy within a footprint:

$$H_{max} = aW_{ext} - b \cdot TI \quad (3.3)$$

This model is based on the waveform extent (Wext) and the terrain index (TI). The incorporation by Lefsky *et al.* [44] of the waveform leading edge extent in equation (3.3) resulted in a slight improvement in the canopy height estimation:

$$H_{max} = aW_{ext} - b \cdot TI + cLead \quad (3.4)$$

Pang *et al.* [64] introduced a model to estimate forest canopy height by using metrics derived from the waveforms themselves:

$$H_{max} = aW_{ext} - (b(Lead + Trail))^c \quad (3.5)$$

Chen [71] introduced the following model to show how a linear model compares to equation 3.5:

$$H_{max} = aW_{ext} - b(Lead + Trail) \quad (3.6)$$

Finally, Lefsky *et al.* [47] proposed a modification of the Lefsky *et al.* [78] model to produce a better estimation when the leading and trailing edges are small:

$$H_{max} = aW_{ext} - bLead - cTrail \quad (3.7)$$

In addition, to quantify the contribution of Lead and Trail in the canopy height estimation models, two additional models were analyzed: one that replaces Lead with Trail in equation 3.4 and one that removes Lead in equation 3.6 (model IDs 7 and 8, respectively, Table 3.1). Finally, each of the eight models was tested with an added intercept (the bis models, Table 3.1). The coefficients of the different models were fitted with least squares regressions using the canopy height estimates from the LD dataset. The reference LD canopy height estimate for model calibration were chosen as the closest points to GLAS footprints, with a distance not exceeding 50 m (near to the 35 m average radius of the GLAS footprints). The least squares regression is an approach for modeling the relationship between a dependent variable (i.e. canopy height) and one or more explanatory variables (i.e. Lead, Trail, Wext, TI, etc.). The technique is based on fitting a straight line (regression line or line of best fit) to the observed data (plotted as a scatter plot) this technique aims to derive a good relationship (the best fit) that may be used to predict future values of one variable

when the value of the other is known. It is named as such because, in its computation, the sum of the squared deviations of the computed (future) values from the observed (past) values of the variables is minimized. Devised by the French mathematician Adrien-Marie Legendre (1752-1833), this technique is applicable to single line functions with any number of independent variables and, under certain assumptions, is the best statistical estimator.

The best regression model was selected based on the Akaike information criterion (AIC) [88]; the coefficient of determination (R^2), and the root mean square error (RMSE). AIC is a measure of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Hence, AIC provides a means for model selection. However, AIC does not provide a test of a model in the sense of testing a null hypothesis; i.e. AIC can tell nothing about the quality of the model in an absolute sense. If the entire candidate models fit poorly, AIC will not give any warning of that. R^2, is a statistical method that explains how much of the variability of a factor can be caused or explained by its relationship to another factor. Coefficient of determination is used in trend analysis. It is computed as a value between 0 (0 percent) and 1 (100 percent). The higher the value, the better the fit. Coefficient of determination is symbolized by R^2 because it is square of the coefficient of correlation symbolized by R or r. The coefficient of determination is an important tool in determining the degree of linear-correlation of variables ('goodness of fit') in regression analysis. The RMSE is square-root of the sum of the squared of all the rows in the error vector divided by the number of observations. And this error vector is obtained by doing $e = y - y_{est}$; where ‘e’ is the residual vector, y is a vector of measured dependent vector, and y_{est} is the vector of the estimated values. Finally, to assess how the model results will generalize to an independent data set, a 10-fold cross validation was used. Large k-fold values mean less bias towards overestimating the true expected error (as training folds will be closer to the total dataset). Cross validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. In a prediction problem, a model is usually given a dataset of known data on which training is run (training dataset), and a dataset of unknown data (or first seen data) against which the model is tested (testing dataset).

Table 3.1. Regression models' fitting statistics calculated with 10-fold cross validation for estimating forest height. R = root mean square error, AIC = Akaike information criterion.

Model	ID	R^2	RMSE (m)	AIC
$H_{max} = H_b - H_g$	1	0.50	7.9	3126
$H_{\max} = 0.6527W_{ext} - 0.0184TI$	2	0.72	4.9	2221
$H_{\max} = 0.5405W_{ext} - 0.0262TI + 6.427$	2bis	0.73	4.4	2185
$H_{\max} = 0.6682W_{ext} - 0.0029TI - 0.0261Lead$	3	0.73	4.7	2223
$H_{\max} = 0.5395W_{ext} - 0.2557TI - 0.0115Lead + 6.8876$	3bis	0.73	4.6	2187
$H_{\max} = 0.7555W_{ext} - \{0.0994(Lead + Trail)\}^{1.5903}$	4	0.80	3.9	2084
$H_{\max} = 0.6908W_{ext} - \{0.1315(Lead + Trail)\}^{1.3109} + 3.3309$	4bis	0.80	3.9	2081
$H_{\max} = 0.7965W_{ext} - 0.2707(Lead + Trail)$	5	0.79	3.9	2096
$H_{\max} = 0.6972W_{ext} - 0.2461(Lead + Trail) + 4.1452$	5bis	0.79	3.9	2083
$H_{\max} = 0.6739W_{ext} - 0.0751Lead - 0.2959Trail$	6	0.85	4.0	2064
$H_{\max} = 0.6739W_{ext} - 0.0751Lead - 0.2959Trail + 4.1823$	6bis	0.85	3.9	2056
$H_{\max} = 0.7377W_{ext} + 0.0235TI - 0.3192Trail$	7	0.81	3.8	2063
$H_{\max} = 0.6656W_{ext} - 0.0026TI - 0.28899Trail + 3.679$	7bis	0.81	3.7	2051
$H_{\max} = 0.7494W_{ext} - 0.3184Trail$	8	0.81	3.8	2064
$H_{\max} = 0.6654W_{ext} - 0.2904Trail + 3.6344$	8bis	0.81	3.8	2056
$H_{max} = a_1PC_1 + a_2PC_2 + \cdots + a_{13}PC_{13}$	9	0.52	5.9	2373
Most important PCs (PC1, PC2, PC4, PC11) from ID 9	9bis	0.47	6.2	2478
$H_{max} = W_{ext} + a_1PC_1 + a_2PC_2 \ldots + a_{13}PC_{13}$	10	0.80	3.8	2047
Most important PCs (PC1, PC2, PC4, PC11) from ID 10	10bis	0.79	3.9	2075
$H_{max} = 0.63W_{ext} - 0.10PC_1 + 0.05PC_2 + 0.02PC_3 + 1.3$	11	0.73	4.4	2174
$H_{max} = WC_i + a_1PC_1 + a_2PC_2 \ldots + a_{13}PC_{13}$	12	0.78	4.0	2064
Random Forest using: Wext + Lead + Trail + TI	13	0.82	3.4	-
Random Forest using: Wext + Lead + TI	14	0.80	3.6	-

Random Forest using: Wext + Lead	15	0.80	3.6	-
Random Forest using: Wext + TI	16	0.82	3.6	-
Random Forest using: Wext	17	0.73	4.4	-
Random Forest using: First 13 PC	18	0.70	4.7	-
Random Forest using: PC1 + PC2 + PC4 + PC11	18bis	0.69	4.8	-
Random Forest using: Wext and the first 13 PC	19	0.83	3.6	-
Random Forest using: Wext +PC1 + PC2 + PC4 + PC11	19bis	0.82	3.6	-
Random Forest using: WC and the first 13 PC	20	0.81	3.7	-
Random Forest using: WC +PC1 + PC2 + PC4 + PC11	20bis	0.81	3.7	-

3.2.2.3 Proposed techniques for canopy height estimation

– ***Multiple regression models using principal components***

The previous section introduced a number of regression models developed in various studies for the estimation of canopy height. However, these models require several metrics derived from GLAS footprints, such as ground peak, canopy-top peak, leading and trailing edge extents, and metrics derived from ancillary data (SRTM DEM), such as terrain index. Moreover, the extraction of some metrics from GLAS waveforms, such as the location of the ground peak, is error-prone, especially in dense forests, such as those in French Guiana. Processing the GLAS data revealed that a considerable number of waveforms taken only over dense forests had the canopy-top location easily identified. In fact, canopy penetration of the waveform in densely vegetated areas was sometimes insufficient to reach the ground; thus, either the ground peak was unidentifiable or the waveform did in fact reach the ground but the return signal was not strong enough for reliable detection. These difficulties in the detection of the ground peak affect the estimation of the trailing edge extent and, ultimately, the estimation of the canopy height. Therefore, a statistical model for canopy height estimation based only on the waveform samples might be an interesting alternative. In this section, a principal component analysis of GLAS waveforms was conducted. A stepwise linear regression model was built for canopy height estimation using the principal components (PCs). A regression model using PCs takes advantage of

model building on orthogonal variables. The regression model using the 13 first PCs for canopy height estimation could be written as follows:

$$\boldsymbol{H_{max} = a_1 PC_1 + a_2 PC_2 + \cdots + a_{13} PC_{13}} \quad (3.8)$$

where PC_i are the principal components, and a_i are the coefficients to be applied to the principal components.
This model based on principal component analysis for canopy height estimation will be compared to the regression models developed in the previous section to quantify the benefits of using waveform data (PCA model) instead of metrics extracted from the waveform.

– ***Random forest regressions using GLAS and DEM metrics***

In Section 3.2.2.2, linear regressions were developed to estimate the canopy height for each GLAS footprint. These regressions linked the canopy height estimated from the LD data to the GLAS and SRTM metrics (waveform extent, leading edge, trailing edge, and terrain index). In this section, the Random Forest (RF) technique was evaluated using the following different configurations:

(1) All the metrics were used to estimate the canopy height (waveform extent "Wext", leading edge extent "Lead", trailing edge extent "Trail", and terrain index "TI");

(2) The Trail metric was removed because in densely forested areas, such as tropical forests, the LiDAR echo seldom reaches the ground, making the ground peak difficult to identify; thus, the Trail metric is often inaccurate;

(3) To study the effects of Trail and TI on the canopy height estimates, the TI and Trail metrics were removed (only the Wext and Lead were used). This case shows promise in the use of the SRTM DEM in a low relief area;

(4) Only Wext and TI were used to assess the impact of the Lead and Trail metrics on the performance of Random Forest for canopy height estimates;

(5) Only Wext was used. This case evaluated the impact of using Lead, Trail, and TI with Wext for canopy height estimates. The relative importance of the different metrics used in Random Forest for the canopy height estimates was also analyzed. Variable importance is based on two measures. The first is a measure of accuracy obtained by quantifying the mean squared error increase in the model by the removal of a variable. The other importance measure is the Gini index, which quantifies the degree to which a variable produces terminal nodes in the classification forest. Finally, to validate the generalization performance of the Random Forest regressions, the error in the estimation of the canopy height was assessed using a 10-fold cross validation. The performance of the different configurations was assessed by comparing the canopy height estimates from Random Forest regressions and the canopy heights extracted from the LD dataset, which were used as the reference data.

Several studies have shown that, for many applications, the Random Forest technique is extremely powerful in estimating biophysical parameters (e.g. [89]; [90]; [91]; [92]). Random Forest can be used as a classifier or a regression algorithm consisting of an ensemble of regression or decision trees but takes a different approach [93]. A decision tree is a graph-like structure that uses a branching method to illustrate every possible outcome of a decision. The tree elements are called nodes (starting at a root node). The lines connecting elements are called branches. A node extending from another node is called a child node. Nodes without children are called leaf nodes, end-nodes, or leaves. Nodes with children nodes, are also called internal nodes. At each internal node, one feature is selected to make a separating decision. That is, the feature that separates instances to classes with the best possible purity is selected. This purity is measured by entropy, Gini index or information gain. Each branch represents the outcome of the test and each leaf node represents a class label (decision taken after computing all features). The paths from root to leaf represents classification rules. However the problem about the decision tree algorithm, it is sensitive to slight changes of the data since these changes are able to change the tree drastically (Decision trees resemble deterministic structures). Therefore, the final structure, and as such, the final decisions are bugled. Random Forest uses decision trees, but takes a different approach. Random Forests, rather than growing a single, very deep tree that is carefully overseen by an analyst, Random Forest relies on

aggregating the output from many shallow trees that are tuned and pruned without much oversight. Some of these trees may have been grown from samples that said a certain variable was the more important feature. Other trees may find completely different features to be relevant. The idea is that the errors from many shallow trees will wash out when aggregated and lead to a more accurate prediction.

3.2.2.4 Random forest regressions using principal components

Similar to the previous section, the first 13 principal components were used in the Random Forest regression to link the canopy heights estimated from LD data to these PCs. This model based on principal component analysis and Random Forest regressions was compared to other the models performed in this study.

3.3 Results

3.3.1 Direct method

The comparison between the canopy height estimates from GLAS waveforms using the direct method and the canopy height estimates from the LD dataset showed a high RMSE of 7.9 m for the estimation of the GLAS canopy height and a low R^2 of 0.50 (Figure 3.5a). This result can be explained by the fact that most of the footprints were in an area with a slope between 5° and 10°.

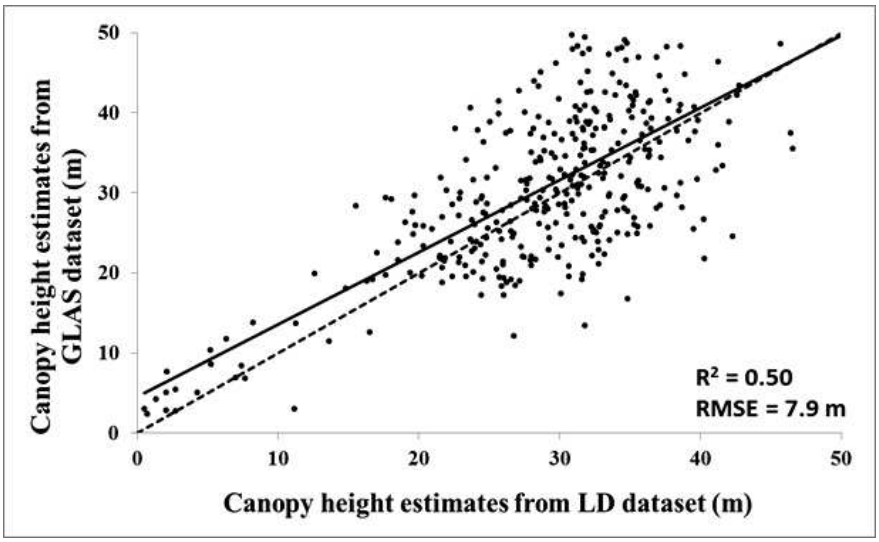
3.3.2 Multiple regression models

3.3.2.1 Using GLAS and DEM metrics

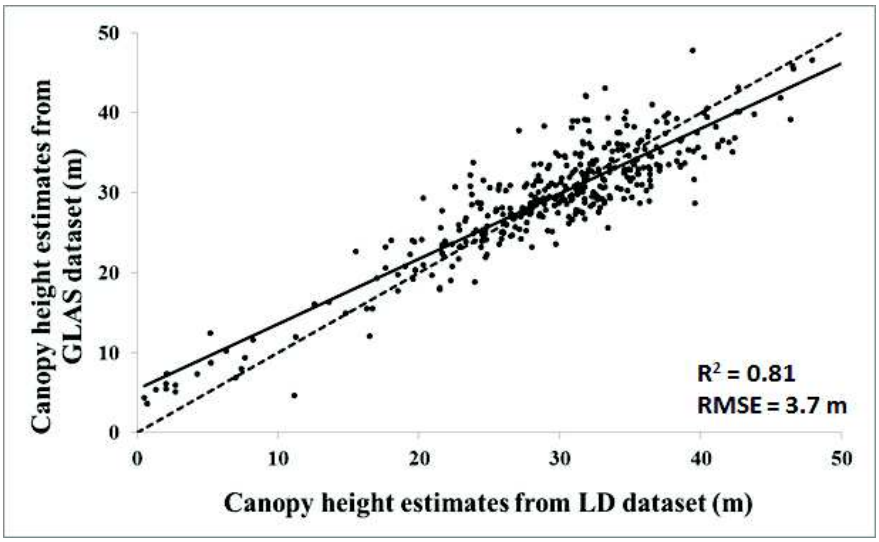
The results of the regression models with 10-fold cross validation showed that the regression models using the trailing edge extent (model IDs 4 to 8, Table 3.1) provided slightly better estimations of canopy height. For these models, AIC ranged between 2051 and 2096, RMSE ranged between 3.7 and 4.0 m, and R^2 between 0.79 and 0.81. The best results in estimating forest height were obtained with model ID 7bis (Table 3.1, Figure 3.5b). The contribution of the leading edge extent appeared to be weak in comparison to the trailing edge extent when estimating the maximum canopy height. Indeed, model IDs 7 and 7bis, which used Trail, had better results than model IDs 3 and 3bis (Table 3.1), which used Lead. Moreover, the use of information calculated from a DEM (terrain index)

alone in the regression models had the lowest estimation accuracy for the canopy height (model ID 2, Table 3.1, Figure 3.5c) (RMSE of 4.9 m and R^2 of 0.72).

Figure 3.5. Canopy height estimates from GLAS data in comparison to estimated canopy heights from the LD dataset: (a) using the direct method (model ID 1, Table **3.1), (b) using the model with Wext, TI and Trail (model ID 7bis,** Table **3.1), and (c) using the model with Wext and TI (model ID 2,** Table **3.1).**

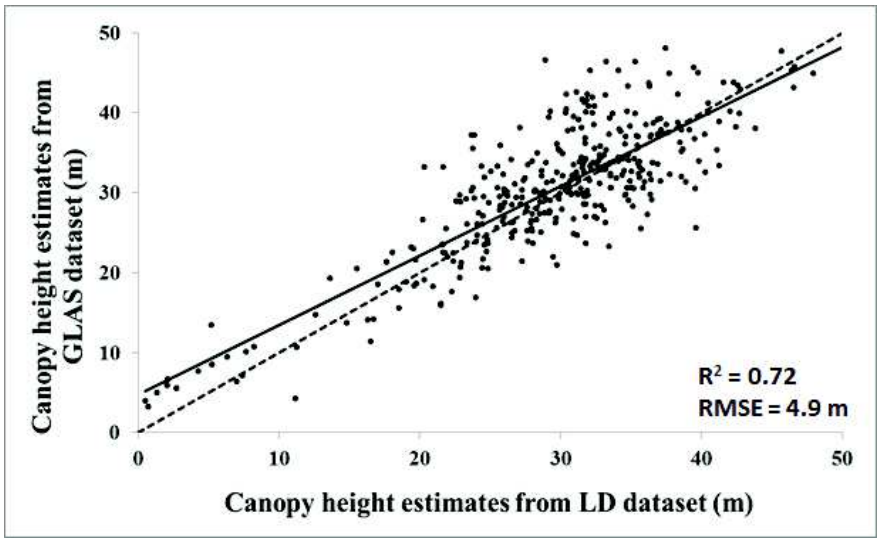


(a)



(b)

Figure 3.5 (cont.)



(c)

3.3.2.2 Using principal components

The results of the PCA model for canopy height estimation showed an estimation accuracy with an R^2 of 0.52 and an RMSE of 5.9 m (Figure 3.6a). To reduce the number of PCs involved in the PCA model, stepwise regression was used to extract the most important principal components. The resulting model, which used 6 principal components containing 76.3% of the waveforms' inertia, showed an R^2 of 0.47 and an RMSE of 6.2 m. Figure 3.6a shows that the PCA model appeared to overestimate canopy heights for canopies with heights lower than 20 m. To improve height estimation of these canopies, a regression model incorporating both the first 13 principal components and the waveform extent was performed:

$$\boldsymbol{H_{max} = W_{ext} + a_1 PC_1 + a_2 PC_2 + \cdots + a_{13} PC_{13}} \quad (3.9)$$

The new PCA regression model for canopy height estimation accounting for the waveform extent showed better canopy height estimation results in comparison to the PCA model without information on the waveform extent, with an RMSE of 3.8 m and an R^2 of 0.80 (Figure 3.6b). Using only the seven most important components from the stepwise regression, the R^2 decreased to 0.79 and the RMSE increased to 3.9 m. Furthermore, using only the first three principal components with the waveform extent, the R^2 decreased to 0.73, and the RMSE increased to 4.4 m.

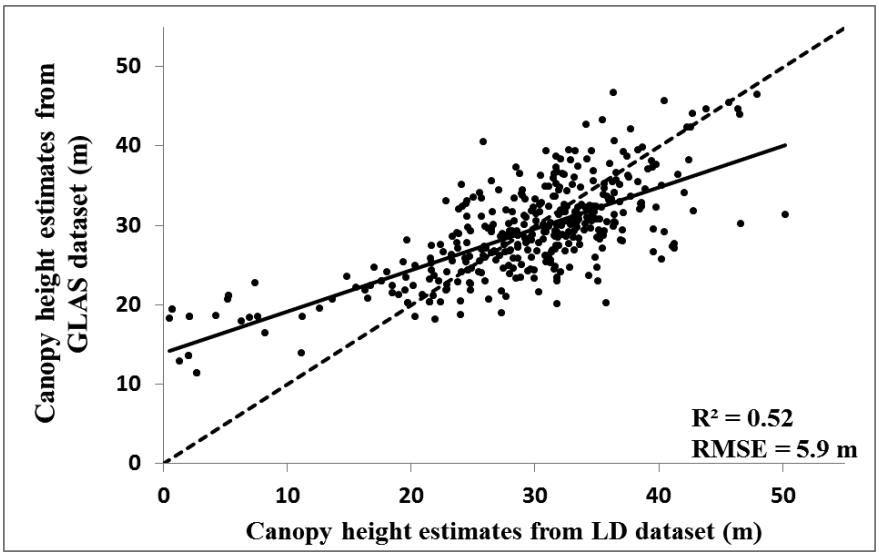
Next, the waveform extent was replaced by a waveform extent factor class (WC): (1) WC_1 for waveform extents lower than 20 m, (2) WC_2 for waveform extents between 20 and 40 m, and (3) WC_3 for waveform extents higher than 40 m. The resulting regression model using all the principal components and the WC has the following form:

$$\boldsymbol{H_{max} = WC_i + a_1 PC_1 + \cdots + a_{13} PC_{13}} \quad (3.10)$$

Where WC_i is the intercept to be applied to the model depending on the waveform extent (i = 1, 2, or 3). The values of WC_i are 7.78, 25.83 and 32.01 for WC_1, WC_2, and WC_3, respectively.

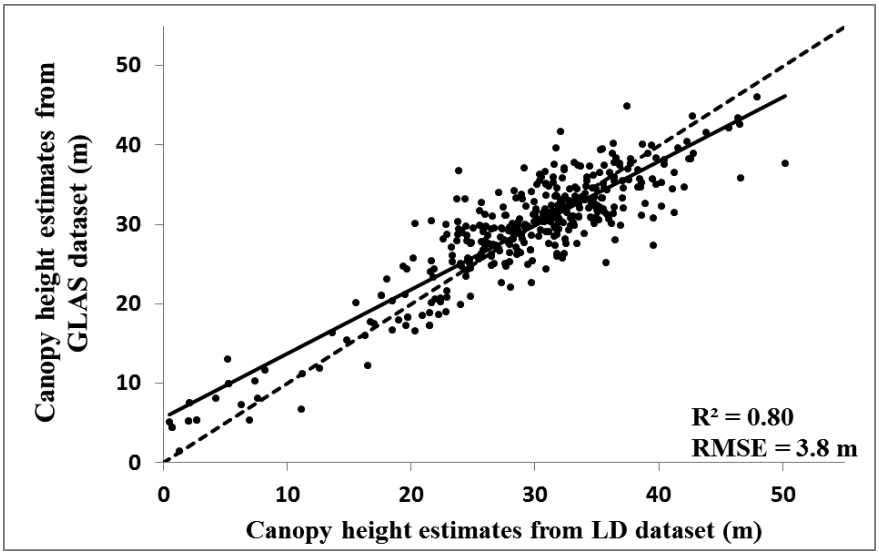
The new PCA regression model for canopy height estimation with information on the waveform extent showed slightly less canopy height estimation accuracy in comparison to model ID 9 (Table 3.1), with an RMSE of 4.0 m and an R^2 of 0.78 (Figure 3.6c).

Figure 3.6. Comparison between canopy height estimates using the PCA regression models and those estimated from low-density airborne LiDAR data (LD) (a) using the first 13 PCs, (b) using the first 13 PCs with the waveform extent, and (c) using the first three PCs with the waveform extent.

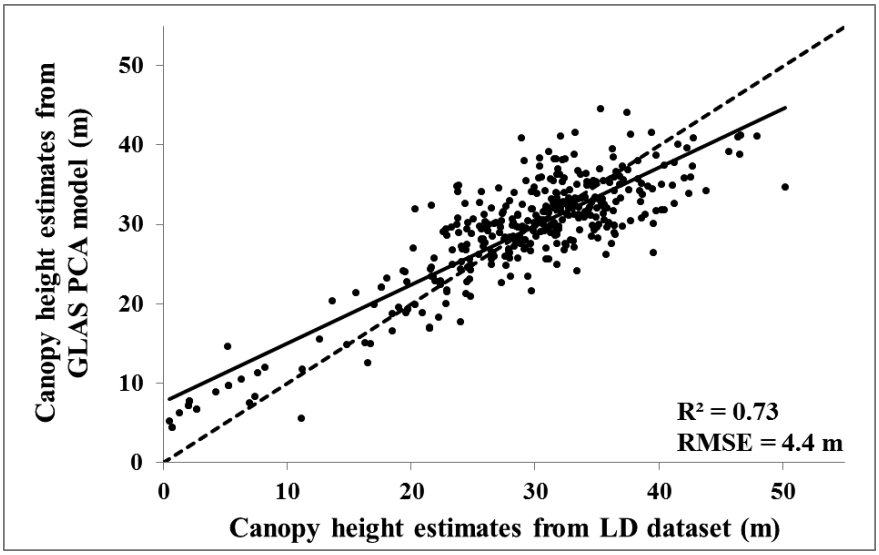


(a)

Figure 3.6 (cont.)



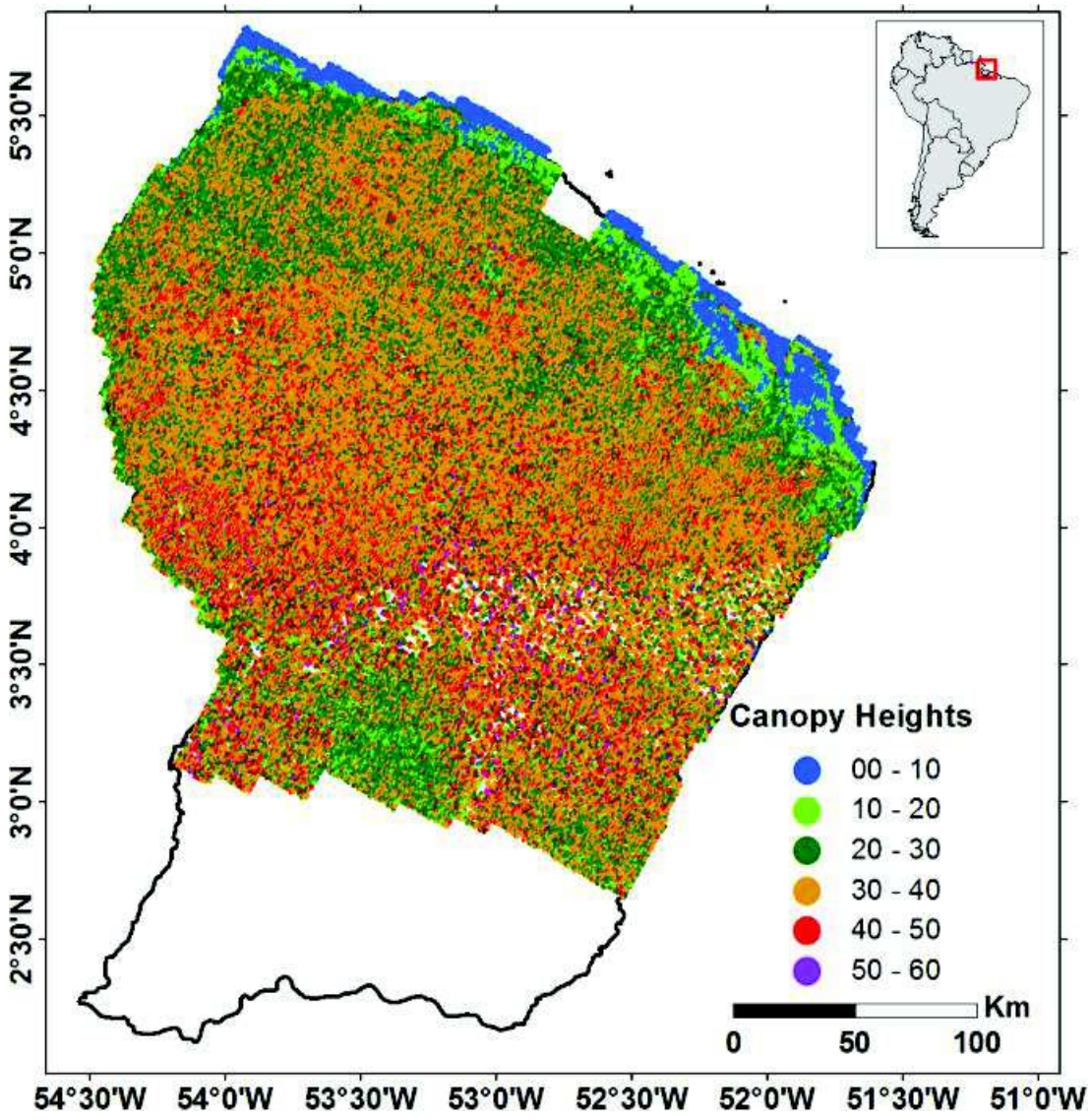
(b)



(c)

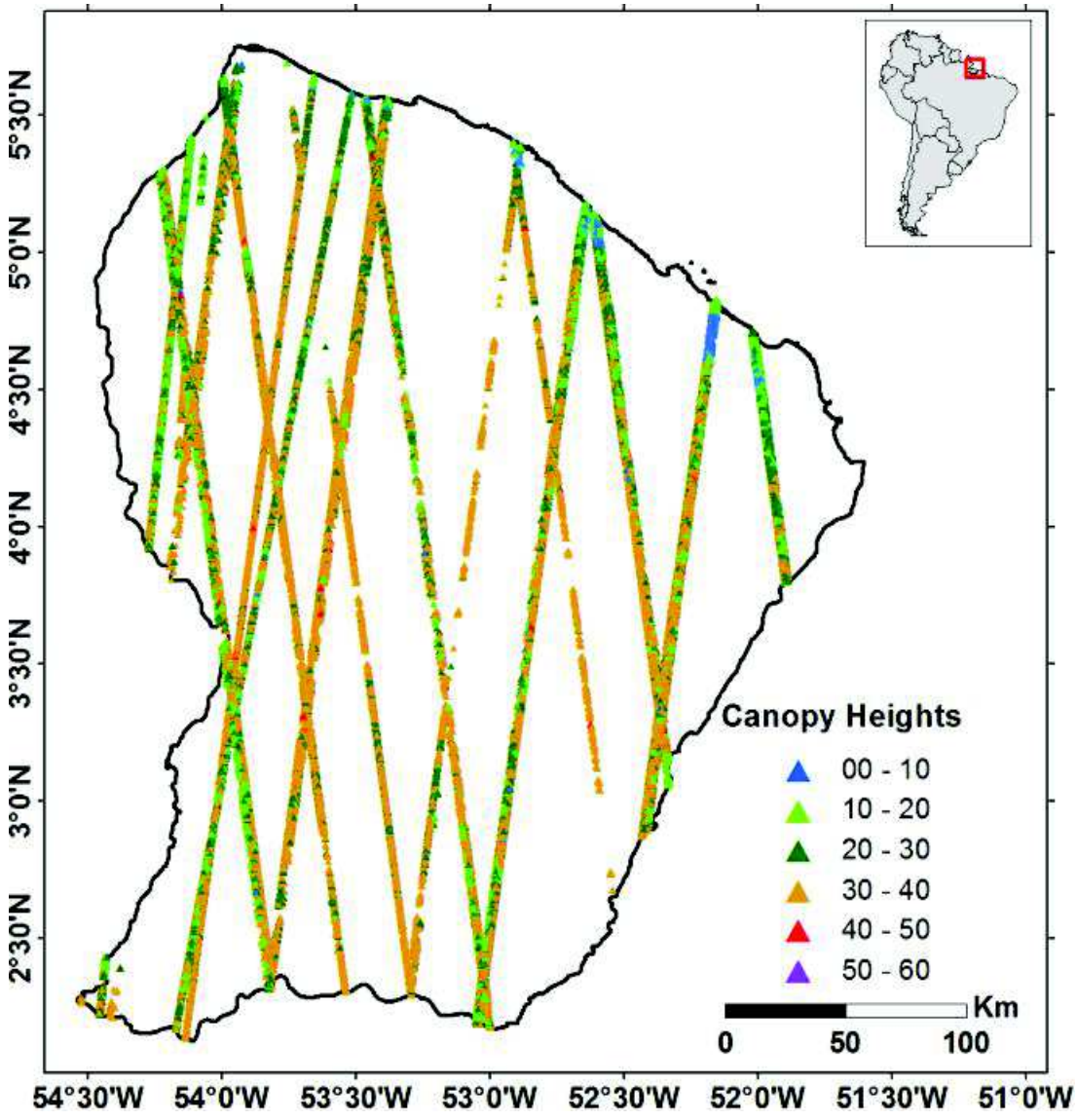
Like previously, stepwise regression was used to extract the most important PCs. The resulting model using using PCs and containing 76.3% of the waveforms' inertia showed slightly lower performance in comparison to the PCA model that used all the PCs and the WC factor, with an RMSE of 4.2 m and an R^2 of 0.76. Figure 3.7 shows the canopy height estimates from the LD and GLAS datasets. Good agreement was observed between the two canopy height maps.

Figure 3.7. (a) Map of canopy heights estimated from the LD dataset; (b) Map of canopy heights estimated from the GLAS dataset using the PCA model; (c) Overlapping of the two maps over a small area of French Guiana.



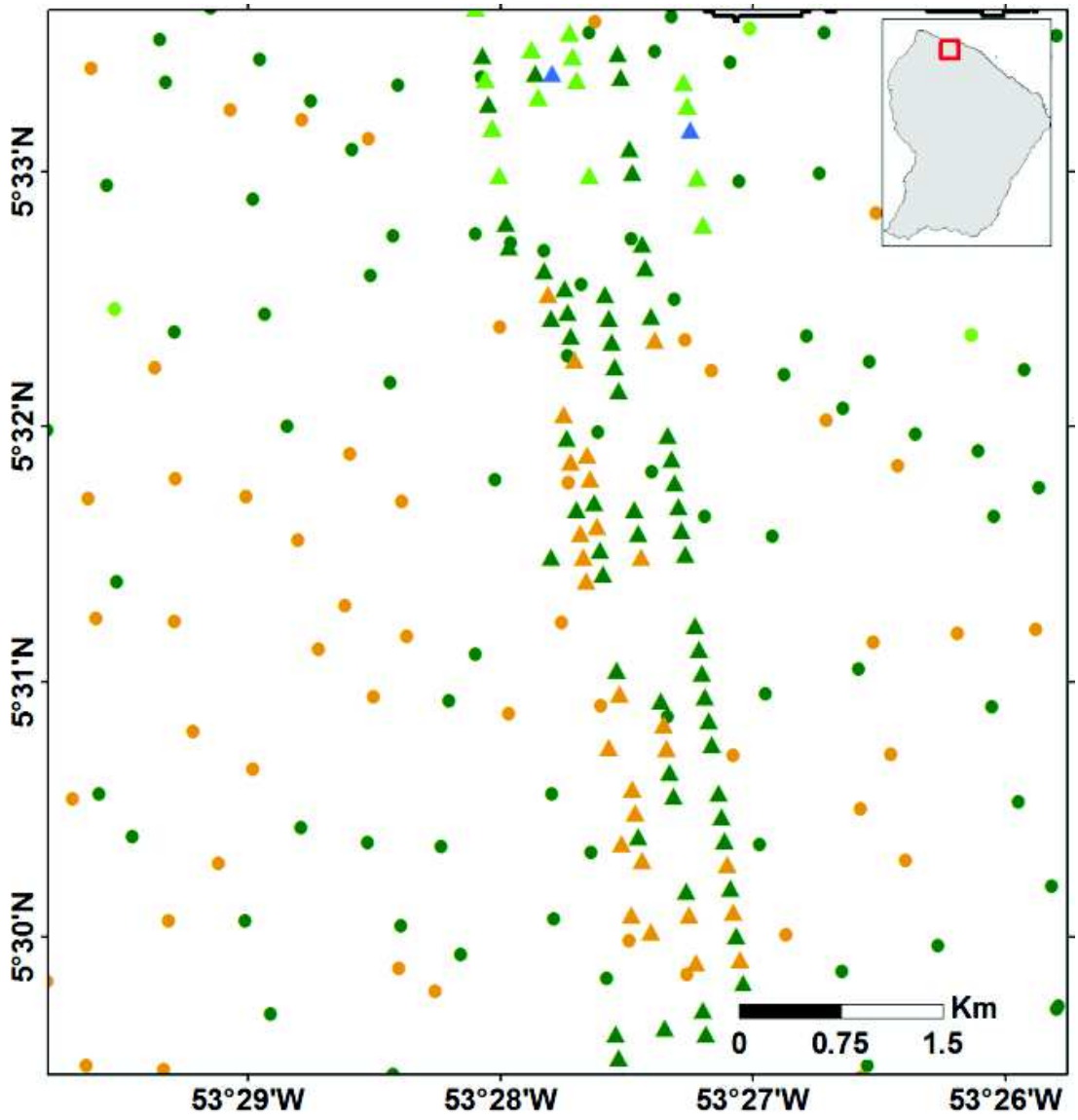
(a)

Figure 3.7 (cont.)



(b)

Figure 3.7 (cont.)



(c)

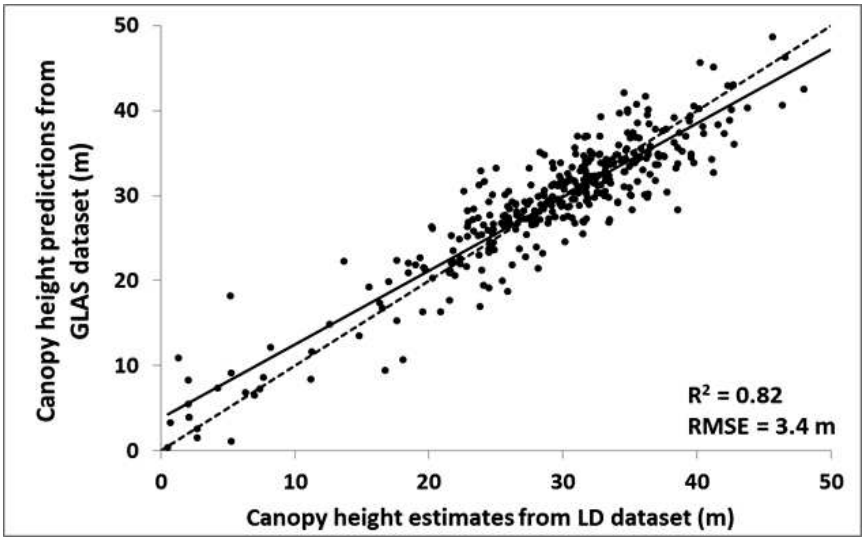
3.3.3 Random forest regressions

3.3.3.1 Using GLAS and DEM metrics

To analyze the precision of the canopy height estimation using Random Forest, several configurations were tested, and the results reveal that the best configuration for canopy height estimation is the one that uses all the metrics: waveform extent, leading edge, trailing edge, and terrain index (model ID 13, Table 3.1). The difference between the GLAS canopy height estimates and those estimated from the LD (reference data) in the first configuration had an RMSE of 3.4 m and a coefficient of determination R^2 of 0.82

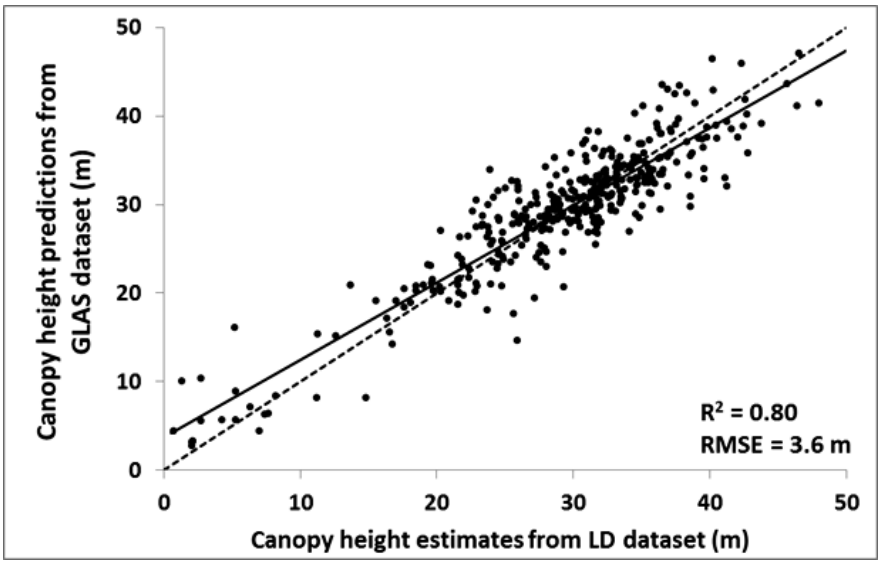
(Figure 3.8.a). Moreover, the variable importance test of the metrics showed that the GLAS canopy height is best explained using Wext, with an importance factor almost three times higher than those for the other three metrics; meanwhile, the other metrics (Trail, Lead, and TI) had almost the same importance. Other configurations using Wext, Lead, and TI (Figure 3.8.b); Wext and Lead; and Wext and TI (model IDs 10, 11 and 12, respectively, Table 3.1) showed a slightly lower precision in the canopy height estimation (RMSE) (approximately 3.6 m). The estimation of the GLAS canopy height using only Wext had an RMSE of 4.4 m with an R^2 of 0.73 (Figure 3.8.c). These results show that, in a low relief area, the use of other metrics in addition to the waveform extent only slightly improved the precision of the estimation of canopy height regardless of which metric was used. The use of one metric (among Trail, Lead and TI) in addition to Wext improved the estimation of canopy heights by approximately 1 m. Moreover, the use of more than one of these metrics in addition to Wext did not improve the estimation of canopy heights.

Figure 3.8. Comparison of estimated canopy heights using Random Forest regressions and estimated canopy heights from the LD dataset for three metrics configurations: (a) Wext + Lead + Trail + TI; (b) Wext + Lead + TI; and (c) Wext.

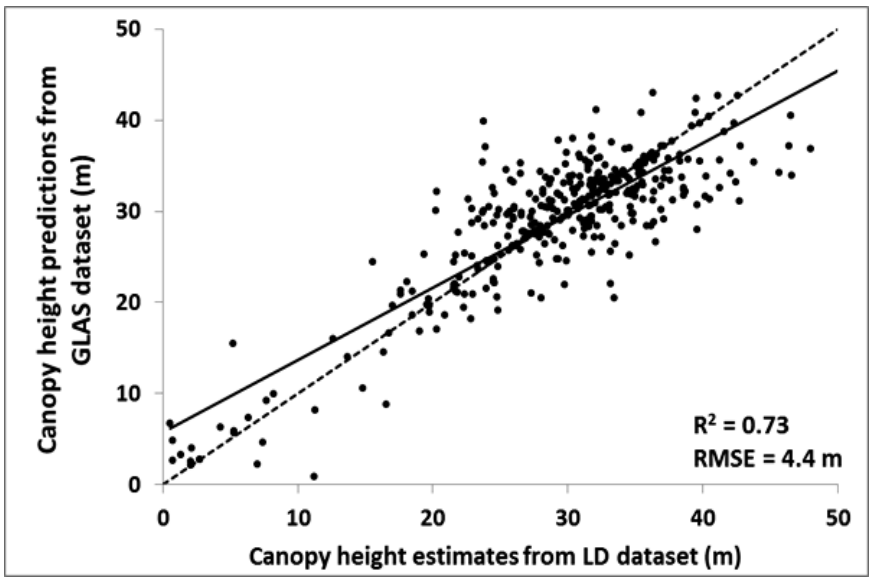


(a)

Figure 3.8 (cont.)



(b)



(c)

3.3.3.2 Using principal components

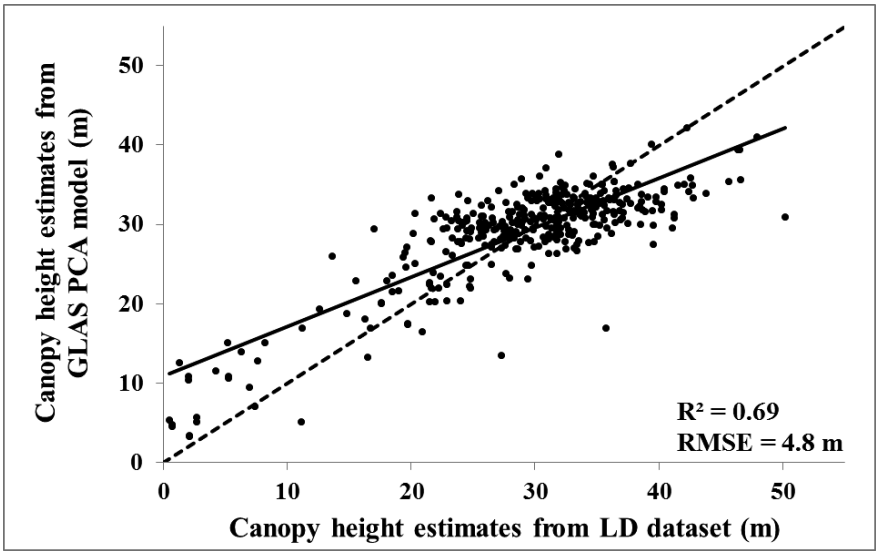
In this section, canopy height estimations with Random Forest regressions using PCs were performed with different configurations. Using the first 13 PCs in the Random Forest regression resulted in better canopy height estimation precision (RMSE = 4.7 m, $R^2 = 0.7$) in comparison to the linear regression model that used the first 13 PCs in Section 3.3.31 (RMSE = 5.9 m, $R^2 = 0.52$). The variable importance test showed that GLAS canopy height is best explained using *PC1*, *PC2*, *PC4*, and *PC11* (variance 62.38%). Using only

these four PCs in the Random Forest model had a similar result (RMSE = 4.8 m, R^2 = 0.69). Next, the incorporation of the waveform extent in addition to the first 13 principal components greatly improved the precision of the canopy height estimation (RMSE = 3.6 m, R^2 = 0.83) in comparison to the RF regressions without Wext. In addition, this result is slightly better to the one obtained using a linear multiple regression with the first 13 PCs and Wext (RMSE = 3.8 m). Using the most important variables (Wext, *PC1*, *PC2*, *PC4*, and *PC11*) in the RF regression yielded similar results, with an RMSE of 3.6 m and an R^2 of 0.82. Finally, replacing the waveform extent by the waveform extent factor class (WC) in addition to the first 13 PCs in the Random Forest regression for canopy height estimation showed similar results (RMSE = 3.7 m, R^2 = 0.81). Similar findings were noted when retaining only the most important variables (WC, *PC1*, *PC2*, *PC4*, and *PC11*), with an RMSE of 3.7 m and an R^2 of 0.81. Figure 3.9 shows examples of the comparison between the GLAS canopy heights using PCs and the Random Forest technique and the reference canopy heights estimated from the LD dataset.

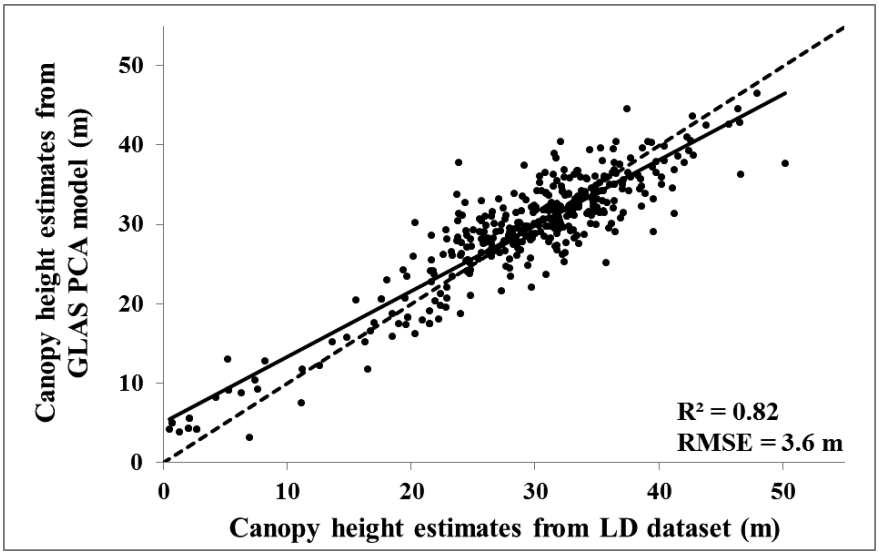
3.3.4 Model performance in different forest conditions

In previous sections, different models were applied on GLAS footprints over French Guiana in order to estimate forest canopy heights. Most models performed well, with an estimation precision lower than 4.5 m on the estimation of canopy heights. In this section, the two best models (model 7bis and 19, Table 3.1) were tested for different slopes and forest types, in order to analyze how the models would adapt in different forest conditions.

Figure 3.9. Comparison between canopy height estimates using the most important PCs in Random Forest regression models and those estimated from Low Density airborne LiDAR data (LD) using (a) the most important PCs (*PC1*, *PC2*, *PC4*, and *PC11*) and (b) the most important PCs with the waveform extent.



(a)



(b)

In our study site, the distribution of the slopes shows that 80% are lower than five degrees, 17% between five and 10 degrees and 3% higher than 10 degrees. Based on these results, GLAS footprints were divided into two slope categories: GLAS footprints that fall on slopes lower than five degrees and GLAS footprints that fall on slopes higher than five degrees. Because the slopes are relatively weak in French Guiana, model validation for

high slopes was not possible. Model validation over these two slope categories showed that the RMSE on the estimation of canopy heights slightly increased from 3.3 to 4.0 m and from 3.5 to 4.8 m for PCA and the linear regression model, respectively (models 7bis and 19, Table 3.1). However, the PCA model is slightly better at correcting the effects of the slopes in comparison to the linear regression model with a 0.7 m increase in the RMSE *vs.* 1.3 m for the linear model.

Forest landscape classes in French Guiana were defined in a previous study carried out by Gond *et al.* [77]. Gond *et al.* [77] interpreted 33 remotely sensed landscape types (LTs) using VEGETATION/SPOT images. Five of the 33 classes occupied 78% of the forests in the area. The method utilized in their study used a multivariate analysis of remote sensing data, field observations and environmental data. The defined LTs were LT8, LT9, LT10, LT11 and LT12. Model application over the different LTs showed that the RMSE on the estimation of canopy is consistent across the four LTs (LT8 to LT11). The RMSE ranged between 2.8 and 3.6 m for the PCA model (model 19, Table 3.1), and between 3.5 and 3.9 m for the linear regression model (model 7bis, Table 3.1).

3.3.5 Error on the estimation of biomass

The objective of this section is to analyze the impact of the canopy height estimation precision on the Above ground Carbon Density (ACD) and Above Ground Biomass (AGB) estimation precision. Asner *et al.* [48] proposed a general plot aggregate allometry in order to estimate the above ground carbon density (ACD):

$$ACD = aH^{\alpha}.BA^{\beta}.W_D^{\gamma} \quad (3.11)$$

Were H is the LiDAR derived top-of-canopy height, BA the basal area and W_D the wood density. Moreover, Asner *et al.* [40] showed that the basal area (BA) and the wood density (W_D) were dependent on the LiDAR derived top-of-canopy height for all the studied tropical forests (Hawaii, Madagascar, Peru, Panama, and Colombia). Hence, according to their study the previous allometric relation could be written as:

$$\mathbf{ACD = aH^{\alpha}(b.H)^{\beta}(c+d.H)^{\gamma}} \quad (3.12)$$

The relationship between the precision on the estimation of canopy heights and the precision on the estimation of ACD and AGB can be written as:

$$\frac{\Delta AGB}{AGB} = \frac{\Delta ACD}{ACD} = \left[(\alpha + \beta) + \left(\frac{d.\gamma}{d + \frac{c}{H}}\right)\right]\frac{\Delta H}{H} \quad (3.13)$$

Where ΔAGB/AGB is the relative precision on the estimation of above ground biomass, ΔACD/ACD is the relative error on the estimation of the above ground carbon density. The coefficients α, β, and γ were estimated by Anser *et al.* [40] using 754 field plots across five tropical countries (Hawaii, Madagascar, Peru, Panama, and Colombia) and many vegetation types. The coefficients c and d were also estimated by Anser *et al.* [40] for different regional forests. In our analysis of the canopy height estimation precision impact on the AGB and ACD estimation precision, the chosen coefficients c and d were those estimated from the moist Colombian forest [40]. These coefficients were chosen due to the fact that the French Guiana's forest is a moist tropical forest and close in location to the Colombian forest. Finally, accuracy on the estimation of canopy heights of 3.6 m will lead to a relative error on the estimation of the ACD and AGB of about 14.1% (for a mean canopy height of 30 m). The United Nations Program on Reducing Emissions from Deforestation and forest Degradation (REDD) recommends biomass errors within 20 Mg/ha or 20% of field estimates for evaluating forest carbon stocks, but should not exceed errors of 50 Mg/ha for a global biomass map at a resolution of 1 ha ([94]; [95]). Finally, in the case of high relief, where the precision on the estimation of canopy height exceeds 5 m, the precision on the estimation of biomass will be at best 20%.

3.4 Discussion

Our findings regarding the strong correlation between the waveform extent and the in-situ canopy heights are in accordance with the studies of Lefsky *et al.* [44]; Hilbert and Schmullius [62]; and Baghdadi *et al.* [72]. They found that this metric is one of the most important metrics used in canopy height estimation models. However, waveform extent is not the sole metric used for canopy height estimation, as it can be affected by external sources such as terrain relief. Thus, in order to obtain more precise canopy height

estimation results, additional metrics are required. Previous studies developed metrics such as the trail, lead, and the terrain index (TI) in order to increase the canopy height estimation precision. The TI index was first developed by Lefsky *et al.* [62] and the lead and trail were first introduced in Lefsky *et al.* [78]. These metrics were later used in many other studies like Hilbert *et al.* [62]; Pang *et al.* [64]; Chen *et al.* [71] and Baghdadi *et al.* [72]. These metrics which were mainly used for the correction of the slope, proved to be very useful, as they increased significantly the precision on the estimation of canopy height models ([44]; [71]; [62]). Moreover, the waveform extent and the trail metrics proved to be very successful in estimating canopy heights even in low relief areas like our study site. Indeed, the linear regression models which used the waveform extent and the trail metric showed a decrease in RMSE of at least 3.9 m in comparison to the direct method (for example, RMSE reaches 3.7 m in using the linear model with Wext, Trail and TI in comparison to RMSE of 7.9 m for the direct method). In contrast, the contribution of the lead in the canopy height estimation models seemed to be weak in this study. Similar findings were noted in the study of Baghdadi *et al.* [72] which also estimated canopy heights over flat terrain.
Our results also demonstrated that canopy height estimation using random forest regressions is better in comparison to the linear models, even when using the same metrics. Indeed, the random forest model which uses only the waveform extent and the terrain index (TI) showed a 1.3 m decrease in RMSE in comparison to the linear model which uses the same metrics. This is probably due to the fact that the relation between the GLAS metrics and canopy heights is not strictly linear.

The metric based estimation methods applied in this study include some potential error sources. These error sources are related to the precision of the extracted GLAS metrics especially metrics extracted using vegetation or ground information such as the lead and trail. Indeed, over dense vegetated areas, the precision on the localization of the ground peak decreases significantly, and this will lead to lower precisions on the estimation of the trail metric and ultimately on the canopy height estimation. To solve this issue, another technique used in this study for canopy height estimation was the principal component analysis (PCA) of the waveform. This technique does not require metrics to be extracted from the GLAS waveform in order to estimate canopy heights, as it works using the principal components of the raw LiDAR waveforms. The results of the PCA based models for canopy height estimation showed promising results when estimating canopy heights

using either linear regressions or random forest regressions, with an RMSE of 5.9 and 4.7 m for the linear regressions and RF models, respectively. In addition, adding the waveform extent metric to these models showed slightly better estimation results in comparison to the metric based methods, with an RMSE ranging between 3.8 to 4.4 m for the linear regression models and around 3.6 m for the random forest models.

Other sources of error on the estimation of canopy heights are terrain slopes. Indeed over sloping areas, canopy height estimation precision decreases with the increase of the slope ([63]; [72]; [71]). In our study area of low relief, an increase in RMSE of 0.7 m for the best PCA model and 1.3 m on for the best metric based model were noted in 5° to 10° slope areas in comparison to flat areas (0 to 5° slopes). However, over higher slopes (> 10°), the error on the estimation of canopy heights is expected to be higher. In this study, the SRTM 90 m DEM was used, which is the only available DEM over large areas. The future availability of finer DEMs such as the SRTM 30 m or the TanDEM-X 12 m might improve the estimation of canopy heights.

Results showed that the canopy height estimation error using ICESat/GLAS (RMSE about 3.6 m in this study) leads to a relative error on the estimation of aboveground biomass of about 20%. This relative error will increase to more than 34% for canopy height estimation precision of 5 m or higher. Thus, the United Nations Program on Reducing Emissions from Deforestation and forest Degradation (REDD) recommendations may not be satisfied over forested areas with steep slopes because the canopy height estimation precision will be higher than those estimated in this study.

3.5 Conclusions

Estimating forest canopy height from large-footprint satellite LiDAR waveforms is challenging given the complex interaction between LiDAR waveforms, terrain, and vegetation, especially in dense tropical and equatorial forests. In this Chapter, canopy height in French Guiana was estimated using multiple linear regression models and the Random Forest technique (RF). This analysis was either based on LiDAR waveform metrics extracted from the GLAS (Geoscience Laser Altimeter System) spaceborne LiDAR data and terrain information derived from the SRTM (Shuttle Radar Topography

Mission) DEM (Digital Elevation Model) or on Principal Component Analysis (PCA) of GLAS waveforms. Results show that the best statistical model for estimating forest height based on waveform metrics and digital elevation data is a linear regression of waveform extent, trailing edge extent, and terrain index (RMSE of 3.7 m). For the PCA based models, better canopy height estimation results were observed using a regression model that incorporated both the first 13 principal components (PCs) and the waveform extent (RMSE = 3.8 m). Random Forest regressions revealed that the best configuration for canopy height estimation used all the following metrics: waveform extent, leading edge, trailing edge, and terrain index (RMSE = 3.4 m). Waveform extent was the variable that best explained canopy height, with an importance factor almost three times higher than those for the other three metrics (leading edge, trailing edge, and terrain index). Furthermore, the Random Forest regression incorporating the first 13 PCs and the waveform extent had slightly-improved canopy height estimation in comparison to the linear model, with an RMSE of 3.6 m. In conclusion, multiple linear regressions and RF regressions provided canopy height estimations with similar precision using either LiDAR metrics or PCs. However, a regression model (linear regression or RF) based on the PCA of waveform samples with waveform extent information is an interesting alternative for canopy height estimation as it does not require several metrics that are difficult to derive from GLAS waveforms in dense forests, such as those in French Guiana. Nevertheless, such approach was only applied on a terrain with weak slopes (<5^{o}), and should be tested over sloping areas to test its effectiveness.

CHAPTER 4:

FOREST CANOPY HEIGHT MAPPING OVER FRENCH GUIANA USING SPACE AND AIRBORNE LIDAR DATA

4.1 Introduction

In this chapter, airborne and spaceborne LiDAR canopy height estimates in combination to ancillary data were used to create a canopy height map covering the entire French Guiana. The LiDAR canopy height datasets used were the airborne LiDAR canopy height estimates covering 4/5 of French Guiana, as well as GLAS canopy height estimates. For the ancillary datasets, globally available datasets of vegetation indices, precipitation, terrain indices, geological, and forest landscape types were selected. The vegetation indices were derived from the Enhanced Vegetation Index (EVI) product. The precipitation data were provided from the NASA Tropical Rainfall Measuring Mission (TRMM). The terrain indices were derived from the Shuttle Radar Topography Mission (SRTM) data. The geological and the forest landscape type maps were provided by Delor *et al.* [76] and Gond *et al.* [77] respectively.

For the creation of the wall-to-wall canopy height maps, a two-step procedure was implemented using well established techniques. First, the ancillary data were modelled to the reference datasets using the Random Forest (RF) regression. Next, the regression-kriging technique was used, first in order to krige the canopy height residuals (reference canopy heights – RF estimated canopy heights) and then to add the results to the wall-to-

wall maps obtained from the RF regressions. Finally, all the created maps were validated using an independent dataset of airborne LiDAR canopy heights estimates. A description of the methodology used for the creation of the wall-to-wall canopy height maps is given in Section 4.2, followed in Section 4.3 by the results. Finally, the discussion is presented in section 4.4 and the conclusions in section 4.5.

4.2 Materials and methods

In order to estimate canopy heights at un-sampled locations by LiDAR data (GLAS or LD datasets) a two-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 and ordinary kriging (e.g. [66]; [65]; [96]; [97]). The LD and GLAS datasets will not be used conjointly in the canopy height mapping procedure (a map using each dataset will be produced). This is due to two main reasons: (1) To be able to compare the precision of the maps obtained by using airborne LiDAR to those obtained using spaceborne LiDAR; (2) The LD dataset is denser, and more precise in comparison to the GLAS dataset, therefore, adding the GLAS dataset to the LD dataset will not improve the precision of the obtained map. Finally, for the validation of the models, the canopy heights estimated from the transverse lines from LD datasets (about 7% of the LD dataset, named LD_val), and the canopy height estimates from the HD dataset will be used. The remainder of the LD canopy height estimates (93% of the LD dataset, named LD_cal), and the canopy height estimates from the GLAS dataset will be used for model building. The datasets for model building will be used separately.

4.2.1 Canopy height mapping using regression-kriging

We mapped canopy height from LiDAR data at a 250 m resolution using the regression-kriging technique (RK). 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) ([98]). 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 ([99]) thus preventing the stationarity of the studied autocorrelated variables, like the RF residuals

that are centred at 0. Finally, regression kriging or RK, is the technique which fits separately the trend and the residuals and then sum them ([100]; [101]). RK can be expressed as follows:

$$\hat{z}(\boldsymbol{s_0}) = \hat{\boldsymbol{m}}(\boldsymbol{s_0}) + \hat{e}(\boldsymbol{s_0}) = \hat{\boldsymbol{m}}(\boldsymbol{s_0}) + \sum_{i=1}^{n} \boldsymbol{\lambda_i}.\boldsymbol{e}(\boldsymbol{s_i}) \quad (4.1)$$

Where $\hat{m}(s_0)$ is the fitted trend, $\hat{e}(s_0)$ is the kriged residual, λi are the kriging weights determined by the spatial dependence structure of the residual and $e(s_i)$ is the residual at location s_i.

4.2.2 Canopy height trend mapping using Random Forest regressions

For the $\hat{m}(s_0)$ component from equation 4.1 we used the Random Forest regression technique (RF). RF which was developed by Breiman [93] 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. [65]; [102]), and biomass estimation ([32]; [96]; [97]). The main advantages of random forest are its incorporation of continuous or qualitative predictors without making assumptions about their statistical distribution or covariance structure [93].

First the 12 predictors described in section 2.2.3 were used in the RF model in an attempt to explain the canopy height estimates from either the LD_cal dataset or the GLAS dataset (Figure 4.1). Next, to select the best predictors for explaining canopy height, the algorithm of Genuer *et al.* [103] which is a two-step procedure was used. The first step is to sort the variables based on their initial scores of importance, and remove the variables of 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 consists of building the nested collection of RF models involving the k variables (k=1 to m where m is the number of remaining variables with the highest score of importance), and selects the variables in the model giving the smallest out-of-bag error. After selecting the best variables, the RF model was calibrated, and applied to all of the study area in order to create a wall-to-wall

canopy heights trend map (Figure 4.1). 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 ([93]). The final prediction decision is based on a voting system of all the predictions from the decision trees that have been created.

4.2.3 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 some sort of spatial correlation between the canopy heights. Thus, a spatial prediction model is required when data are spatially dependent. In this study we used the ordinary kriging (OK, Figure 4.1) 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 the semivariance γ as a function of the distance between samples *h* using the following function:

$$\boldsymbol{\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(S_i) - z(S_i + h)]^2} \tag{4.2}$$

Where $\gamma(h)$ is the semivariance as a function of the lag distance h, $N(h)$ is the number of pairs of data separated by h, and z is the estimated canopy height at locations $\boldsymbol{u}_\alpha$ and $(\boldsymbol{u}_\alpha+h)$ ([104]). 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; (3) the *range* is the distance at which the *sill* is reached. After plotting the sample semivariogram which 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. After model fitting of the sample semivariogram, ordinary krigging is then used, which estimates values Z^* at location u using the following equation:

$$Z^*(S_0) = \sum_{i=1}^{n} \lambda_i z(S_i) \quad (4.3)$$

Where z is the data, in this case the canopy height at location S_i, λ_i are the weights of n neighbouring samples [104]. The number of n neighbouring points is user defined. The weights λ_i depends on the fitted semivariogram function, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location.

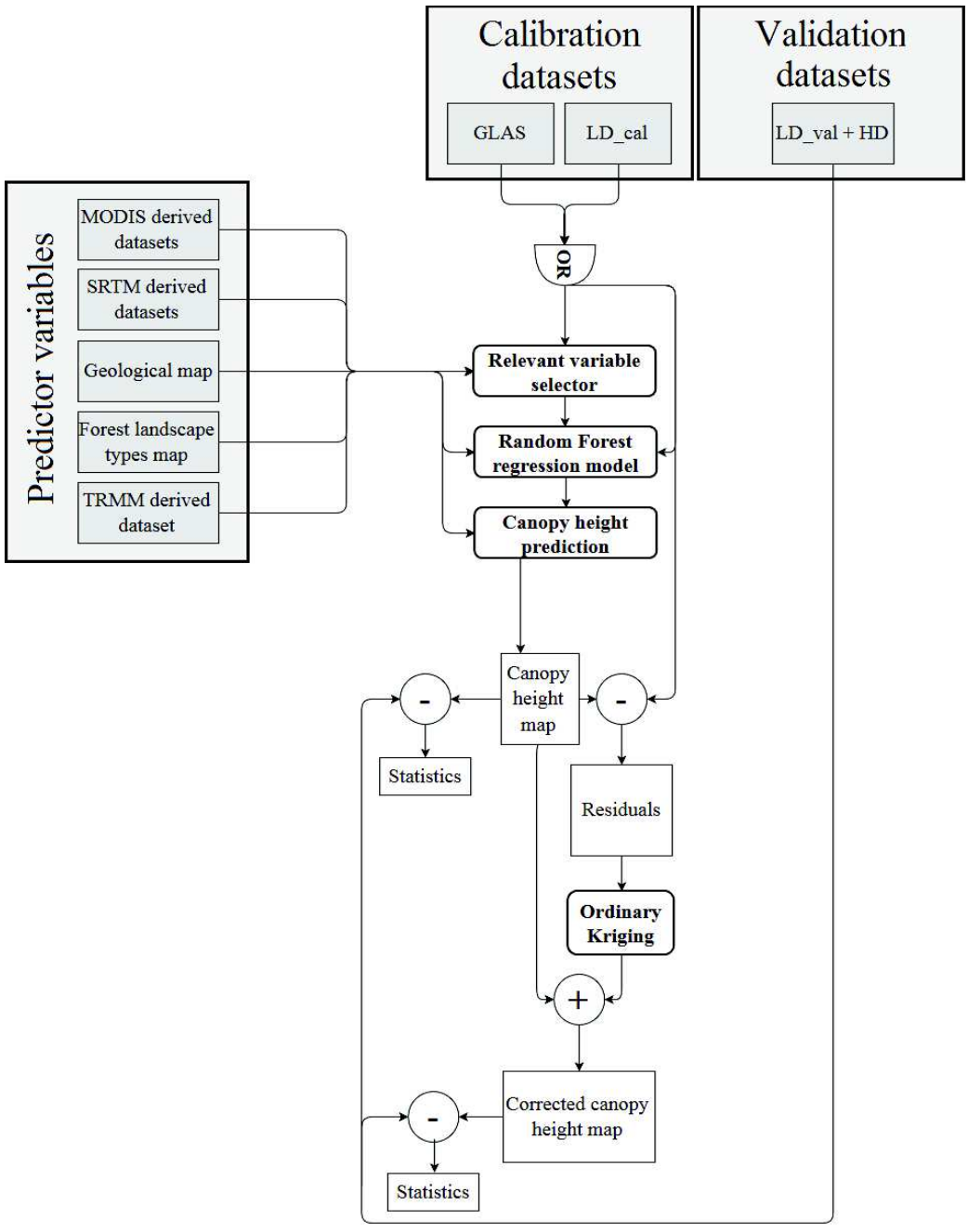
4.2.4 Effects of LiDAR sampling density on precision of the mapped canopy heights.

The purpose of the current analysis was to measure how the canopy height maps accuracy are 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 spacings 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. The canopy height maps were created using the same procedure described in sections 3.1, and 3.2 which consists at 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, 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 subet by comparing these canopy height maps to the validation datasets (HD and LD_val).

Figure 4.1. The procedure used in order to create a wall-to-wall map of canopy heights over French Guiana.



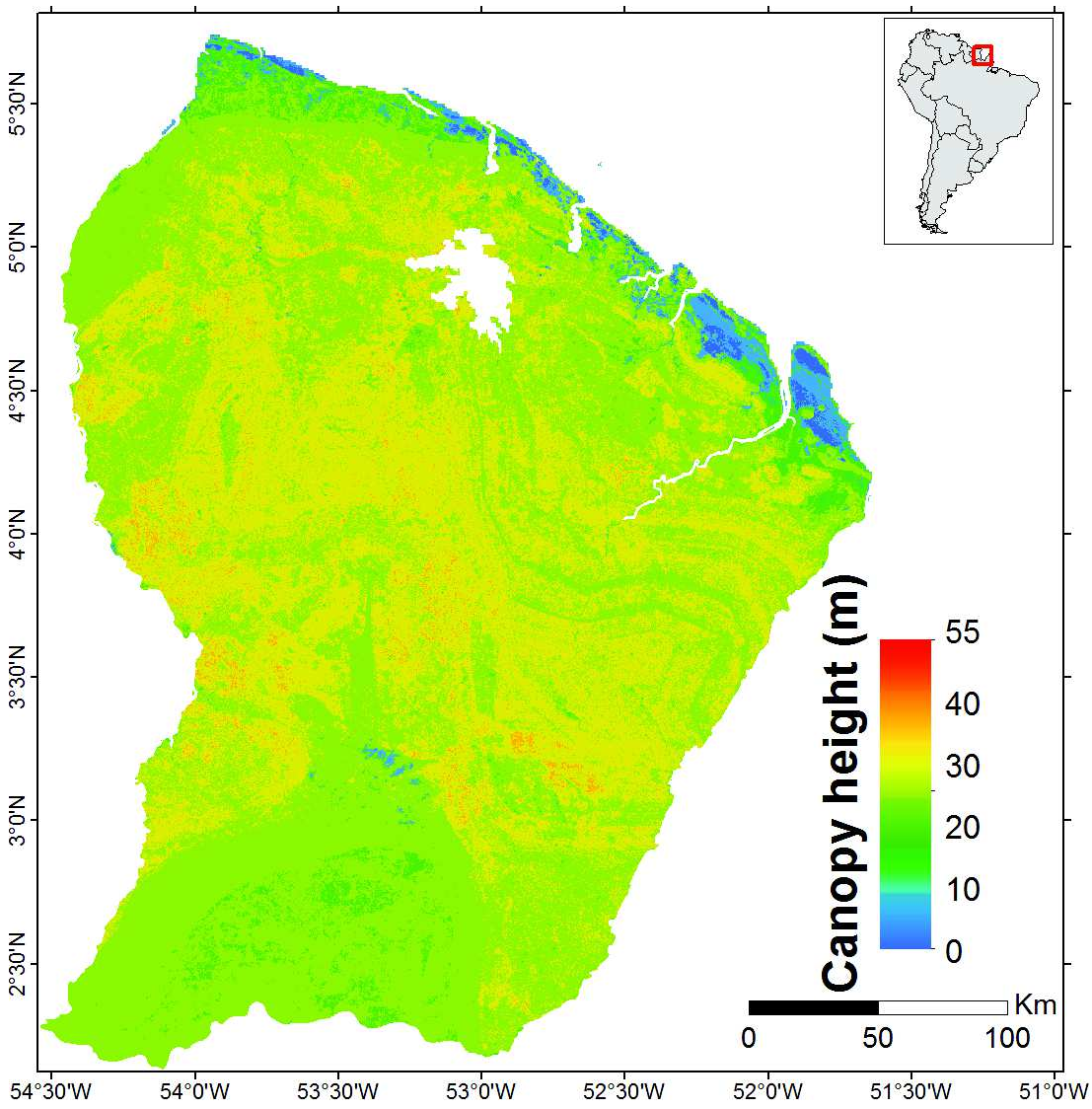
4.3 Results

4.3.1 Canopy height 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 2.2). 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 [105].

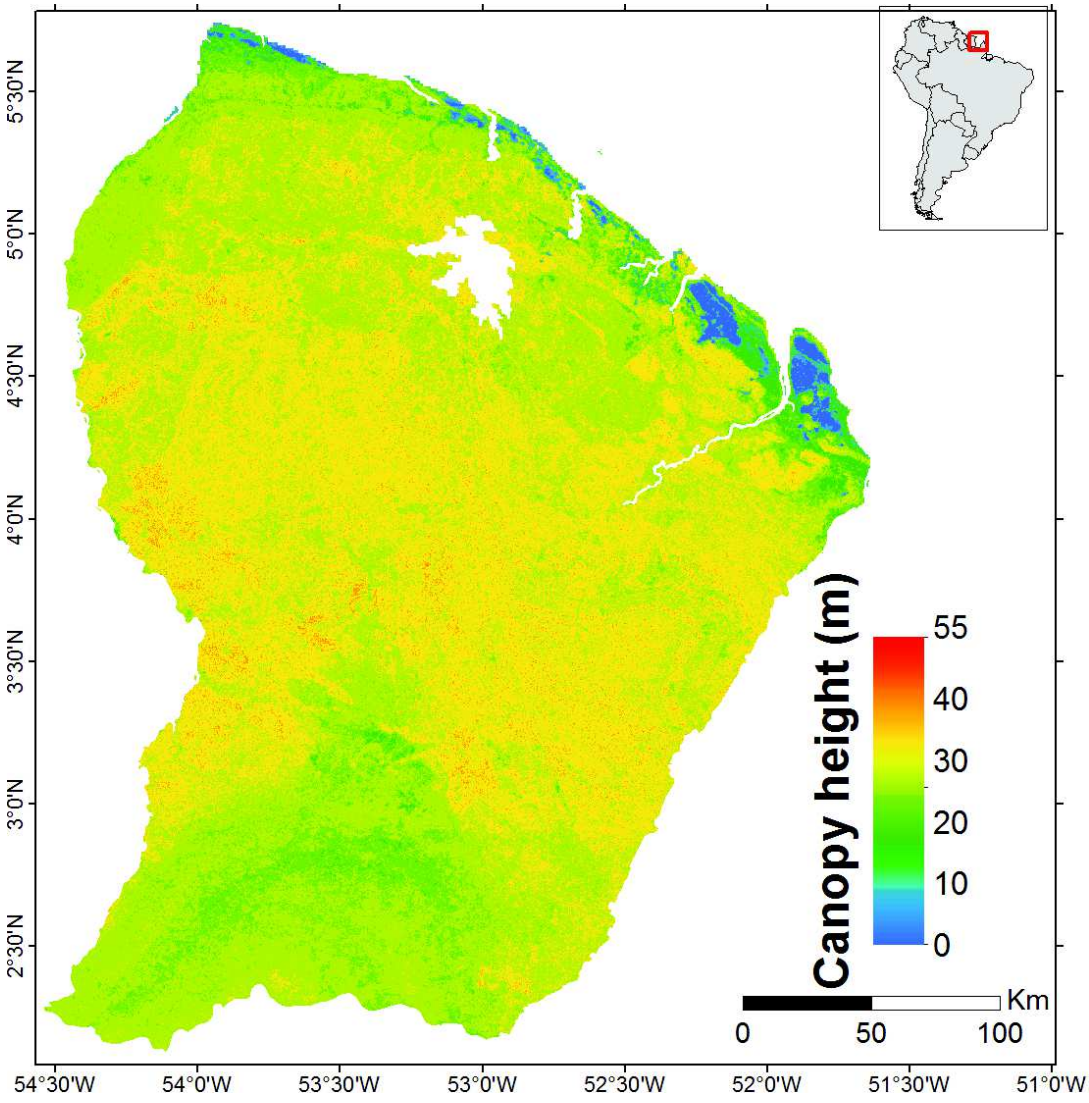
After creating the Random Forest regression models using the twelve predictors, only the predictors that best explained the canopy heights were selected using the variable selector test. Results showed that the best predictors were the same for both calibration datasets. The best variables according 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 4.2). 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 4.3, Table 4.1). 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 (validation canopy heights – estimated canopy heights)) for both the GLAS and LD_cal datasets was very low (< 0.2m).

Figure 4.2. Wall-to-wall map of French Guiana with Random Forest regressions using as reference data the canopy height estimates from: (a) GLAS dataset; (b) LD_cal dataset.



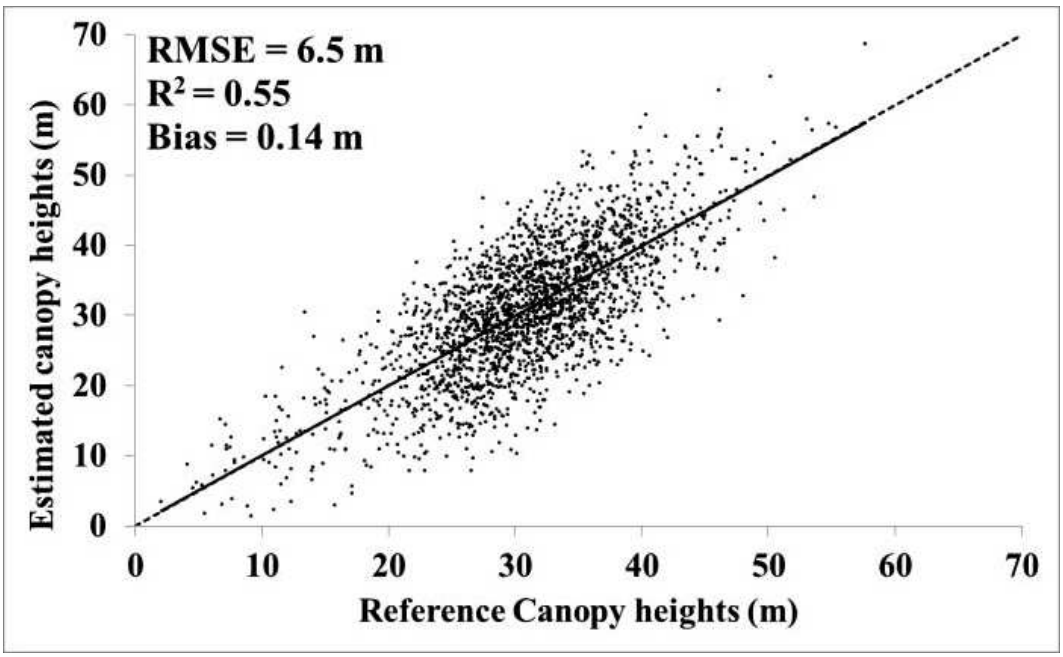
(a)

Figure 4.2 (cont.)

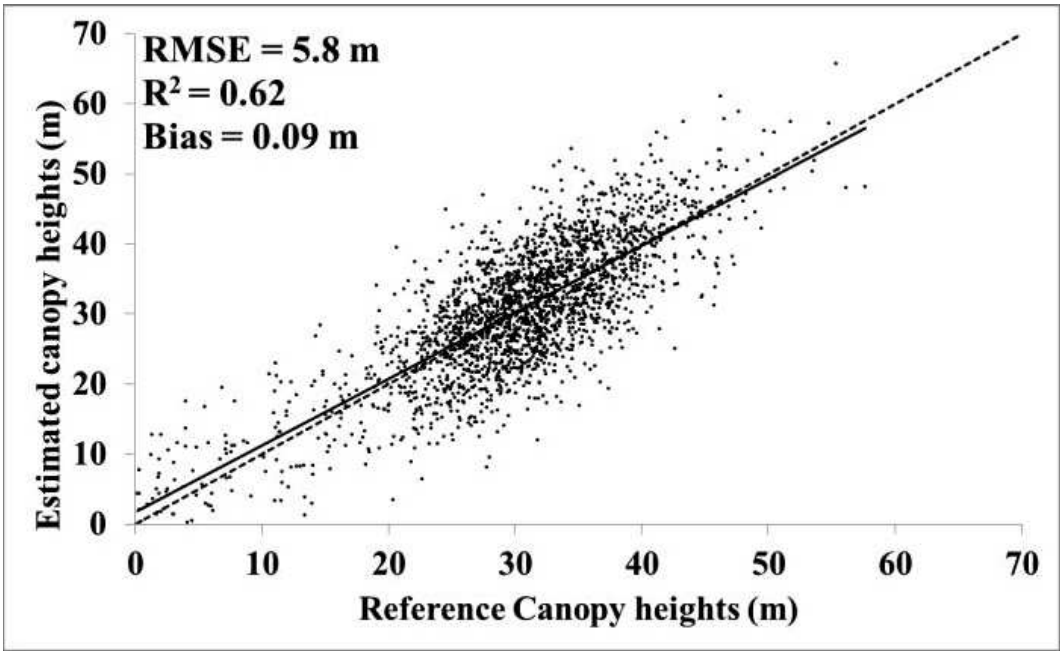


(b)

Figure 4.3. Comparison between the reference canopy heights of the validation datasets and the canopy height estimates using Random Forest regressions: (a) GLAS dataset; (b) LD_cal dataset.



(a)



(b)

4.3.2 Canopy height estimation using regression-kriging

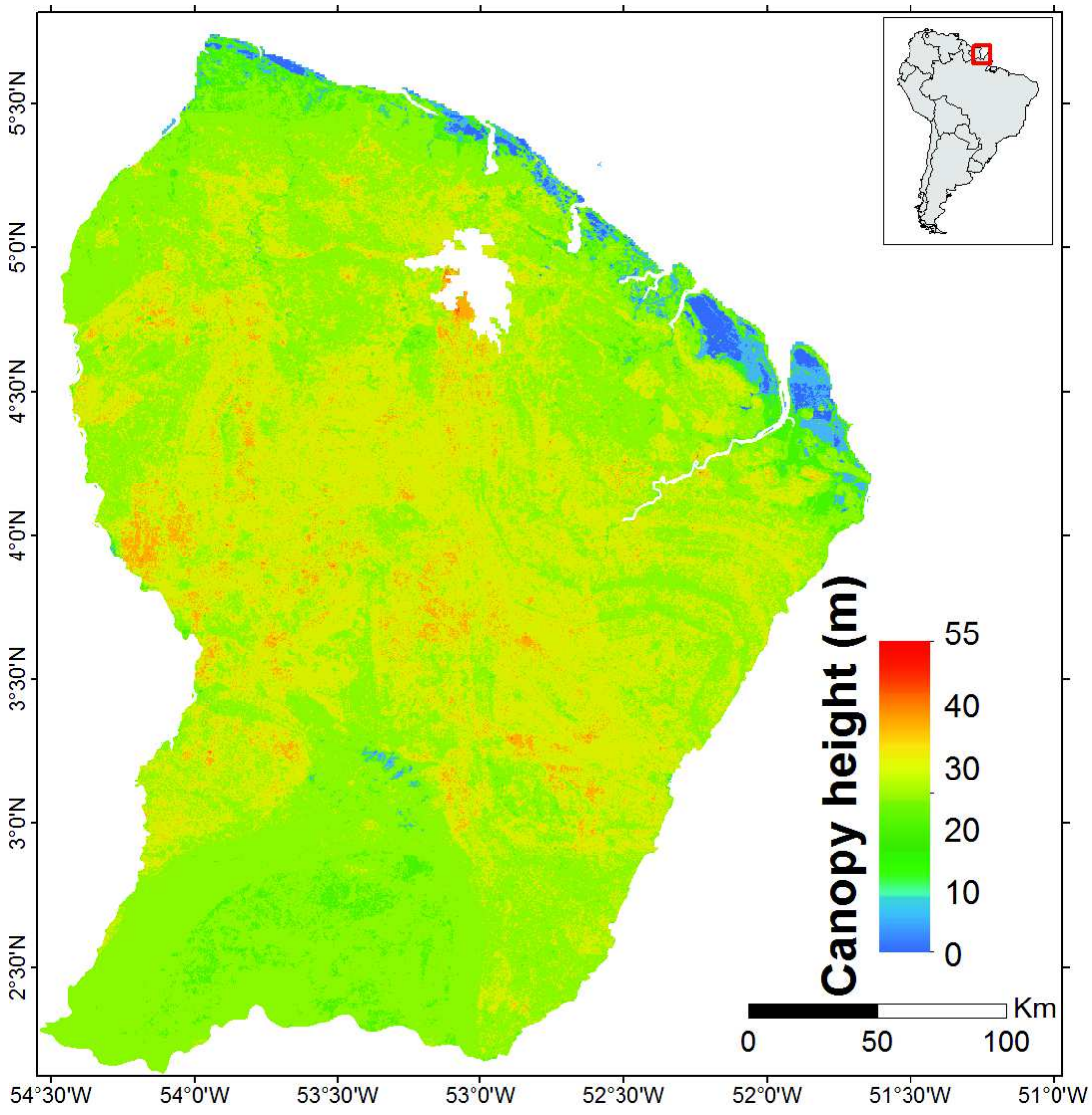
After creating the wall-to-wall maps using the Random Forest regression models, the canopy height residuals (reference canopy height – estimated canopy height) were kriged from each model in an attempt to increase the precision of the canopy height estimates using the Random Forest regressions for both 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:

$$\boldsymbol{\gamma(h) = S^2 + \sigma^2 \left[1 - exp\left(\frac{-h}{a}\right)\right]} \quad (4.4)$$

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, sills, and ranges, which ranged respectively between 15 and 18 m^2, 28 and 32 m^2 and 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 4.4). These maps were then validated using the validation datasets (LD_val and HD) (Figure 4.5, Table 4.1). 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 4.6). For the GLAS dataset, the standard deviation of the canopy height estimates uncertainty ranged between 4 and 7 m (Figure 4.6a). In addition, the standard deviation values appear to be lower near the location of the

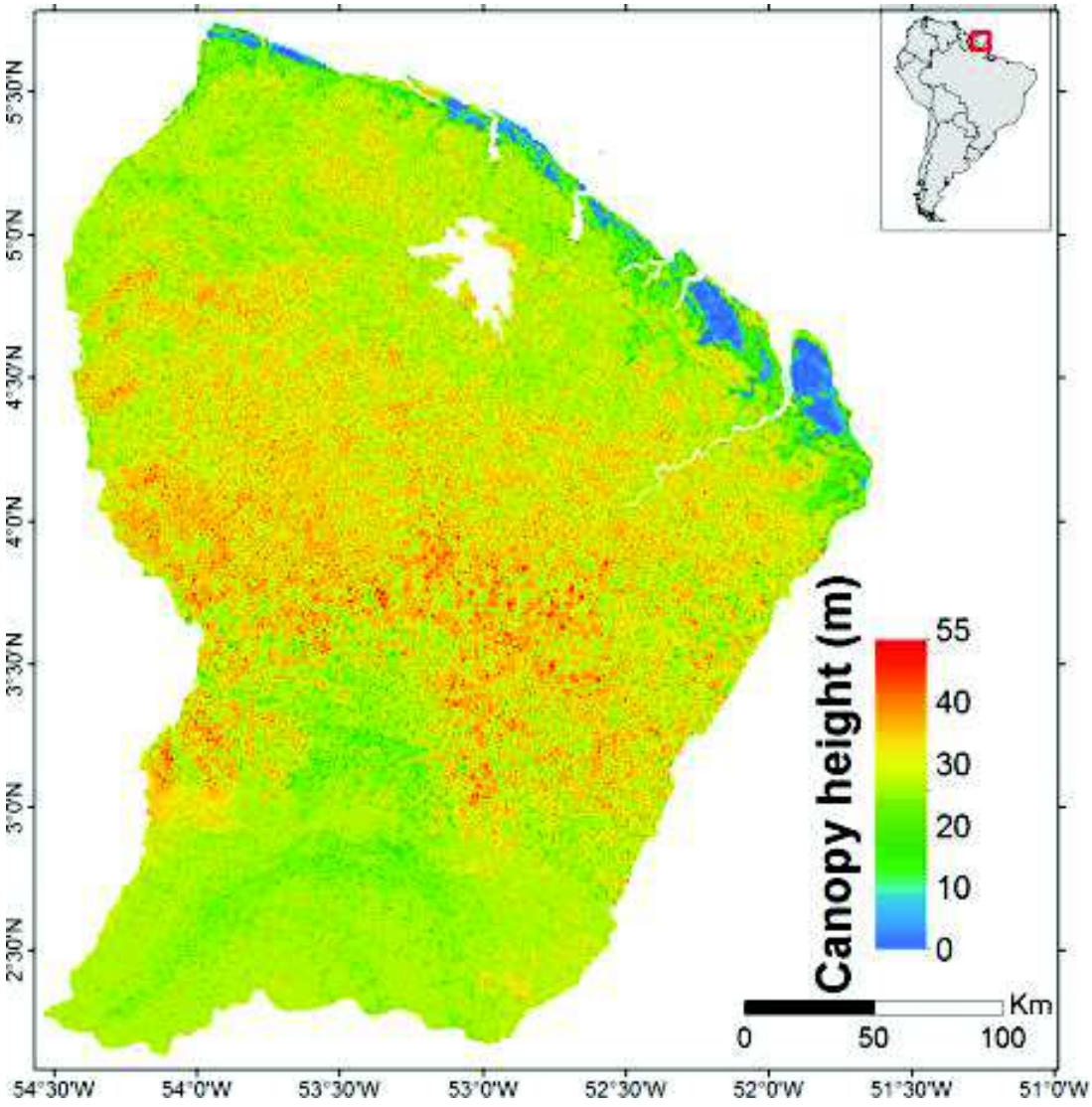
GLAS canopy height estimates, and increases with increasing distance until they reach 7 m. Similar results appear for the LD_cal dataset (Figure 4.6b). With lower standard deviations in areas with denser LiDAR acquisitions (i.e. north of French Guiana) and higher standard deviations with sparser LiDAR acquisitions (i.e. center of French Guiana). However, due to the generally denser dataset in comparison to the GLAS dataset, the standard deviation of the canopy height estimates uncertainty ranged between 1 and 4 m (Figure 4.6b).

Figure 4.4. Wall-to-wall map of French Guiana with Random Forest regressions and residual kriging using as reference data the canopy height estimates from: (a) GLAS dataset; (b) LD dataset.



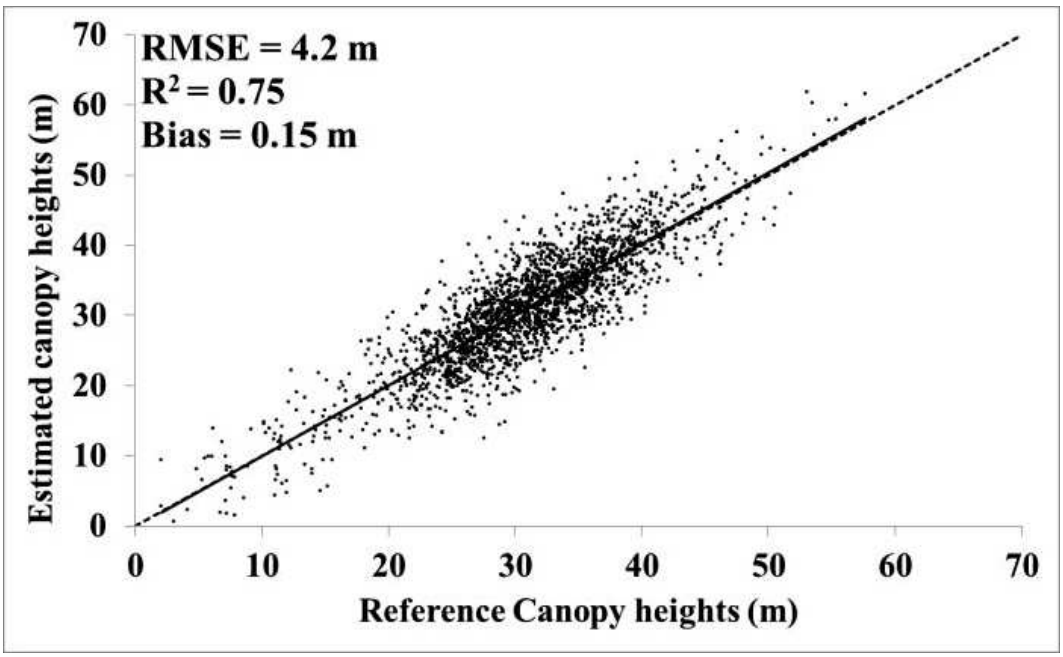
(a)

Figure 4.4 (cont.)

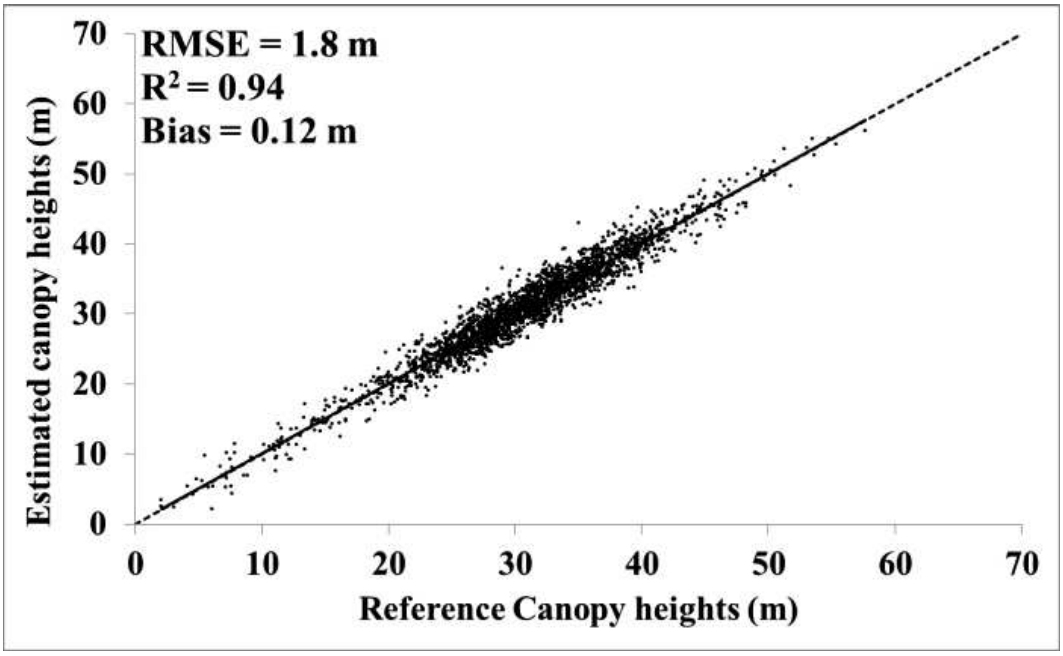


(b)

Figure 4.5. Comparison between the reference canopy heights of the validation datasets and the canopy height estimates using Random Forest regressions and residual-kriging: (a) GLAS dataset; (b) LD_cal dataset.

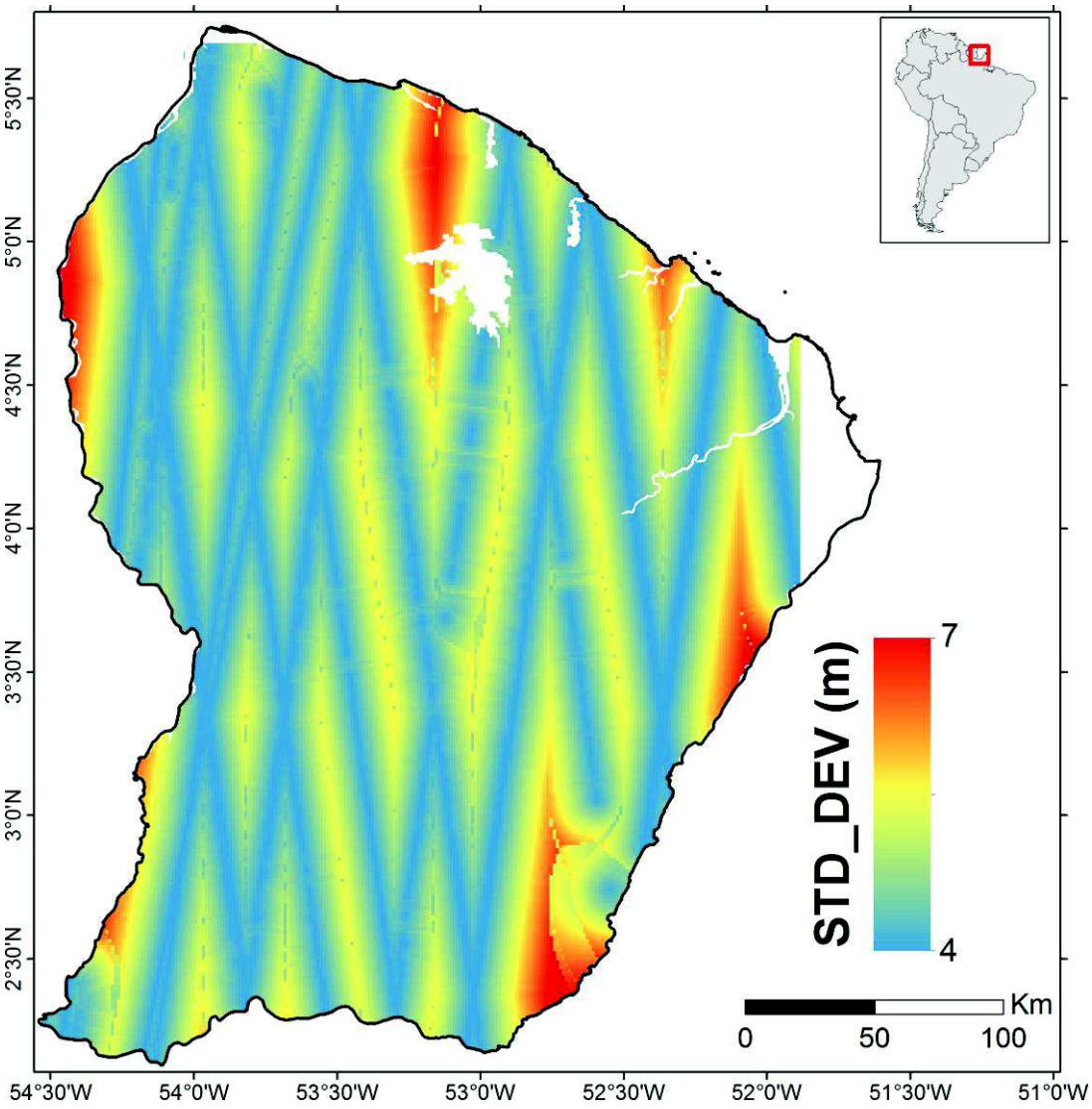


(a)



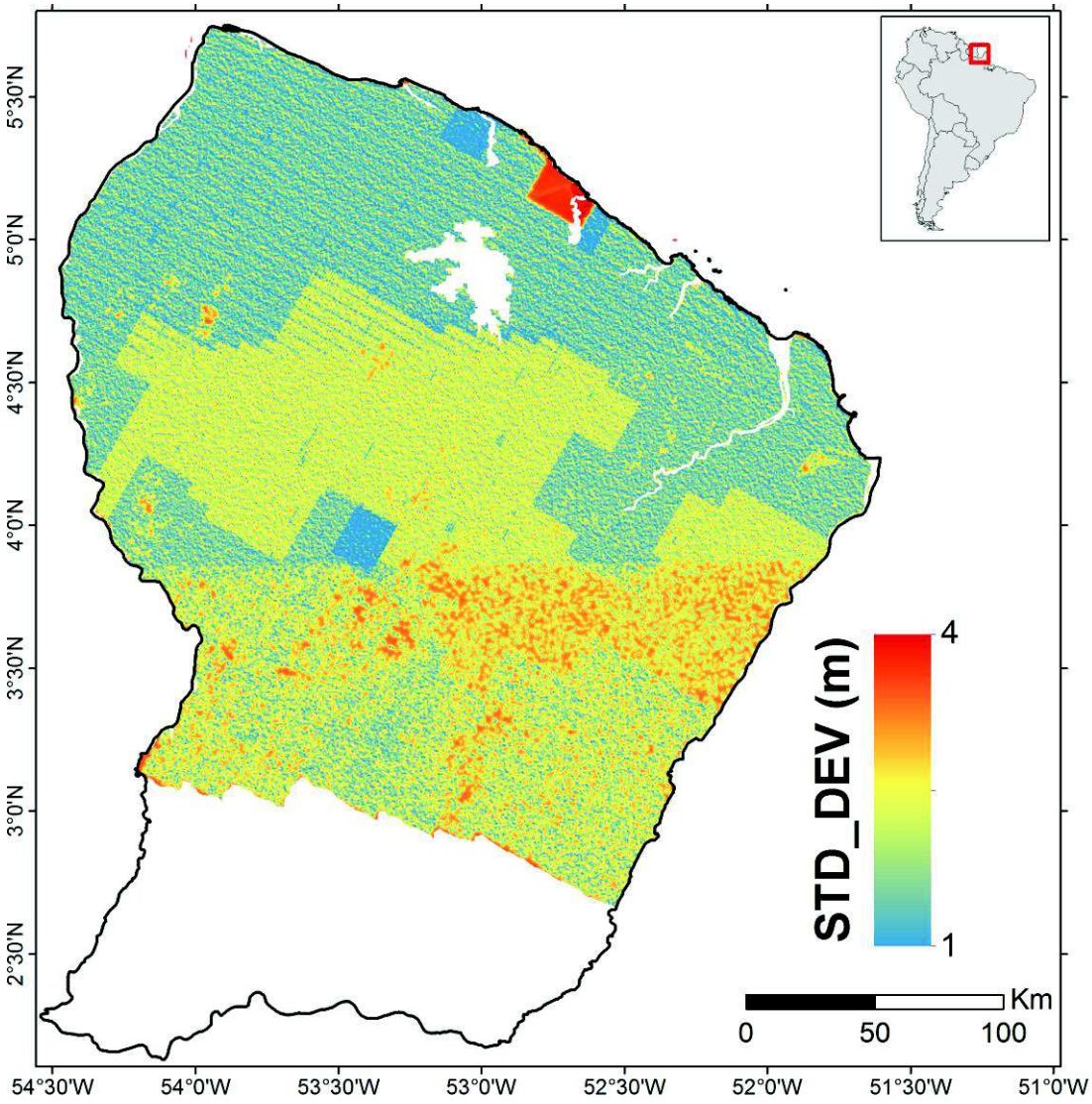
(b)

Figure 4.6. Wall-to-wall standard deviation map (STD_DEV) of the the canopy height estimates uncertainty for: (a) GLAS dataset; (b) LD_cal dataset.



(a)

Figure 4.6 (cont.)



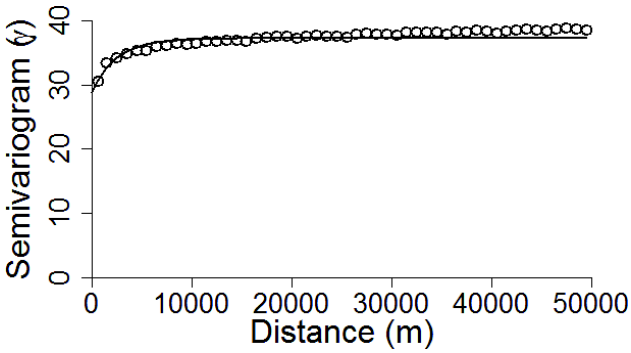
(b)

4.3.3 Relationship between LiDAR flight lines spacing and the accuracy on the kriged canopy height

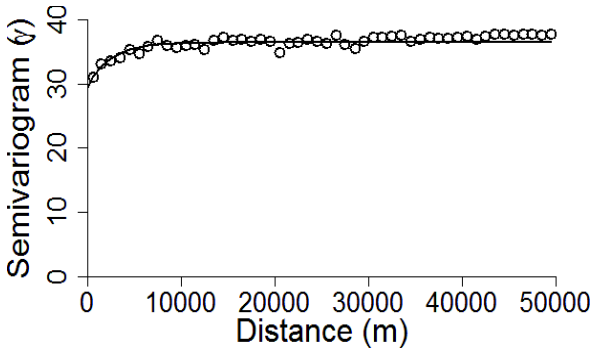
The analysis performed in the previous section showed a significant improvement of the canopy height estimation precision when adding the kriged residuals. This improvement was observed for the two calibration datasets. In the case of the LD_cal dataset, the improvement was the highest from 5.8 m without the height residual kriging to 1.8 m with the residual kriging. This is due to the high density of the 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 the 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 with different LiDAR densities was assessed. The purpose was to analyse 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 4.1). 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 nuggets (~30 m^2) sill (~8 m^2), range (~4500 m) were obtained as those from the canopy height residuals from the GLAS and LD_cal datasets (Figure 4.7). When adding the kriged residuals corresponding to each of the LD_cal subsets (Figure 4.8), the precision on the canopy height estimate maps increased as expected (Table 4.1). This increase in the precision on the canopy height estimation was found to be negatively correlated with the LiDAR flight lines spacing of the LD subsets. For the LD_5 and LD_10 subsets, the precision on the canopy height estimates were 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.

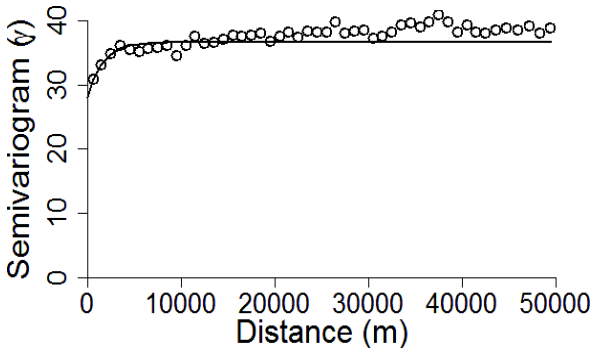
Figure 4.7. Examples of fitted semivariograms of the canopy height residuals from: (a) LD_5; (b); LD_20; (c) LD_50.



(a)



(b)



(c)

Figure 4.8. Examples of wall-to-wall maps of French Guiana with Random Forest regressions and residual kriging using as reference data the canopy height estimates from: (a) LD_5; (b) LD_20; (c) LD_50.

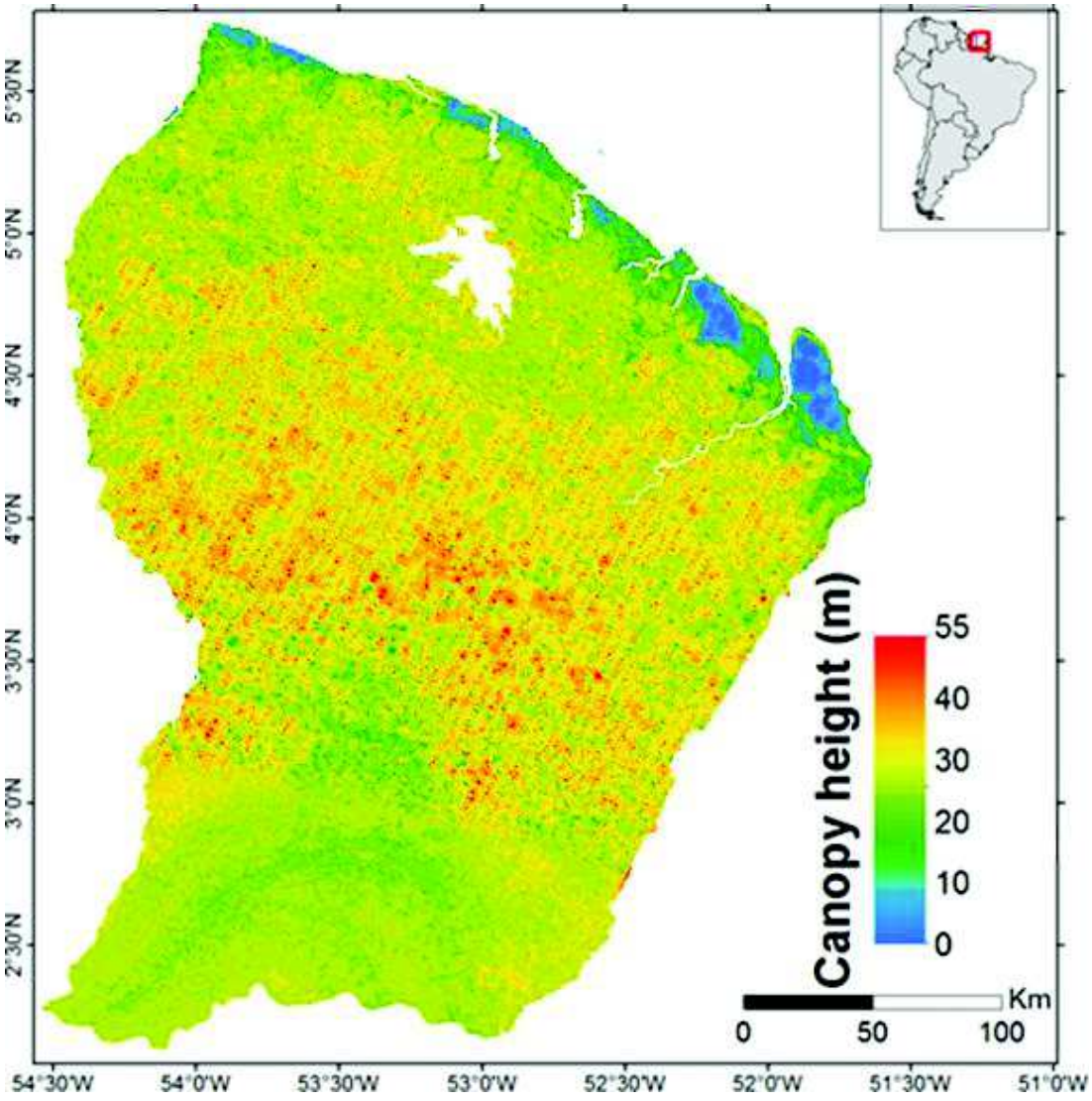
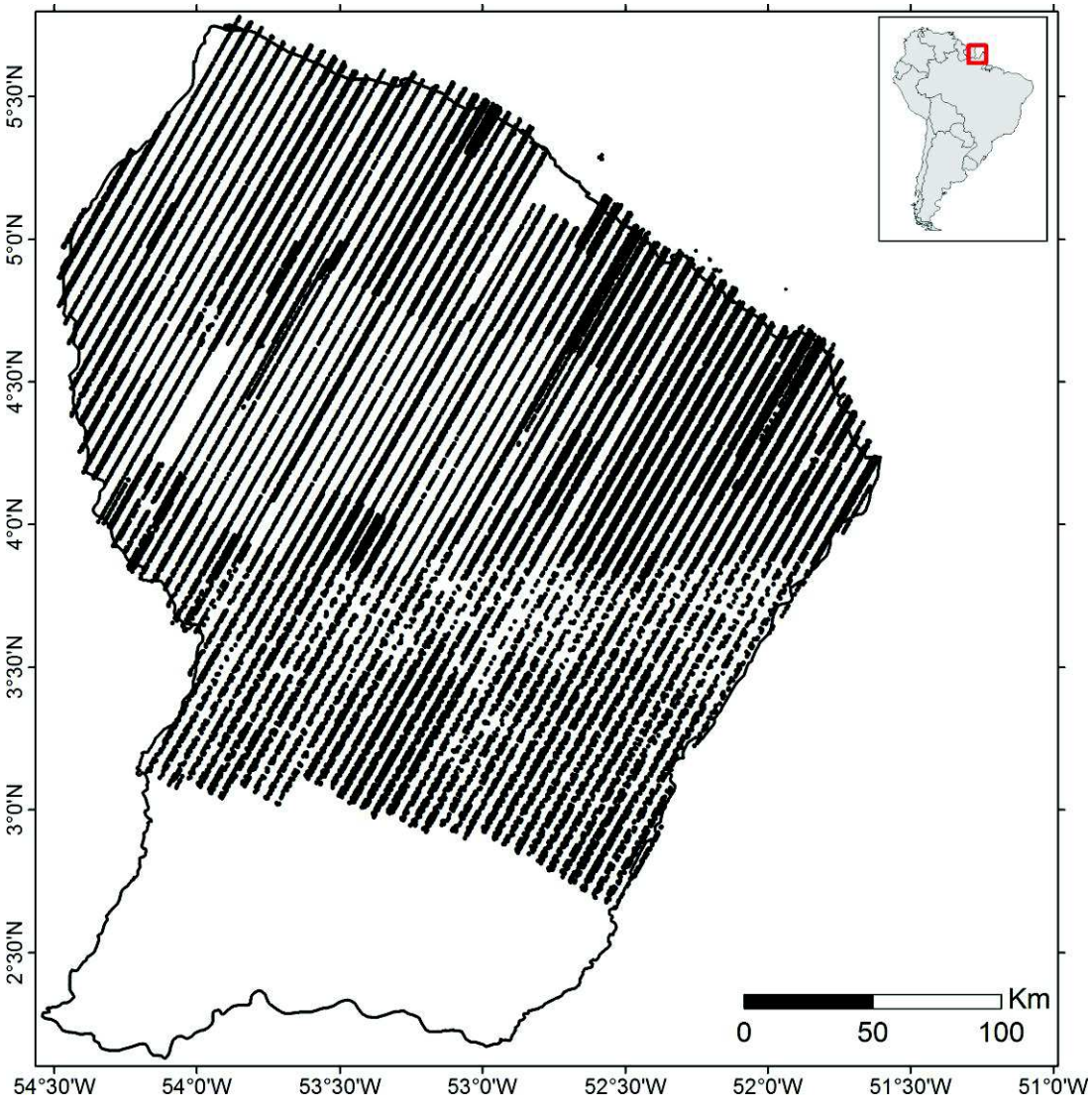


Figure 4.8 (cont.)



(a)

Figure 4.8 (cont.)

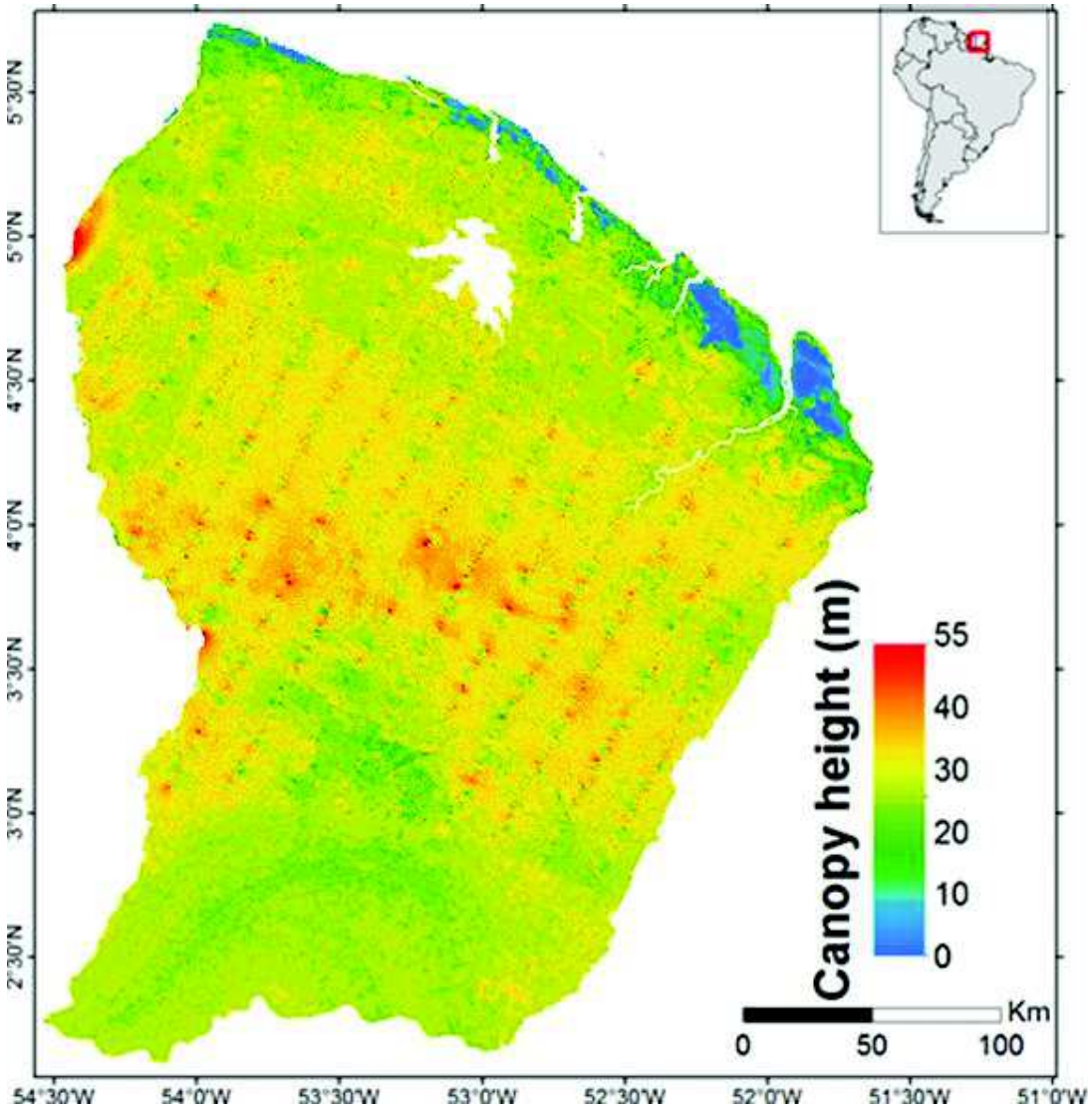
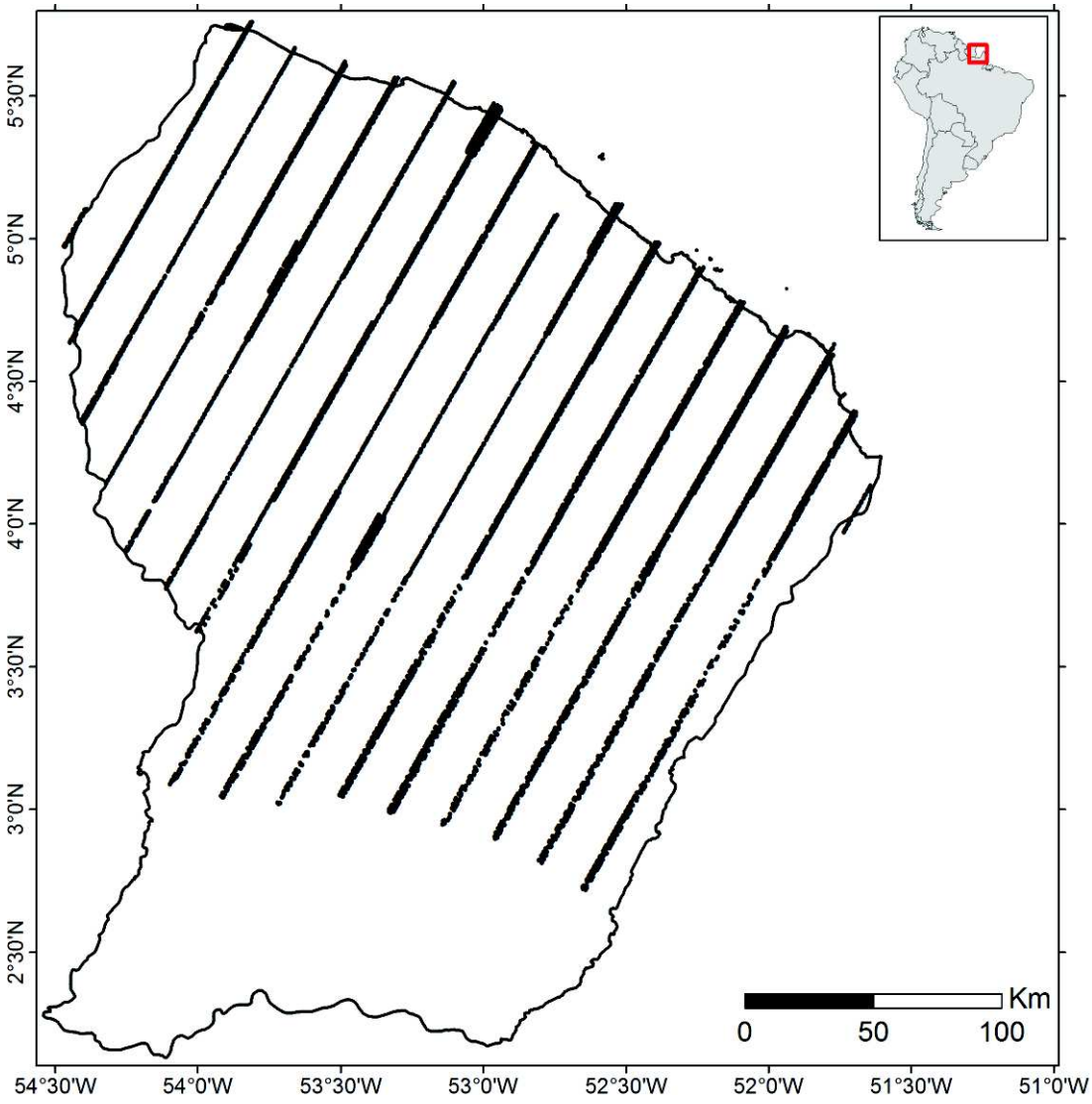


Figure 4.8 (cont.)



(b)

Figure 4.8 (cont.)

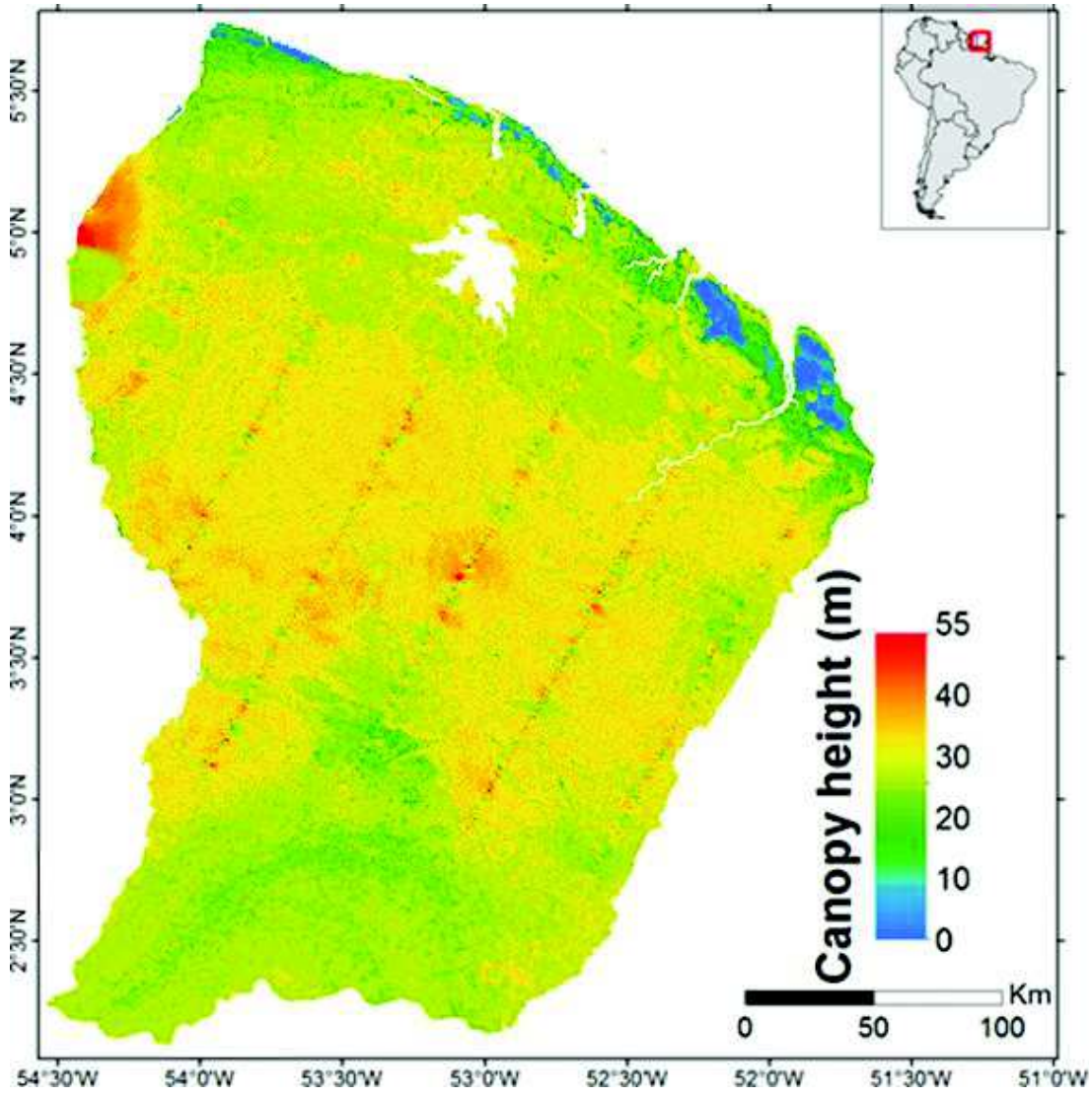
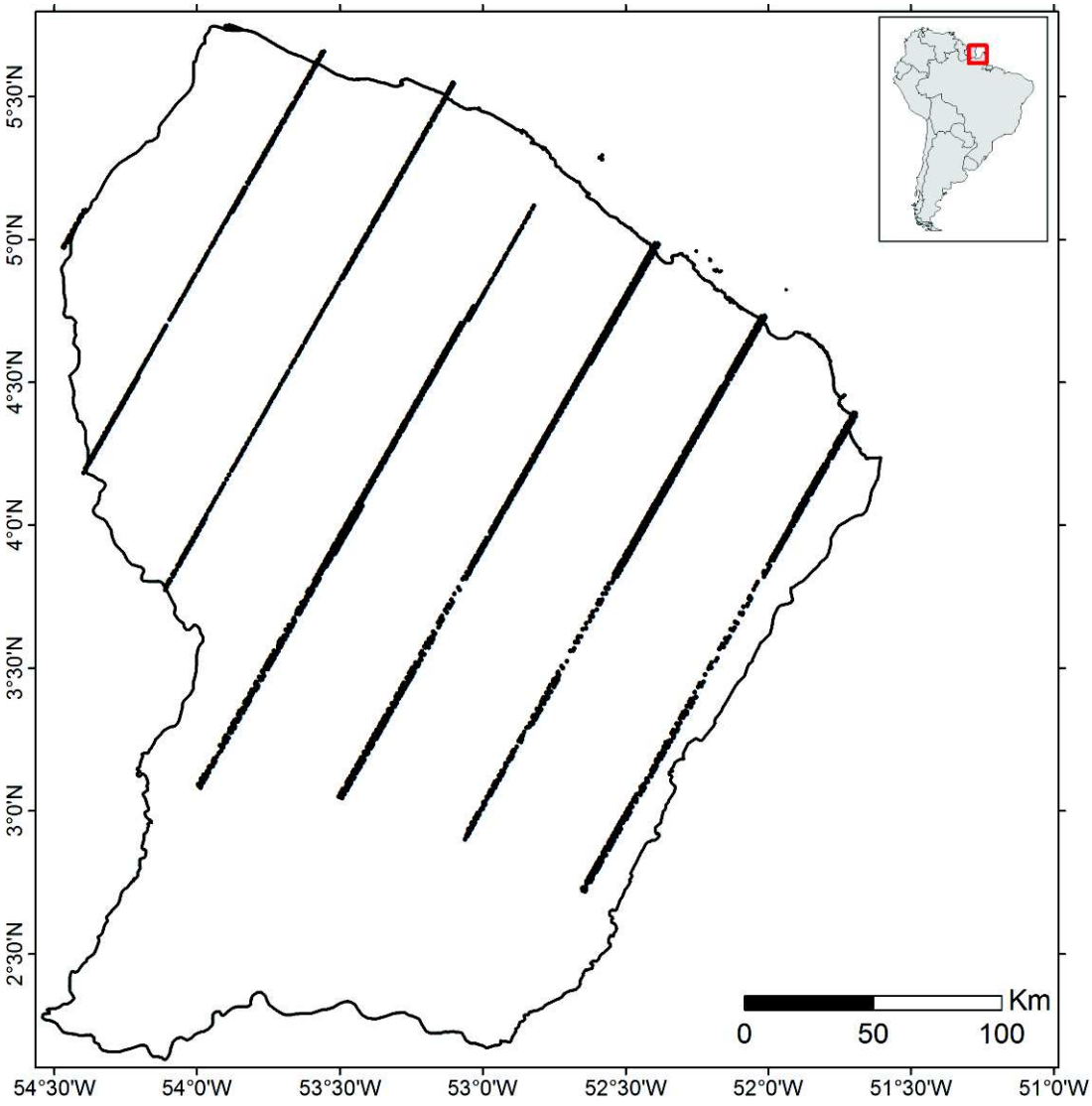


Figure 4.8 (cont.)



(c)

Table 4.1. Comparison between the canopy heights of the validation datasets (LD_val and HD) and the canopy height estimates using Random Forest regressions and residual-kriging.

	Using RF			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_10	0.09	6.0	0.61	0.07	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

4.4 Discussion

For the first time, airborne and space borne Lidar data were used separately 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 and regression-kriging. Because of this approach, both calibration and validation 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 such as the one presented in studies like Hudak *et al.* [66] 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 that 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 their opposite nature (low-bias/high-variance).

Moreover, our canopy height estimates using random forests show precisions slightly higher comparable to a recent study ([65]). The study of Simard *et al.* [65]; 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 validation dataset (LD_val and HD) showed a slightly higher RMSE of 7.3 m (Figure 4.9). 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.* [47]. 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.* [47] 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 ([65]).

However, neither our approach, nor the approaches used in Simard *et al.* [65] or Lefsky *et al.* [47] was able to produce wall-to-wall canopy height maps explaining more than 62% of the total variation of canopy heights (R^2) (Table 4.1). The last 38% should be related to forest dynamic endogeneous processes (i.e. gap processes). In addition, while our canopy height mapping using only RF regressions provides a good canopy height estimates (RMSE about 6 m) at large scales with a medium spatial resolution (250 m), such precision is not optimal when estimating forest biomass with allometries that only use canopy heights (e.g. [39]; [40]). Indeed, an RMSE of about 6 m using only the RF regressions leads to a relative error on the estimation of biomass of about 25%. This precision on the estimation of biomass is higher than the recommended relative error of 20% by the United Nations program on Reducing Emissions from Deforestation and Forest Degradation (REDD) ([95] ; [94]). Hence, to satisfy the UN REDD recommendations on the precision of biomass, improved canopy height estimates are required. Therefore, the canopy height estimation residuals (reference canopy heights – estimated canopy heights by RF) were kriged and used.

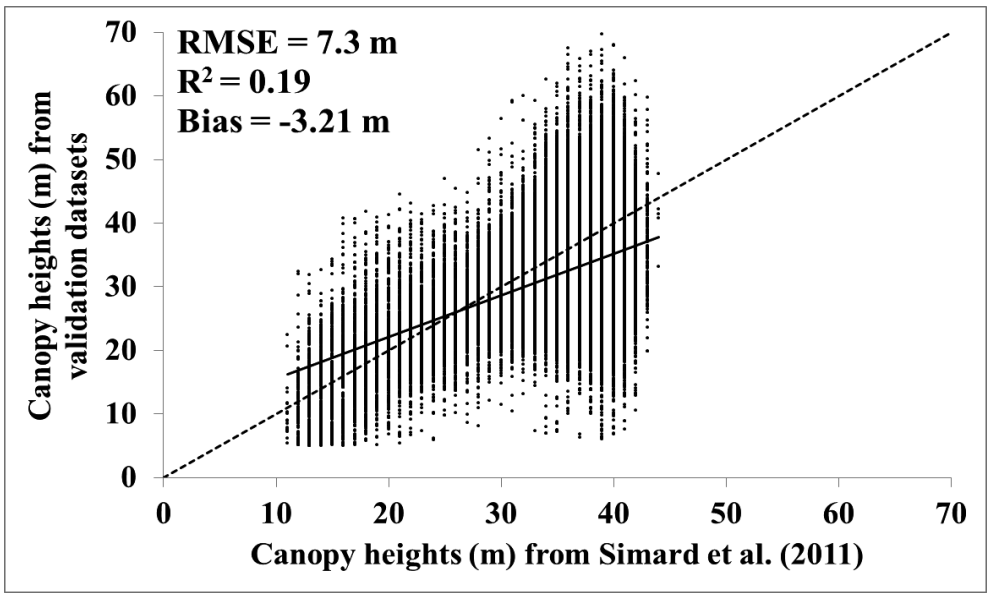
This approach proved very efficient, although highly sensitive to the spatial sampling of the reference LiDAR dataset (flight line spacing). Indeed for the French Guiana, the semivariograms indicated that the autocorrelation in the canopy height residuals did not go beyond 5 km, beyond this distance 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 validation 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 validation 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 validation datasets (bias of -4 m). To analyse 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) won't 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.

Figure 4.9. Comparison between the canopy heights of our validation datasets (LD_val and HD) and the canopy height estimates from the study of Simard *et al.* [65].



4.5 Conclusions

LiDAR (Light Detection And Ranging) remote sensing has been shown to be a good technique for the estimation of forest parameters such as canopy heights and above ground biomass. Whilst airborne LiDAR data are in general very dense but only available over small areas due to the cost of their acquisition, spaceborne LiDAR data acquired from the Geoscience Laser Altimeter System (GLAS) have low acquisition density with global geographical cover. It is therefore valuable to analyse the integration relevance of canopy heights estimated from LiDAR sensors with ancillary data (geological, meteorological phenological etc.) in order to propose a forest canopy height map with good precision and high spatial resolution. In this study, canopy heights extracted from both airborne and spaceborne LiDAR, were first estimated from available mapped environmental data (e.g. geology, slope, vegetation indices, etc.). The estimated canopy height maps using random forest (RF) regression with either the airborne (LD_cal) or GLAS calibration datasets showed similar precisions (RMSE better than 6.5 m). In order to improve the precision of the canopy height estimates regression-kriging (kriging of random forest regression residuals) was used. Results indicated a decrease in the RMSE from 6.5 to 4.2 m for the regression-kriging maps from the GLAS dataset, and from 5.8 to 1.8 m for the regression-kriging map from the airborne LiDAR dataset. Finally, in order to study the impact of the

spatial sampling of future LiDAR missions on the precision of canopy height estimates, six subsets were derived from the initial airborne LiDAR dataset with flight line spacing of 5, 10, 20, 30, 40 and 50 km (corresponding to 0.29, 0.11, 0.08, 0.05, 0.04, and 0.03 points /km² respectively). Results indicated that using the regression-kriging approach, the precision on the canopy height map was 1.8 m with flight line spacing of 5 km and decreased to an RMSE of 4.8m for the configuration for the 50 km flight line spacing.

CHAPTER 5:

COUPLING POTENTIAL OF ICESAT/GLAS AND SRTM FOR THE DISCRIMINATION OF FOREST LANDSCAPE TYPES IN FRENCH GUIANA

5.1 Introduction

Currently, land and forest cover classifications over large areas are made using high temporal frequency data provided by moderate spatial resolution sensors with a spatial resolution ranging from a few hundred meters (MODIS) to one kilometre (VEGETATION/SPOT). Nevertheless, the characterization and quantification of broad-scale forest land cover remains a major challenge for remote sensing scientists ([106]). Mayaux *et al.* [107] produced a land-cover map of Africa using the spectral response and the temporal profile of the vegetation cover. In their study, radar data and thermal sensors were also used for specific land-cover classes. In the Guiana Shield, Gond *et al.* [77] interpreted 33 remotely sensed landscape types (LTs) using VEGETATION/SPOT. Five of the 33 classes occupied 78% of the forests in the area. The method used by Gond *et al.* [77] used a multivariate analysis of remote sensing data, field observations and environmental data. However, due to LiDAR's ability to provide detailed information on the vertical structure of forests (canopy height, tree crown, etc.) in comparison to optical sensors, LiDAR appears to be one of the most applicable remote sensing techniques for forest monitoring ([108]; [109]; [110]). Conversely, optical sensors provide extensive

coverage of forests on the horizontal plane but are less sensitive to forest vertical structure variations ([111]). Generally, to better classify forest structures, canopy information on both the horizontal and vertical planes are required. In fact, studies that use LiDAR datasets in conjunction with optical data show better classification accuracy of forest structures ([111]; [112]; [113]). Indeed, in Mundt *et al.* [112]; the fusion of LiDAR and multispectral data provided an increase in the detection of sagebrush by 15% in comparison to using multispectral data alone. Dalponte *et al.* [113] used a fusion of LiDAR and hyperspectral data to classify complex forest areas with more than 20 tree species, with several similar tree species and with no preordered spatial distribution of trees. In their study, an increase of up to 9% in the classification accuracy was noted when adding LiDAR data. Finally, Ali *et al.* [111] fused LiDAR and multispectral data for the classification of three Eucalyptus types. Their results indicated an increase of 23% in the classification accuracy when using LiDAR data.

Our study uses the interaction between the Shuttle Radar Topography Mission (SRTM) data and vegetation in the five forest landscape types in French Guiana to assess the potential of the SRTM to identify these five forest types. This was accomplished by comparing SRTM elevations with elevations extracted from NASA's Geoscience Laser Altimeter System (GLAS) full waveform data, namely, the highest (most likely canopy top) and centroid (distance-weighted average) elevations.

Comparisons between the GLAS and SRTM elevations have been investigated in numerous studies, mainly for studying the SRTM penetration levels over different landscape types and using different elevation levels within the GLAS waveforms (highest, centroid and lowest) (e.g. [114]; [115]; [116]; [117]; and [70]). Calculating the bias (the difference between the SRTM elevations and the GLAS centroid elevations), Bhang *et al.* [114] found that elevation bias is dependent on the landscape type and the terrain relief. Bias increased from -1.5 m for bare terrain to -1 m for agricultural areas and 0.9 m for forested areas. Rodriguez *et al.* [115] compared the SRTM elevations with field measurements in different regions around the globe and found an elevation bias between the SRTM and field measurements that varied with each location. In their study, they compared the difference between the SRTM and the Land, Vegetation, and Ice Sensor (LVIS) canopy top elevations across five different regions in the USA (Maine, Maryland,

Massachusetts, New Hampshire, and Costa Rica) and found a maximum elevation bias difference of 54% between two regions (Costa Rica and Maryland). Becek [117] found a linear relationship between the magnitude of the elevation bias of the SRTM in comparison to reference field data and the percentage of tree cover. The bias increased from 4.8 m for 0% tree cover to 11 m for 100% tree cover. Carabajal and Harding [70] compared the highest, centroid and lowest GLAS elevations with the SRTM for different regions around the world (Amazon, Africa, Asia, Australia, and Western USA) and found different elevation biases (difference between GLAS highest and SRTM) that varied with region (a maximum difference of 16.6 m was found between Australia and Western USA).

In addition, bias appeared to be correlated with the GLAS waveform extent and the roughness index (bias increases with increasing waveform extent and roughness index). The GLAS waveform extent represents the distance between the laser signal start and the signal end. In vegetated areas, laser signal start corresponds to the highest canopy surface large enough to yield a return signal. Signal end corresponds to the lowest detected ground elevation. In vegetated areas, roughness represents the combined effect of topographic relief (top of canopy) and the interaction of the C-band microwaves with the vegetation ([70]).

Seasonal changes in the GLAS signal over different forest types have also been studied. Duong *et al.* [83] used the differences between overlapping pairs of GLAS footprints in different seasons (winter and summer) to differentiate between different forest structures. Their study showed promising results for identifying conifer, deciduous and mixed conifer/deciduous forests.

The objective of this study was to analyse the potential for forest landscape type mapping using the coupling of GLAS and SRTM data in French Guiana. First, the penetration depth of the SRTM radar wave corresponding to the difference between the GLAS highest elevations and the SRTM elevations, as well as the difference between the GLAS centroid elevations and the SRTM elevations, was analysed over the different forest landscape types obtained in the study of Gond *et al.* [77]. Next, the behaviour of these two variables was studied for the different LTs as a function of the GLAS canopy height and the SRTM roughness index. The classification potential for the five forest landscape types (LTs) using

the coupling of GLAS and SRTM was assessed using the Random Forest algorithm. This classification was conducted using the penetration depth, the difference between the GLAS centroid elevations and the SRTM elevations, the GLAS canopy height and the SRTM roughness index. Finally, using the changes in the GLAS signal in different seasons, the potential for LT discrimination using these changes was studied.

The processing of the GLAS waveforms and the methodology used to assess the potential of GLAS and SRTM in the discrimination of forest landscape types are shown in section 5.2. The results and discussion are given in section 5.3 and the conclusions in section 4.4. This chapter is based on the following published paper: **Fayad I *et al.* (2014) Coupling potential of ICESat/GLAS and SRTM for the discrimination of forest landscape types in French Guiana. International Journal of Applied Earth Observation and Geoinformation 33:21-31.**

5.2 Materials and methods

5.2.1 Methodology

To assess the potential of the GLAS and SRTM data to discriminate the five main forest landscape types in French Guiana, the difference between the GLAS highest elevations and the SRTM elevations was investigated, as well as the difference between the GLAS centroid elevations and the SRTM elevations. First, the differences between the GLAS (highest and centroid) and SRTM elevations were analysed for each forest landscape type. Next, because of the influence of canopy height (Hc) and the roughness index (R) on the penetration depth of the SRTM radar wave in the canopy, the differences between the GLAS and SRTM elevations were studied according to the classes of canopy height and the roughness index. Four canopy height classes ($Hc < 10$ m, $10 \text{ m} \leq Hc < 20$ m, $20 \text{ m} \leq Hc < 30$ m, and $Hc \geq 30$ m) and three roughness index classes ($R < 5$ m, $5 \text{ m} \leq R < 10$ m, $R \geq 10$ m) were chosen for each forest landscape type.

To analyse the potential for discrimination of the five main forest landscape types (LTs) using the coupling of the GLAS and SRTM data, a classification of the GLAS footprints based on the Random Forest algorithm was conducted using the penetration depth, the

difference between the GLAS centroid elevations and the SRTM elevations, the GLAS canopy height and the SRTM roughness index.

Several classifiers, such as CART (Classification And Regression Trees), SVM (Support Vector Machines), logistic regression, and the Random Forest classifier, were tested in this study. However, Random Forest represents the statistical mode of many classification and regression trees (CART); hence, it is a more robust model than a single tree ([93]). In addition, Random Forest does not over-fit, even if more trees are added, it always converges, it produces error estimates of the predictions and of the importance of the variables, and it handles weak explanatory variables. The variable importance index it produces is very important, as it allows an understanding of the relative values of the predictors used in the classification and therefore removes unnecessary predictors. Variable importance is based on two measures ([118]). The first is a measure of accuracy obtained by quantifying the mean squared error increase in the model by the removal of a variable. The other importance measure is the Gini index, which quantifies the degree to which a variable produces terminal nodes in the classification forest. Finally, the Random Forest classifier is less sensitive to outliers and noise (the 10 m vertical accuracy of the SRTM data in our case) in comparison to other classification routines ([93]).

The Random Forest (RF) algorithm is also known to be a powerful classification method that is becoming widely used by the remote sensing community for land-cover classification (e.g. [119]; [120]; and, [90]). RF is designed to produce accurate and robust predications without over-fitting the data while being insensitive to outliers and noise in comparison to single classifiers ([93]). Random Forest is called an ensemble classifier because it uses a tree-based classifier multiple times and aggregates the results. However, each tree is grown using a randomized subset of predictors. The final prediction decision is based on a voting system of all the predictions from the decision trees that have been created. Furthermore, because of the inability to examine the decision trees directly, Random Forest is considered more of a “black box” approach. However, several metrics are available to aid in the interpretation, one of which is the variable importance, which is evaluated based on the increase in the error in the prediction when removing a certain variable.

Finally, a comparison between the GLAS waveforms acquired at the same location in the dry and wet seasons was carried out to analyse its potential in the discrimination between the different forest LTs. The overlapping GLAS footprints, one from the wet season and one from the dry season, were compared against their corresponding SRTM elevations. However, to quantify the changes between the dry-wet season pairs, further processing of the waveforms was required. The first step was to obtain the waveform pairs. This was based on the geographic coordinates of each footprint (ellipse centre) found in the GLA14 product. Two footprints from the wet/dry seasons, according to Duong *et al.* [83]; were considered a pair if the distance between their centres was equal to or less than the sum of the footprints' radii divided by 2 (the two footprints were partly overlapping). Next, due to the different intensity returns caused by the laser output and/or different atmospheric conditions, the waveform pairs were normalized to enable comparison between them ([83]). The normalization procedure requires the division of each received bin voltage (V_i) by the total energy of the waveform V_t, where $V_t = \sum_{i=1}^{N} V_i$, with N being the number of waveform bins (544 or 1000 bins, depending on the GLAS mission). Finally, due to technicalities with the receiver, the recording of a pair of waveforms did not start at the same local time, even if they were similar in structure, thus producing a time lag, and hence, a shifting operation was needed. According to Hofton and Blair [121]; the shift operation can be performed on the complete waveform.

The time shift needed to match a pair of coincident waveforms was determined by the maximum of the cross correlation $\hat{R}(m)$ defined by (m=1,...,2N-1):

$$\widehat{\boldsymbol{R}}(\boldsymbol{m}) = \begin{cases} \sum_{\boldsymbol{i=0}}^{\boldsymbol{N-m-1}} \boldsymbol{W_w(i)W_d(i+m)} & \boldsymbol{for\ m \geq 0} \\ \widehat{\boldsymbol{R}}(\boldsymbol{-m}) & \boldsymbol{for\ m} < 0 \end{cases} \tag{5.1}$$

W_w and W_d represent the normalized signals from the wet and dry season, respectively. The results of the shifting algorithm are shown in Figure 5.1. As illustrated in Figure 5.1, two waveforms recorded in different seasons (wet and dry) and at approximately the same location do not match. The waveform from the dry season was shifted 43 ns to the right of the waveform from the wet season. Therefore, using the cross-correlation technique

described above, it was possible to match them for further comparisons. The flowchart summarizing the processing of the SRTM and GLAS data is shown in Figure 5.2.

Figure 5.1. Typical GLAS waveform acquired during the wet season (grey) used as a reference for the shifting of the same waveform acquired during the dry (dashed), and the shifted waveform from the dry season (black).

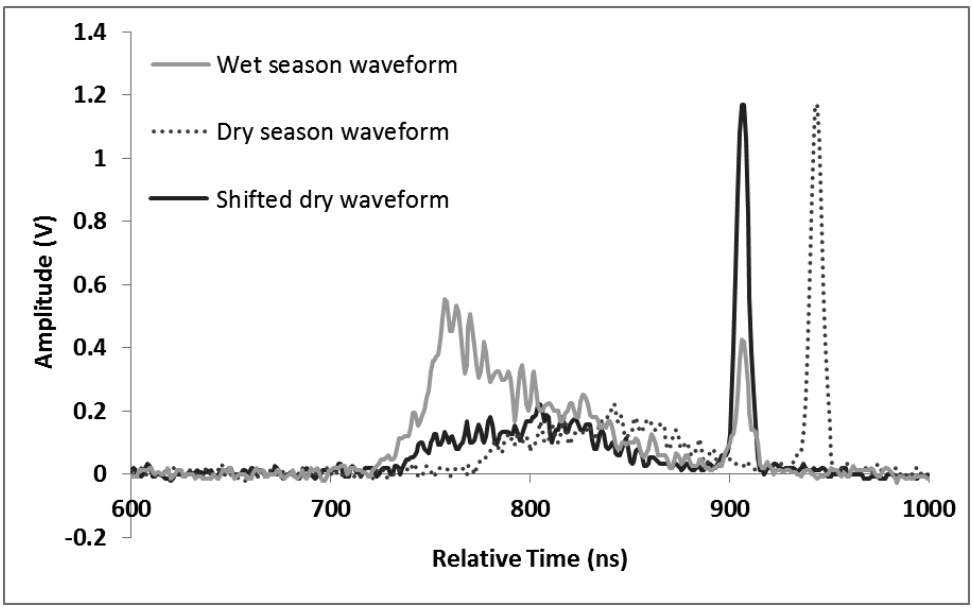
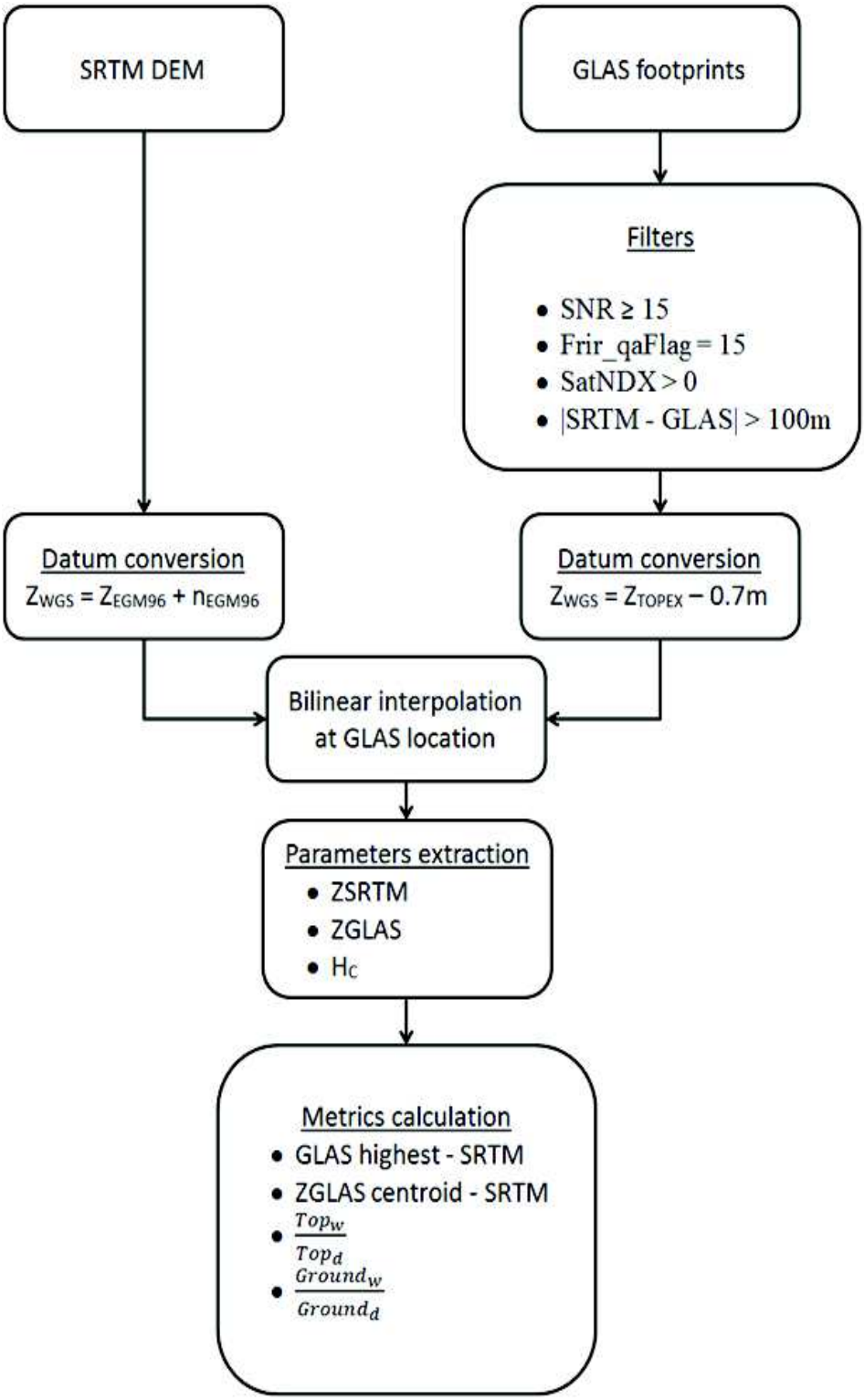


Figure 5.2. Flowchart of the processing steps for the GLAS and SRTM DEM data



5.2.2 GLAS waveform processing

To conduct a full comparison between the GLAS and SRTM elevations, several parameters needed to be extracted from the GLAS waveforms: signal start and end, ground peak position, highest, centroid and lowest elevations, and tree heights.

GLAS's signal start and end are defined as the first and last locations where the waveform intensity exceeds a certain threshold level (n.σb, where σb is the standard deviation of the background noise) above the mean background noise (μb) (e.g. [44]; [71]). Both μb and σb are found in the GLA14 product. However, there are no consistent optimal thresholds that can be used for all study areas ([71]). Different thresholds have been used in different studies, including 3σb in [82]; 4σb in [44] and 4.5σb in [72]. In this study, a threshold of 4.5σb was used. The difference between the signal end and signal start is called the waveform extent. The ground peak is identified using either the last peak (e.g. [82]) or the strongest in amplitude between the last two peaks (e.g. [81]; [121]). After close examination of the GLAS waveforms in French Guiana, the ground peak was identified using the Gaussian peak representing the highest amplitude from the last two peaks.

The GLAS product only provides the centroid elevation in a footprint. To estimate the highest and lowest elevations, the following approach was used. First, the position of the centroid within the waveform over the relative time axis was determined. Then, to determine the highest elevation, the difference between the position of the centroid and the signal start was added to the centroid elevation. Similarly, the lowest elevation was determined by subtracting the difference between the position of the centroid and the ground peak from the centroid elevation. The lowest elevation is less accurate than the top elevation because the identification of the ground peak is more error prone than the identification of the signal start.

5.2.3 Canopy height and roughness index estimations

The GLAS canopy heights (Hc) were estimated using the most commonly used method in areas of low relief, introduced by Lefsky *et al.* [44]; which uses the difference between the signal start (Hs) and the ground peak (Hg).

$$Hc = Hg - Hs \quad (5.2)$$

The roughness index (R), according to Carabajal and Harding [70]; was defined as the standard deviation of the values of the SRTM elevation data in a 3x3 window.

5.3 Results and discussion

In this section, we analyse the possibility for the discrimination of the different forest landscape types using the GLAS and SRTM data. First, in section 5.3.1, the discrimination potential for the different forest LTs is analysed using the SRTM penetration in the canopies (GLAS highest - SRTM). Then, section 5.3.2 analyses the discrimination potential according to two added parameters, canopy height (Hc) and the roughness index (R). The differences between the GLAS and SRTM elevations are grouped for each LT into four canopy height classes and three roughness index classes. Next, in section 5.3.3, all the parameters (Hc, R and the differences between the GLAS and SRTM elevations) are used in the Random Forest classifier to classify the GLAS footprints into the five different forest LTs. Finally, section 5.3.4 uses the variation in the GLAS waveforms captured from two seasons (wet and dry) to classify the different forest LTs.

5.3.1 Global analysis of the differences between the GLAS and SRTM elevations

Several studies, such as Bhang *et al.* [114]; Rodriguez *et al.* [115]; and Hofton *et al.* [121]; have demonstrated that the penetration depth of the SRTM wave (GLAS highest – SRTM elevation) is affected by the type of surface it interacts with (forest, agricultural areas, etc.). In this section, the differences between the GLAS (highest and centroid) and SRTM elevations are tested for each forest landscape type (LT) to ascertain the potential of the SRTM to discriminate between these five LTs.

Table 5.1 shows that using either the penetration depth of the SRTM radar signal or the difference between the GLAS centroid elevations and the SRTM elevations alone is not sufficient for discriminating the five forest landscape types (LTs). Indeed, penetration is similar for LT8 and LT10 (approximately 11.0 m) and for LT9 and LT11 (12.5 m). However, the characterization of these two classes (LT8 and LT10) is very different in

terms of structure. LT8 is closed, regular canopy, and LT10 is composed of high canopy with large emergent trees making it very irregular. Moreover, forest LT12 has a slightly lower penetration, with an average of 9.3 m. The difference between the GLAS centroid elevations and the SRTM elevations is also of the same order for LT8, LT9, LT10 and LT11 (approximately -5 m) and is lower for LT12 (-7.1 m). The same order of penetration for all LTs can be attributed, according to Carabajal and Harding (2006), to the following reasons: (1) the penetration of the SRTM varies with canopy height, which is not the same for all forest LTs, and (2) the roughness index plays a major role in the variability of the SRTM penetration. In the next section, the differences between the GLAS and SRTM elevations will be analysed according to the GLAS canopy height (Hc) and the SRTM roughness index (R) to investigate if the use of additional metrics improves the discrimination between the different forest LTs.

Table 5.1. Statistics (mean ± standard deviation) of the difference between GLAS highest and SRTM elevations for each of the five forest landscape types (LT).

LT	Highest – SRTM (m)	Centroid – SRTM (m)	Count
8	11.0±12.3	-4.2±12.1	1421
9	12.2±13.0	-5.0±13.1	7151
10	11.0±9.4	-5.5±11.4	1195
11	12.8±12.2	-5.1±12.0	2228
12	9.3±7.4	-7.1±7.2	243

5.3.2 Analysis of the differences between the GLAS and SRTM according to Hc and R

Section 5.3.1 showed that it was impossible to discriminate the five forest landscape types using the differences between the GLAS and SRTM elevations alone. According to some studies, other variables, such as canopy height (Hc) and the roughness index (R), might contribute to the variability in the SRTM signal's penetration depth. Carabajal and Harding [70]; Bhang *et al.* [114]; and Huang *et al.* [59] studied the penetration depth of the SRTM as a function of canopy height, and their results showed that penetration depth is dependent on canopy height (increases with an increase in Hc). Carabajal and Harding [70] and Bhang *et al.* [114] also studied the behaviour of penetration depth as a function of the roughness index, which appear to be positively correlated. To better analyse the effect of canopy height and roughness index on the discrimination of different LTs, the GLAS footprints

were regrouped for each LT, first into four canopy height classes and then into three roughness index classes.

5.3.2.1 Differences between the GLAS and SRTM according to Hc

The results indicate that the discrimination of the five forest LTs is not possible using only canopy height in the analysis of the difference between the GLAS highest elevations and the SRTM elevations (same penetration of the SRTM signal for the five forest LTs in each Hc class). Figure 5.3a shows that the difference between the GLAS highest elevations and the SRTM elevations increased with increasing Hc. This difference increased in LT8, LT9, LT11 and LT12 from approximately 3 m when Hc was less than 10 m to approximately 14 m when Hc was greater than 30 m. LT10 showed an increase in the SRTM penetration from 4.4 m when Hc < 10 m to approximately 14 m when Hc > 30 m. This slightly higher SRTM penetration for LT10 with Hc < 10 m was due to the mean canopy height being greater than that in the other LTs (8.5 m versus 5.5 m). The increase in the SRTM penetration with the increase in canopy height was due to the C-band phase centre (the position of the dominant backscattering level), which, on average, becomes increasingly biased below the canopy top with increasing waveform extent ([117]). These results comply with the study by Carabajal and Harding [70]. Similar findings were observed in the analysis of the difference between the GLAS centroid elevations and the SRTM elevations according to Hc.

Figures 5.4a and 5.4b show, respectively, the spatial distributions of the canopy height (Hc) and the penetration percentage (penetration depth divided by canopy height) in French Guiana. Figure 5.4a shows that over the coastal area, canopy heights tend to be no higher than 20 m (mangrove forest). The highest canopies are concentrated in the centre of French Guiana, with heights mostly above 30 m. In the south of French Guiana, canopy heights are shorter, ranging between 20 and 30 m, and are mostly classified as LT11 and LT12. Figure 5.4b shows that the penetration percentage is the highest (> 30%) in the centre of French Guiana, which mostly contains classes LT8, LT9 and LT10. For the coastal area and the south of French Guiana (mostly LT11 and LT12), the lowest penetration percentage (< 30%) was observed.

5.3.2.2 Differences between the GLAS and SRTM according to R

In addition, the results also show that it is possible to discriminate LT10 from LT8, LT9 and LT11 when R > 10 m (no available data for LT12 with R > 10 m). It is also possible to discriminate LT12 from the other classes when R < 5 m. This ability to discriminate LT12 from the other LTs is due to the fact that LT12 is structurally different from the other LTs (open forest with shorter canopy heights, on average). Figure 5.3b, which represents the difference between the GLAS highest elevations and the SRTM elevations as a function of the roughness index, shows that LT12 presents the lowest SRTM penetration when R is less than 5 m (4.9 m versus ~8 m for the other LTs). This can be explained by the denser and relatively shorter canopy forest represented within the fragmented horizontal structure of LT12 in comparison to other forest LTs.

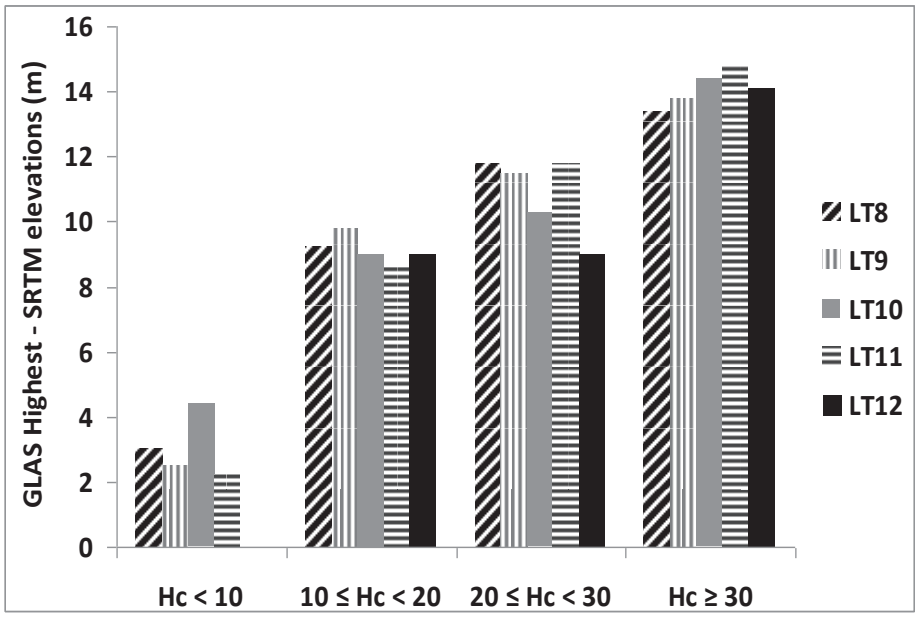
Furthermore, the difference between the GLAS highest elevations and the SRTM elevations increased significantly with increasing R for all forest LTs. For LT8, LT9 and LT11, the mean difference between the GLAS highest elevations and the SRTM elevations increased as a function of R from approximately 8 m when R was less than 5 m to approximately 16 m when R was greater than 10 m. For LT10, the average difference between the GLAS highest elevations and the SRTM elevations increased from approximately 8 m when R was less than 5 m to approximately 12 m when R was greater than 10 m. This is most likely due to the irregular heights of the canopies in this forest LT (high forest with disrupted canopy). Moreover, the average difference between the GLAS centroid elevations and the SRTM elevations showed low dependency on the roughness index, with an average between -4 m and -7 m for different R classes and forest LTs.

The spatial distribution of the roughness index presented in Figure 5.4c shows that the lowest roughness index values were observed on the coastal area, ranging mostly below 5 m, where they are attributed mostly to LT8 (very regular canopy roof). Low to moderate roughness index values, ranging between 0 and 10 m, were located in the south, where they are attributed to LT11 and LT12. The centre of French Guiana presents the highest values of the roughness index (> 10 m).

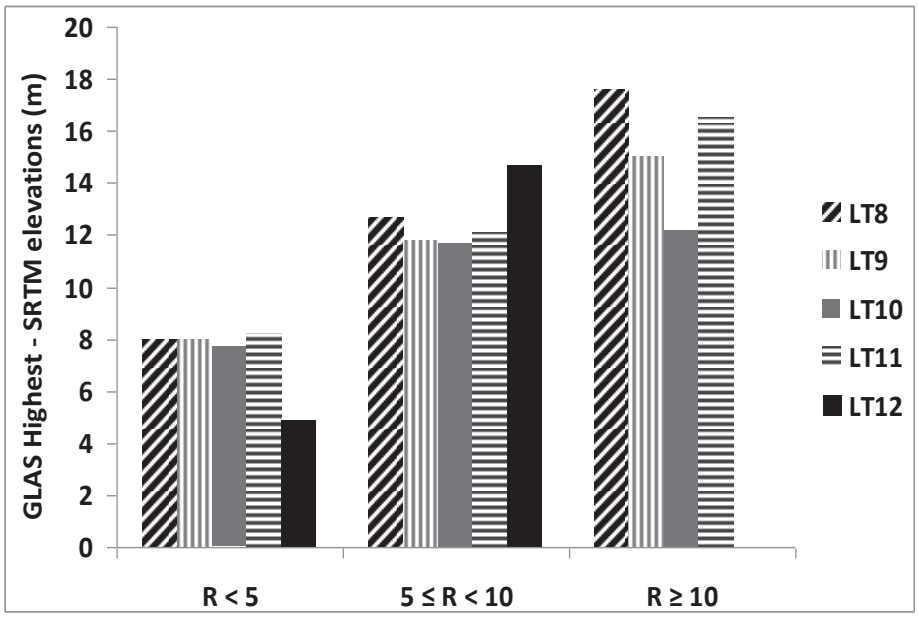
In conclusion, these results confirm that the discrimination between the five forest LTs requires the combination of several variables. Using the difference between the GLAS and

SRTM elevations, it was not possible to classify the different LTs as a function of Hc because of the effects of the roughness index. The same thing applies when attempting to classify different LTs as a function of R because of the effects of canopy height.

Figure 5.3. Differences between GLAS elevations (highest and centroid) and SRTM elevations for each forest landscape type (LT) according to four canopy height (H_c) classes (a) and three roughness index (R) classes (b). Only statistics with a count greater than 20 were used.

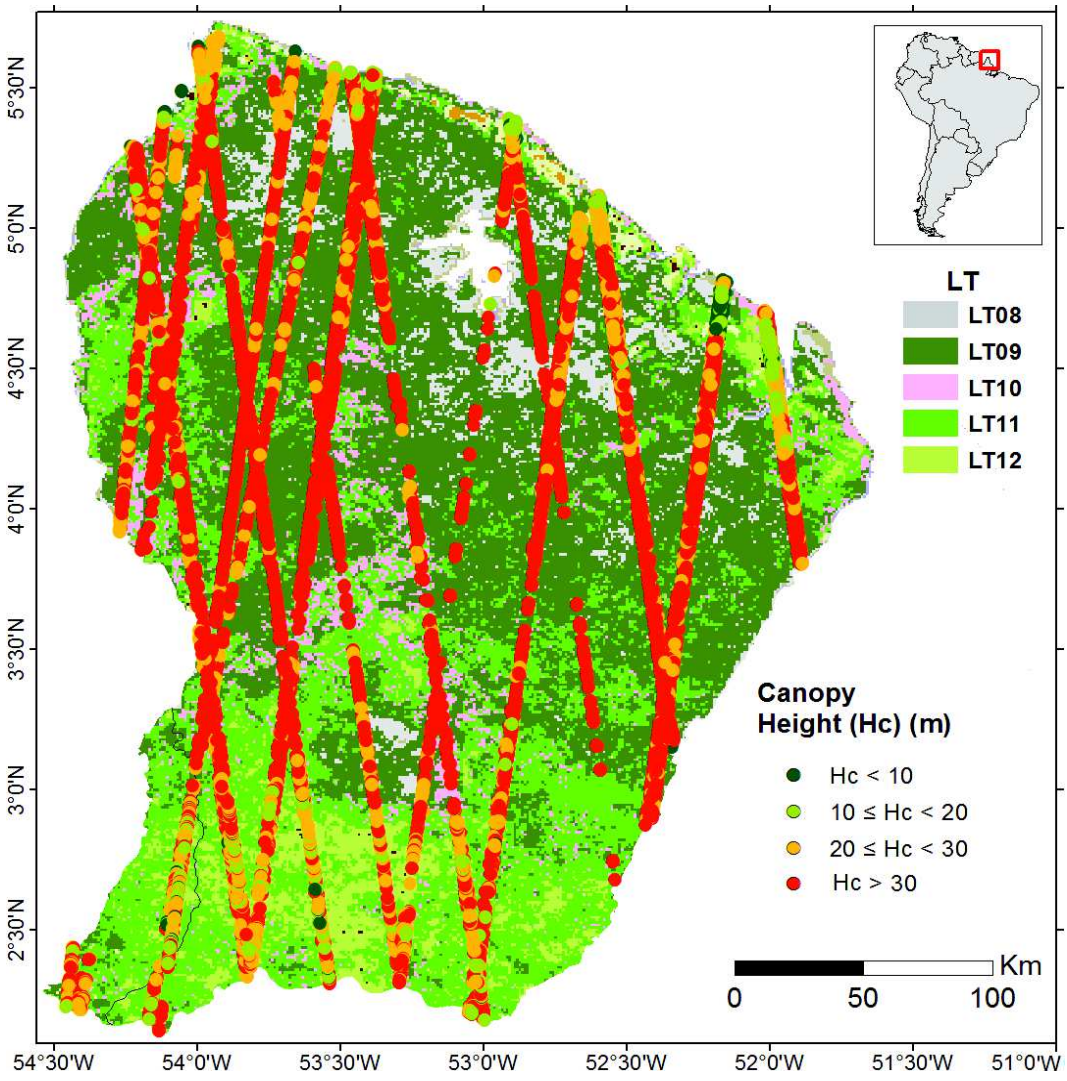


(a)



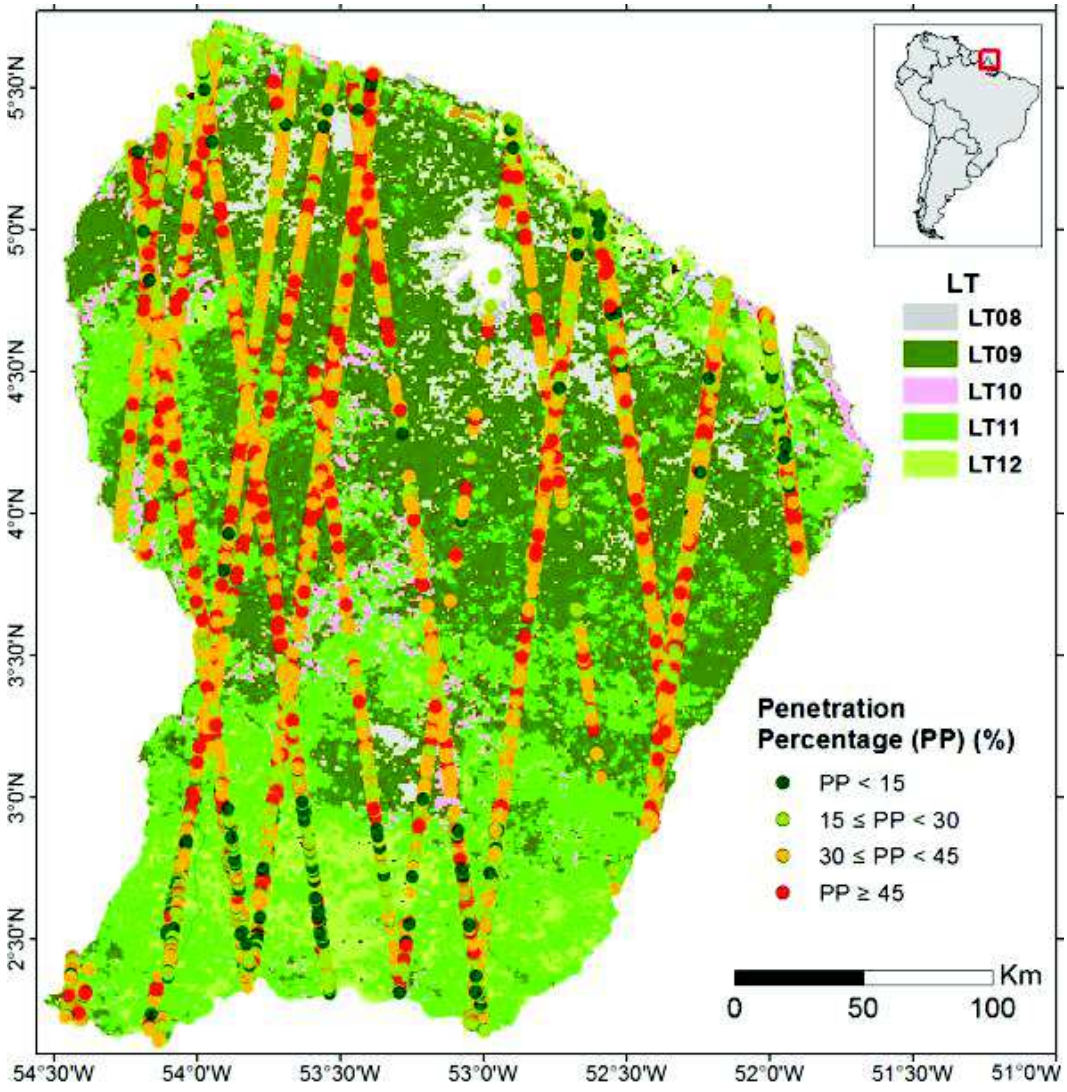
(b)

Figure 5.4. Spatial distribution of canopy heights (a), penetration percentage (b), and roughness index (c) over French Guiana.



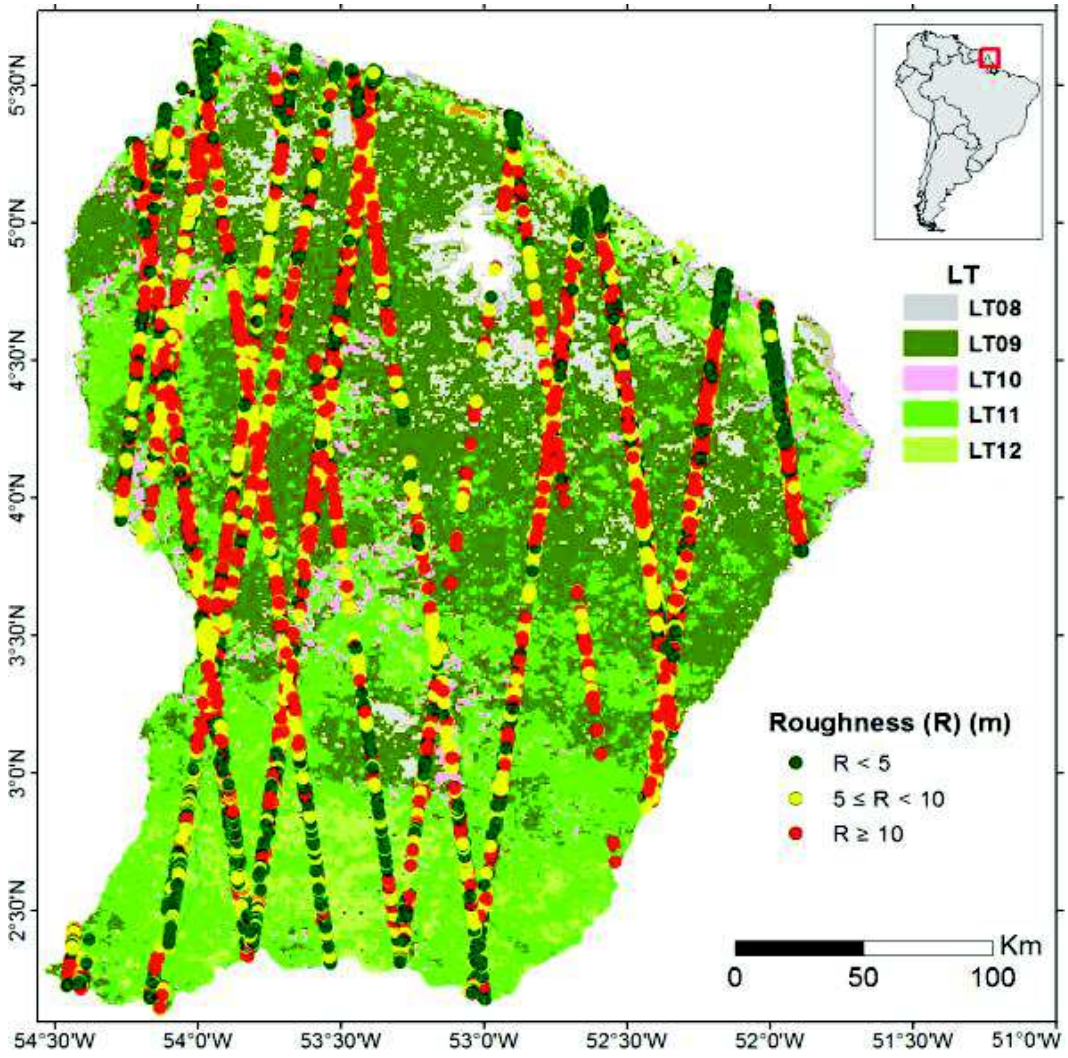
(a)

Figure 5.4 (cont.)



(b)

Figure 5.4 (cont.)



(c)

5.3.3 Random Forest classification results

The results of sections 5.3.2.1 and 5.3.2.2 showed that in order to discriminate between the five forest LTs, it is necessary to combine several variables, as it was not possible to classify the different Lts as a function of Hc or R alone. Therefore, in this section, the discrimination of the different LTs will be attempted using Hc, R, GLAS (highest and centroid) and the SRTM elevations with the Random Forest (RF) classification. The results show that all forest LTs were well classified with good accuracy, according to the map by Gond *et al.* [77]. The Random Forest classification results summarized in Table 5.2 show an overall accuracy of 83.3% (kappa coefficient of 0.75). Moreover, the producer's accuracies ranged between 78.4% (LT11) and 97.5% (LT12), and the user's accuracies

ranged between 77.3% (LT9) and 96.3% (LT12). The results show that the coupling between the GLAS and SRTM elevations allows better distinction between forests that are quite different (LT8, LT10, and LT12), and the most misclassifications were observed between LT9 and LT11 (approximately 12%).

The observed misclassifications between LT9 and LT11 could be explained by their similar characteristics and proximity and by the LT spatial distribution map used as a reference in our analysis ([77]), which has a sample size of 1 square-kilometre, whereas the GLAS footprints are, at most, 100 m in diameter. This difference in spatial scale could have had an effect on the classification results.

Our dataset contained forest LTs with uneven sample counts (LT9 represents more than 58% of the total dataset). The use of a dataset with uneven class sizes will result in a classifier biased towards the majority class ([122]). Therefore, a random under-sampling technique is often used ([123]). This technique balances the dataset by removing samples randomly from the majority class. However, the elimination of samples from a class could eliminate useful samples. Thus, it is recommended that the majority class be under-sampled into several subsets ([124]). Then, the classifier is trained and validated using each of the subsets, and the results of all the classifiers are averaged. In this study, the majority class LT9 was under-sampled into four subsets (each with 1788 points).

After the sub-setting, the subsets were each randomly divided into 90% training and 10% validation data samples. The prediction error based on a 10-fold cross-validation was estimated to validate the generalization performance of the Random Forest algorithm. The importance of the variables used in the Random Forest algorithm was also assessed. The results show that the difference between the GLAS highest elevations and the SRTM elevations was the most important variable, followed by canopy height and the roughness index. The difference between the GLAS centroid and the SRTM showed the lowest importance.

The influence of the size of the training dataset on the behaviour of the Random Forest classifier was subsequently assessed using three cases: (1) only 20 samples were used for each forest LT (the draw of the 20 random samples for each LT was repeated 100 times),

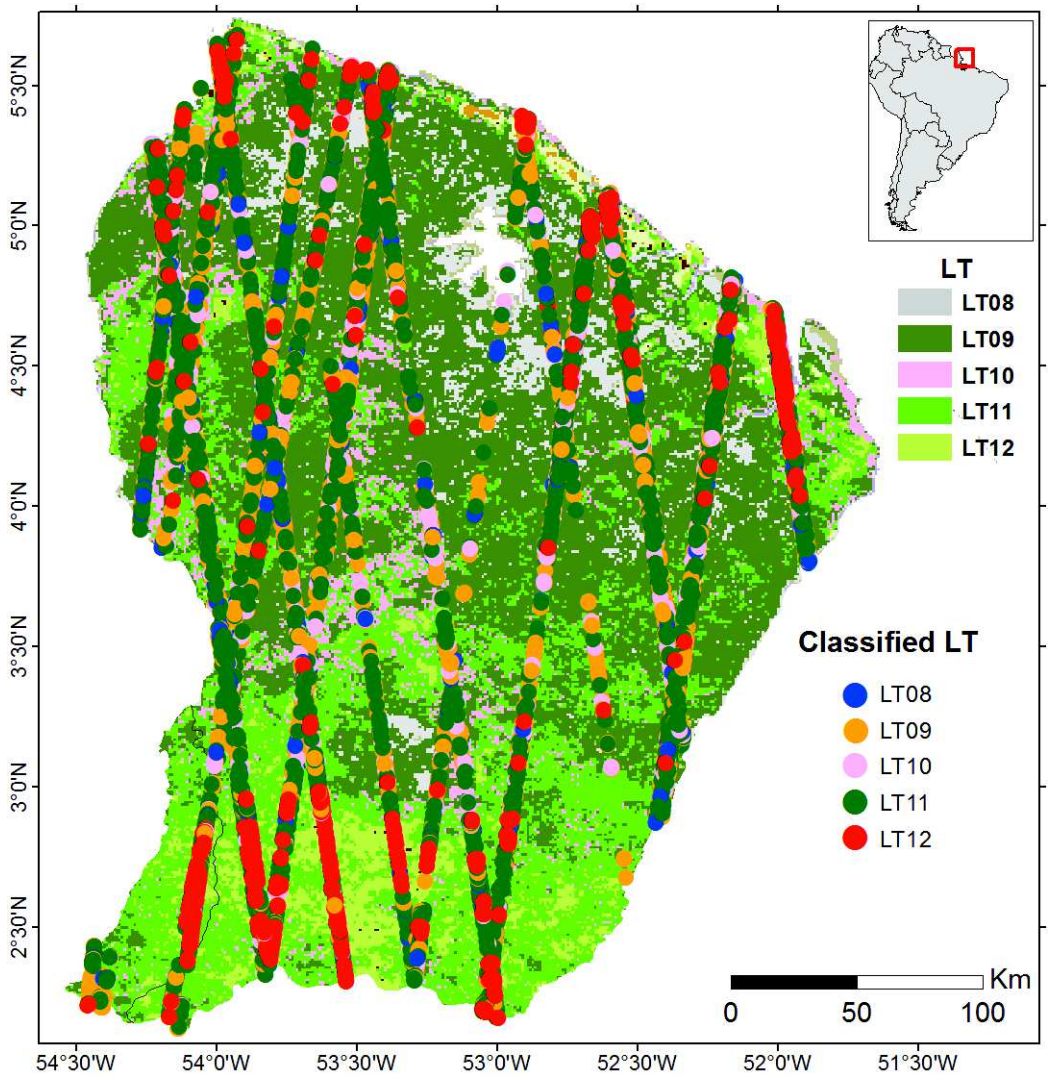
(2) 243 samples were used for each forest LT, corresponding to the lowest class size in our dataset (LT12), and (3) all samples available in our dataset were used with uneven class sizes between the LTs, with 243 samples from LT12 and 7151 samples from LT9. The results show that RF has a low sensitivity to the training dataset size reduction, with an overall classification accuracy slightly lower for the case with 20 samples for each LT (case 1) in comparison to the other two cases (approximately 78.0% for case 1 and 84% for cases 2 and 3). The kappa coefficient was also of the same order for the three cases, with values of approximately 0.7 for cases 1 and 3 and 0.8 for case 2. In addition, the producer's and user's accuracies were similar for all forest LTs, except for LT10 and LT11, where the accuracies were lower by approximately 20% for case 1 in comparison to cases 2 and 3 (approximately 64% for case 1 and 84% for cases 2 and 3). This result shows that LT10 and LT11 most likely have high intra-class variability, and for this reason, it is necessary to use a larger number of training samples for these forest LTs ([119]).

Finally, to ensure that RF is not over-fitting, an additional test was carried out. First, for each of the three cases described above, the explanatory variables (differences between the GLAS and SRTM elevations, the roughness index, and canopy height) were randomly permuted. Then, the classifier was trained and tested on the new dataset. Next, the predictions obtained from the 1000 datasets resulting from the permutation allowed the calculation of the 95% confidence interval for the different elements in a confusion matrix. The results show that at a 95% confidence interval, the user's and producer's accuracies were less than 25%, except for LT9 in case 3, where the user's and producer's accuracies were approximately 54%, and 91%, respectively. Classifiers trained with unequal class sizes have a tendency to classify the majority of the samples in the majority class to lower the classifier's error rate ([124]). In this study, LT9 contains approximately 58% of all the samples, so the classifier, in order to obtain the lowest classification error rate, classified most of the samples in LT9. The classification results are shown in Figure 5.5.

Table 5.2. Random Forest classification results for the five forest landscape types.

	Reference Classes						
Classified Classes	LT8	LT9	LT10	LT11	LT12	Classification overall	Producer's accuracy (Omission error)
LT8	1166	79	53	123	0	1421	82.1%
LT9	78	1382	58	267	3	1788	77.3%
LT10	52	77	966	99	1	1195	80.8%
LT11	100	221	66	1839	2	2228	82.5%
LT12	0	3	1	5	234	243	96.3%
Truth overall	1396	1762	1144	2333	240	6875	
User's accuracy (Commission error)	83.5%	78.4%	84.4%	78.8%	97.5%		
Overall classification accuracy = 81.3% and kappa $\kappa = 0.75$							

Figure 5.5. Classification of GLAS footprints into five forest landscape types using Random Forest algorithm.



5.3.4 Effect of the GLAS acquisition season

A study by Duong *et al.* [83] demonstrated the potential to classify broad-leaved, mixed and needle-leaved trees using GLAS footprint pairs taken from two seasons (dry and wet seasons). Two waveforms, one from the dry season and one from the wet season, were considered a pair if the distance between the footprint centres was less than or equal to the sum of their radii divided by 2. In this section, waveform pairs from the wet (December to June) and dry (August to December) seasons are compared to show the potential for discriminating different forest LTs using five criteria: (1) difference between the GLAS highest elevations and the SRTM elevations, (2) difference between the GLAS centroid elevations and the SRTM elevations, (3) penetration percentage in each season, (4) ratio of

the canopy top energy in the wet season to the canopy top energy in the dry season, and (5) ratio of the ground energy in the wet season to the ground energy in the dry season. The canopy top energy is defined by Duong *et al.* [83] as the energy from signal start to end from the vegetated part of the canopy, and the ground energy corresponds to the energy from the ground peak.

The analysis of the difference between the GLAS (highest and centroid) and the SRTM elevations, conducted separately for the GLAS footprints acquired during the wet and the dry seasons, showed similar differences for both seasons (Table 5.3). The difference between the GLAS highest elevations and the SRTM elevations for all LTs varied from 9.1 to 12.3 m in the wet season and from 9.6 to 12.9 m in the dry season. The difference between the GLAS centroid elevations and the SRTM elevations varied between -3.1 and -6.7 m in the wet season and between -4.7 and -7.7 m in the dry season. This slight difference between the two seasons could be due to some trees losing leaves in the dry season, meaning the GLAS waveform penetrates more into the canopy in the dry season compared to the same canopy in the wet season.

The waveform pairs were next separated into two categories. The first, called leaf-on corresponds to the waveforms without changes between the two seasons (trees with leaves in both seasons). The second category, called leaf-off, represents trees that shed their leaves in the dry season. Leaf-off trees were identified from the GLAS waveform pairs that showed changes in the amplitudes of the canopy top and ground peaks from one season to another (the distance between footprint pairs should be lower than half the sum of their radii). The threshold was set to be half of the reference top and ground peaks. In general, when trees shed their leaves in the dry season, the GLAS waveform reflections from the ground are more prominent. This causes an increase in the amplitude of the ground peak with a decrease in the amplitude of the canopy top in comparison to the waveforms from the wet season. In total, 71 waveform pairs were identified for LT8 (all leaf-on), 62 pairs for LT10 (all leaf-on), and 70 pairs from LT12 (7 leaf-on and 33 leaf-off) (Table 5.4).
The ratio of the energies from the canopy tops in the wet and dry seasons $\left(\frac{Top_w}{Top_d}\right)$, as well as the ratio of the energies from the ground in both seasons $\left(\frac{Ground_w}{Ground_d}\right)$, were calculated (Table 5.4). On average, the largest change was detected in the leaf-off pairs of LT12, with

a $\left(\frac{Top_w}{Top_d}\right)$ of 2.4, as opposed to 1.1 for LT8, 0.9 for LT10, and 1.2 for the leaf-on LT12. Moreover, Table 5.4 shows that the ratio $\left(\frac{Ground_w}{Ground_d}\right)$ is similar for LT8, LT10 and the leaf-on LT12, with a value of approximately 1.1. A lower ratio value was obtained for the leaf-off LT12 (0.7) because canopies in the leaf-off class of LT12 appear to shed their leaves in the dry season, allowing a greater reflection from the ground surface. Table 5.4 also quantifies the changes in the waveform centroid elevations in comparison with the SRTM. For the pairs in LT8, LT10 and the leaf-on LT12, the difference between the GLAS centroid elevations and the SRTM elevations is similar for the data in both the dry and wet seasons, with values between -4.3 and -6.3 m for the wet season and between -4.7 and -5.8 m for the dry season. Conversely, this difference increases for the leaf-off pairs of LT12, from -5.8 m in the wet season to -9.6 m in the dry season. This difference means that the GLAS is able to penetrate the forest to a deeper depth in the dry season if the forest characteristics (leaves) change between the seasons. In conclusion, LT12, a deciduous forest type, could efficiently be discriminated based on the seasonal variation of the GLAS signal. Pennec *et al.* [153] found that in comparison to other LT types, LT12 has the highest enhanced vegetation index (EVI) all year round. This high EVI could be caused by under-canopy activity. This result is very important because it shows the utility of multi-season LiDAR data for mapping forest types that lose their leaves in the dry season, which is not possible with the optical imagery typically used. Figure 5.6 shows the locations of the leaf-on and leaf-off pairs in French Guiana, revealing that the majority of the leaf-off pairs are located in the same area in the south of French Guiana.

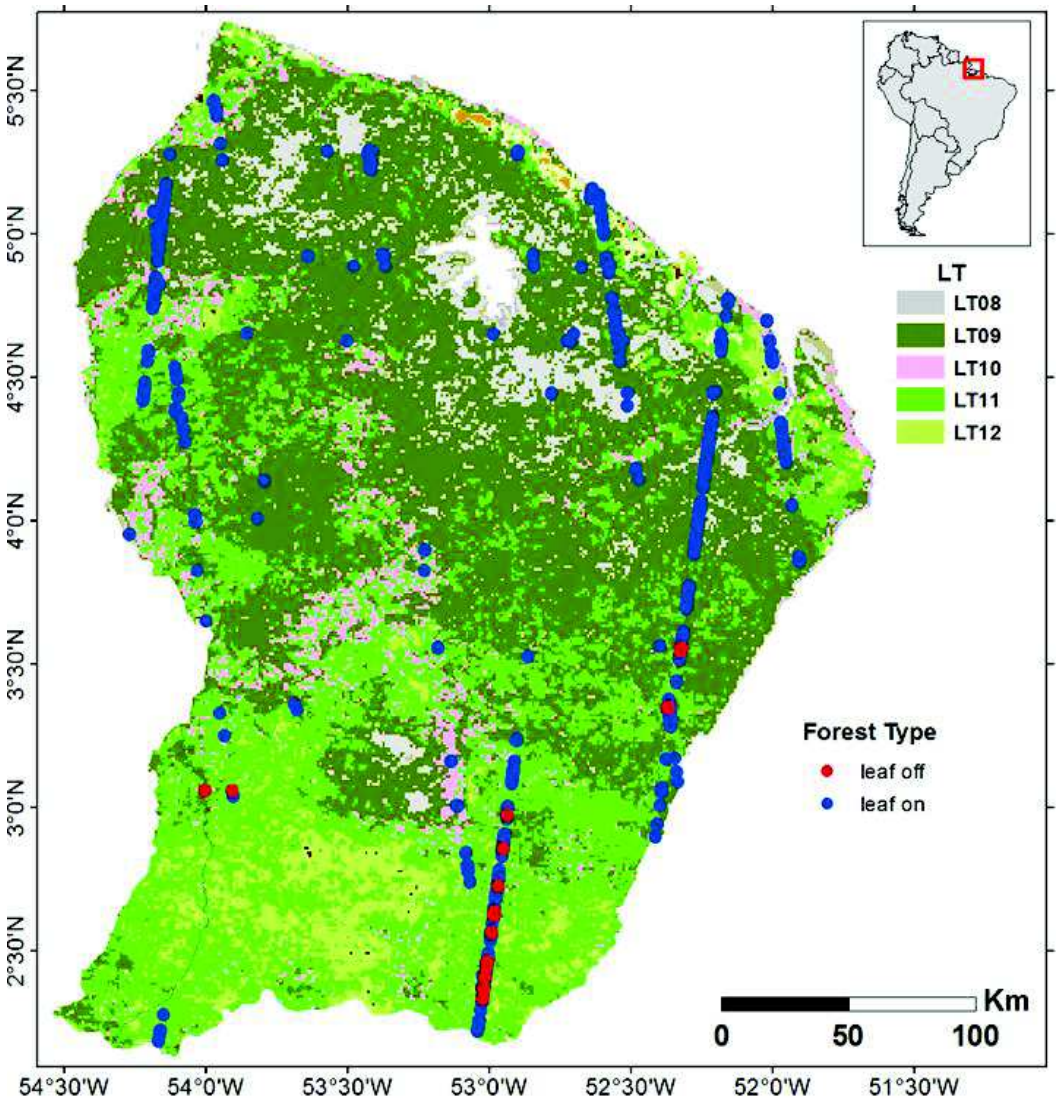
Table 5.3. Statistics (mean ± standard deviation) of the difference between GLAS (highest and centroid) and SRTM elevations for each of the forest landscape type (LT) in each season.

Wet season			
LT	Highest – SRTM (m)	Centroid – SRTM (m)	Count
8	11.4±12.1	-3.1±11.1	430
9	11.8±12.7	-4.7±12.7	1885
10	10.9±9.8	-5.3±9.4	374
11	12.3±11.7	-4.8±11.7	584
12	9.1±7.5	-6.7±8.0	152
Dry season			
LT	Highest – SRTM (m)	Centroid – SRTM (m)	Count
8	10.9±12.4	-4.7±12.4	991
9	12.3±13.0	-5.1±13.2	5266
10	10.9±10.3	-6.0±11.1	821
11	12.9±12.4	-5.2±12.1	1644
12	9.6±7.4	-7.7±5.4	91

Table 5.4. Comparison between wet and dry seasons for different forest LTs (no data for LT9 and LT11). Top_w and Top_d represent the energy of the signal reflected from the canopy top for the wet and dry seasons, respectively. $Ground_w$ and $Ground_d$ represent the energy of the signal reflected from the ground for the wet and dry seasons, respectively.

LT	Pairs number	Season	Highest – SRTM (m)	Centroid – SRTM (m)	$\frac{Top_w}{Top_d}$	$\frac{Ground_w}{Ground_d}$
8	71	Wet	11.5±15.0	-4.3±14.9	1.1±0.5	1.2±0.6
		Dry	11.2±14.4	-4.7±14.6		
10	62	Wet	10.9±14.3	-6.3±13.9	0.9±0.5	1.1±0.5
		Dry	11.0±14.7	-5.8±14.3		
12 (leaf-on)	7	Wet	9.5±13.3	-5.5±13.0	1.2±0.4	1.1±0.7
		Dry	9.9±15.2	-4.9±15.3		
12 (leaf-off)	33	Wet	9.3±12.3	-5.8±12.1	2.4±1.2	0.7±0.5
		Dry	9.7±14.4	-9.6±14.7		

Figure 5.6. Spatial location of leaf-on and leaf-off GLAS footprint pairs over French Guiana.



5.4 Conclusions

The Shuttle Radar Topography Mission (SRTM) has produced the most accurate nearly global elevation dataset to date. Over vegetated areas, the measured SRTM elevations are the result of a complex interaction between radar waves and tree crowns. In this study, waveforms acquired by the Geoscience Laser Altimeter System (GLAS) were combined with SRTM elevations to discriminate the five forest landscape types (LTs) in French Guiana. Two differences were calculated: (1) penetration depth, defined as the GLAS highest elevations minus the SRTM elevations, and (2) the GLAS centroid elevations minus the SRTM elevations. The results show that these differences were similar for the

five LTs, and they increased as a function of the GLAS canopy height and of the SRTM roughness index. Next, a Random Forest (RF) classifier was used to analyse the coupling potential of GLAS and SRTM in the discrimination of forest landscape types in French Guiana. The parameters used in the RF classification were the GLAS canopy height, the SRTM roughness index, the difference between the GLAS highest elevations and the SRTM elevations and the difference between the GLAS centroid elevations and the SRTM elevations. Discrimination of the five forest landscape types in French Guiana was possible, with an overall classification accuracy of 81.3% and a kappa coefficient of 0.75. All forest LTs were well classified with an accuracy varying from 78.4% to 97.5%.

Finally, differences of near coincident GLAS waveforms, one from the wet season and one from the dry season, were analysed. The results showed that the open forest LT (LT12), in some locations, contains trees that lose leaves during the dry season. These trees allow LT12 to be easily discriminated from the other LTs that retain their leaves using the following three criteria: (1) difference between the GLAS centroid elevations and the SRTM elevations, (2) ratio of top energy in the wet season to top energy in the dry season, or (3) ratio of ground energy in the wet season to ground energy in the dry season.

GENERAL CONCLUSIONS AND PERSPECTIVES

6.1 Conclusions

With the ending of the fifth chapter, it was clear that all the objectives set at the beginning of the thesis have been met. First, waveforms from GLAS have been used to estimate canopy heights. Next, canopy heights estimated from GLAS and airborne LiDAR have been used with environmental predictors to map canopy heights on the entire French Guiana. Finally, due to the role of forest landscape types in AGB estimation methods, forest landscape types have been predicted using information from GLAS waveforms and SRTM data. The main conclusions can be summarized as follows:

In chapter 3, the performance of the most frequently used linear regression models for canopy height estimation, which use metrics extracted from GLAS waveforms, was first evaluated. Then, models based on two seldom-used techniques for canopy height estimation from GLAS waveforms were introduced. The first included regression models using the principal component analysis (PCA) of GLAS waveforms. The second was based on the Random Forest technique. The Random Forest technique first used the metrics derived from the GLAS waveforms and then used PCs. The evaluation of these different models was performed with a large database consisting of GLAS data and canopy heights estimated from small-footprint airborne LiDAR measurements. Within the GLAS

footprints, which fell mostly on flat and sometimes moderately sloping terrain (slope < 15°), the direct method based on the difference between the ground peak and signal start showed an accuracy precision of 7.9 m (RMSE).

The linear regression models that used a combination of waveform extent (Wext), modified trailing and leading edge extents (Trail and Lead) [44] and terrain index (TI) showed better accuracies for canopy height estimation in comparison to the direct method, with an RMSE between 3.7 (using Wext, TI, and Trail) and 4.9 (using Wext and TI) m. In addition, the results reveal that the most relevant metrics in the estimation of forest heights are the waveform extent (Wext) and trailing edge extents (Trail). The linear regression model based on Wext and Trail estimated canopy height with an RMSE of 3.8 m. However, this model requires the Trail metric, which is difficult to extract with good accuracy in densely vegetated forests, such as those in French Guiana, affecting canopy height estimation due to the large contribution of the Trail metric to the linear regression models. The contribution of the leading edge extent (Lead) and TI calculated from the SRTM DEM appears to be very weak.

The linear regression model using the first 13 PCs and incorporating the waveform extent provided canopy height estimates with an RMSE of 3.8 m. The PCA regression models appear to be better in comparison to the other linear regression models using the GLAS waveform metrics with the same precision, as they do not use difficult-to-extract metrics, such as the Trail metric. The PCA model only requires the determination for each GLAS waveform, the class to which the Wext belong to (Wext lower than 20 m, Wext between 20 m and 40 m, or Wext higher than 40 m). Thus, even if the estimation of Wext depends on the signal start and signal end metrics, which are sometimes difficult to calculate with certainty, the error in the estimation of Wext does not affect the estimation of canopy height because the Wext classes are defined in large intervals (20 m).

The Random Forest model using all metrics (Wext, Trail, Lead, and TI) had an RMSE of 3.4 m. Using only one of the Trail, Lead or TI metrics in addition to Wext slightly increased the RMSE to 3.6 m. Using only Wext, which has a relative importance factor almost three times higher than those for the other metrics, produced canopy height estimates with a precision of 4.4 m. Finally, using the first 13 PCs in the Random Forest

regressions showed similar canopy height estimation results in comparison to using the PCs in the linear regression models, with an RMSE of 3.7 m when using the waveform extent and the four most important PCs.

In conclusion, the random forest regression models using the GLAS metrics did not show and improvements in terms of precision on the canopy height estimation in comparison to the linear regression models using the same metrics. In addition, the PCA based approach produced similar precisions on the canopy height estimation in comparison to the metric based approach. However, the advantage of using the PCA based approach is its independence to GLAS metrics that are difficult to extract in dense vegetated areas such as the Trail.

In chapter 4 an approach for canopy heights mapping over French Guiana was presented , given the limited coverage of LIDAR datasets (either airborne of spaceborne). 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 was 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 on the precision of the canopy height maps was observed. For the GLAS dataset the RMSE on the canopy height estimates was improved to 4.2 m and for the airborne LiDAR dataset the RMSE on canopy height estimates was improved to 1.8 m. However, this improvement is positively correlated with the point density of the calibration datasets used. Indeed, the GLAS dataset has a very large flight line spacing (~30 km on average) in comparison to the airborne LiDAR dataset (~0.5 km) Further investigation shows that for the airborne LiDAR subsets (subsets of the

original airborne LiDAR dataset with different flight line spacing) 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.

The procedure presented in chapter 4 which uses the regression-kriging has strong prospects for application to other tropical forests. In comparison, studies such as Simard *et al.* [65] which used only the Random forest regressions were only been able to estimate canopy height at a precision of 6.1 m with a lower resolution maps (1 km). In comparison, our 250 m forest height map with regression-kriging and using GLAS the RMSE on the canopy height estimate was 4.8 m

Finally in chapter 5, the coupling of the GLAS and SRTM DEM elevations (GLAS highest – SRTM and GLAS centroid - SRTM) was assessed in order to analyse the potential of discriminating different forest landscape types in French Guiana. A dataset of 12238 GLAS elevations over French Guiana calculated from GLAS waveforms, namely the highest and centroid elevations was compared to SRTM elevations. Based on VEGETATION-SPOT derived forest landscape types by Gond *et al.* [77]; GLAS footprints and their corresponding SRTM elevations were analysed according to the five forest landscape types.

Results showed that the mean differences between GLAS and SRTM elevations were of the same order for all forest landscape types (LTs). Furthermore, these differences increased as a function of GLAS canopy height, and SRTM roughness index in all forest landscape types (LT). Hence, the discrimination between the different forest LTs requires in addition to the differences between GLAS and SRTM elevations, other variables such as canopy height and roughness index.

A classification based on the Random Forest technique using the differences between GLAS and SRTM elevations as well as the canopy height and the roughness index was conducted. All forest LTs were well classified with accuracies between 78.4% à 97.5%.

Furthermore, the classification was achieved with an overall accuracy of 81.3% (Kappa coefficient of 0.75).

Coinciding GLAS footprints, with one from the wet season and one from the dry season were analysed to study the potential of discriminating different forest LTs according to the changes occurring between the GLAS waveforms pairs. Results show that the open forest (LT12), which is known to be mostly deciduous, was discriminable from other forest LTs using one of the following three criteria: (1) difference between GLAS centroid and SRTM elevations, or (2) ratio of energy from canopy top from the wet season to energy from canopy top in the dry season, or (3) ratio of ground energy from the wet season to the ground energy from the dry season.

6.2 Perspectives

This research opens on a number of perspectives cited below. They encompass subjects from canopy height model improvements to biomass estimation.

6.2.1 Canopy height estimation using GLAS

In chapter 3 a new technique using PCA was presented that allowed the estimation of canopy heights without using metrics extracted from GLAS waveforms. This was beneficial since a lot of GLAS metrics are difficult to estimate in dense vegetated areas. Our approach allowed the estimation of canopy heights with a precision comparable to models requiring the incorporation of GLAS metrics presented in other studies. However, our approach was tested and validated in French Guiana, which is mostly over flat terrain. Generally in sloping areas, the value of the slope must be taken into account with canopy height estimation models that uses GLAS waveforms in order to avoid overestimation of canopy heights. Therefore, it is necessary to test the PCA technique in different forested areas and with different terrain reliefs. If the presence of the slope will affect negatively the canopy height estimates with the use of the PCA technique, several improvements must then be made, such as the incorporation of information on the slope, terrain roughness or relief.

Moreover, given that GLAS is a large footprint satellite LiDAR that aims to map canopy heights at the regional and global scales. Therefore is its necessary to study the generalizability of the PCA approach, both at inter-site and intra-site levels. The inter-site refers how the model calibrated in one site would behave in another site, while intra-site refers how the model would behave when calibrated with one forest community and tested in another forest community within the same site. The generalizability of a model is crucial since every model should be calibrated using in-situ or airborne LIDAR measurements. While we had a large database of airborne LIDAR measurements for our study area, this might not be the case in other study sites.

6.2.2 LiDAR canopy height mapping

In chapter 4, forest canopy height mapping has been carried out with regression-kriging and using airborne and spaceborne LiDAR datasets. While the canopy height estimation results were satisfying, several aspects still need further study.

6.2.2.1 Non-spatial canopy height mapping

For the non-spatial canopy height mapping method, we used the random forest regression. However, since most of the predictors are correlated with canopy heights, inversion regression models ([125]) could be used instead of RF and their performance analysed. In addition, there exist many other predictors other than those used in chapter 4, which also are dependent on canopy heights (i.e. surface radiance) that might be interesting to include in the non-spatial model. Finally, all remotely sensed data including airborne LiDAR are subject to error sources such as, sensor calibration, sensor drift, signal digitization, atmospheric attenuation, etc. An alternative regression model that can be used instead of RF and that accounts for errors in both the predictors and dependent variables is reduced major axis (RMA) regression ([125]; [126]).

6.2.2.2 Spatial canopy height mapping

With regard to the spatial estimation methods, in extrapolation situations such as in this study, the use of universal kriging ([127]; [128]) or ordinary kriging with an external drift or regression-kriging ([129]) are advised. However, in cases where anisotropy exists in the landscape which was not studied in chapter 4, anisotropic kriging models having a

directional component are employed. Therefore, this point should be addressed in future studies.

6.2.2.3 Canopy height map resolution

The final canopy height product produced in chapter 4 had a resolution of 250 m. This was due to the resolution of the most important predictor according to the RF regression. However, several maps with different resolutions should be produced and studied; because, even though, the most important predictor has a resolution of 250 m, three less important predictors with 90 m resolution are also used in the final regression model. Maps with higher resolution allow the capture of finer local canopy height variations.

6.2.2.4 Canopy height mapping sampling scheme

Finally, we found that the sampling scheme plays a crucial role on the precision of canopy height mapping. In addition, over our study site, forest structure varies at the scale of individual stands at a distance of less than 5000 m. This distance poses problems for GLAS which has a sampling scheme in the tens of kilometres. Therefore, it would be beneficial if future spaceborne LiDAR can increase it sampling frequency. ICESat-2 which is estimated to launch in 2017 will have a higher sampling frequency, as it will use a 6 beams laser, arranged into 3 pairs with a distance of 3.3 km between each pair. The laser will also take measurements every 140 m along the track. The system will also feature denser cross-track sampling in comparison to ICESat-1.

6.2.3 Above-ground biomass estimation

The main objective of this thesis was to map forest canopy heights across French Guiana. This objective was met with interesting results. However, canopy heights can be used to estimate other important forest resources such as above-ground carbon stocks (ACD) or above-ground biomass (AGB), either directly or indirectly. In General, AGB is estimated by applying allometric relations for each individual tree by using information about the stand structure, such as its height, diameter at breast height (DBH), and wood density (WD). However, in this study canopy height estimates refer to the maximum canopy height of plots and not individual trees. In addition, parameters such as DBH and WD are not possible to extract directly with LiDAR measurements. While many studies have demonstrated the strong correlation between AGB and LiDAR canopy height estimates

([44]; [39]) and thus developed relations to incur aboveground biomass from LiDAR canopy height estimates. The relations developed presented large uncertainties in large-scale AGB estimation since AGB estimation is not only related to tree height and is site dependent. Therefore in order to estimate plot aggregated AGB, several approaches that account for ancillary variables other than LiDAR canopy height estimates should be used.

Drake *et al.* [39] found that biomass estimates are more related to canopy height at medium energy of the LiDAR waveform (RH50) than it is related to the height at full energy, due to RH50 being more sensitive to changes in both canopy density and vertical arrangement of canopy elements. However, Ni-Meister *et al.* [68] found that RH75 is more correlated to biomass than RH50 in their study area. According to Ni-Meister *et al.* [68] this might be due to topography. Therefore it is worth studying the relation between LiDAR canopy heights at different energies and their correlation to biomass in our study area. Surface topography parameters could also be added to the AGB estimation model in an attempt to improve the estimation precision.

In addition, since biomass correlation with canopy heights is site dependent, the inclusion of forest types in the biomass estimation might be beneficial. Forest type information can be easily extracted from the existing land cover maps derived from optical remote sensing data. Tree cover, could also be used to improve the precision of biomass estimates. Indeed, according to Ni-Meister *et al.* [68]; models using RH100 or RH50 with forest cover perform better than using only RH50. This is due to RH50 with tree cover being highly related to wood volume. Moreover, allometric equations are also dependent on stand age ([130]; [131]; [132]). Therefore, information on stand age should also be used in the biomass estimation models to obtain optimal results.

Asner *et al.* [40]; proposed Plot-aggregate allometry for the estimation of aggregate ACD or AGB using LiDAR canopy height estimates. In their study they postulated that if forest structure and biomass organization follow consistent scaling patterns, simple plot level variables could capture the same information about AGB in comparison to field inventories. Next, since parameters such as DBH and WD are not directly estimated by LiDAR, relations were found between LiDAR canopy heights and these parameters.

Therefore, plot AGB can now be estimated using plot aggregate allometry and using only LiDAR canopy height estimates.

Finally, further investigation is required to have a better understanding of the links between AGB, ancillary variables (forest cover, forest type, forest topography) and other variables such those used in chapter 4 for canopy height mapping, and the allometric relations used in Asner *et al.* [40]; in order to develop a more improved aboveground biomass estimates from LiDAR. Ideally, an approach similar to the one used in chapter 4 should be approached.

RESUME

7.1 Introduction

Le changement climatique attire l'attention de la communauté scientifique depuis les dernières décennies. Cette attention a été suivie par un intérêt pour la quantification de la biomasse terrestre (Above Ground Biomass - AGB), non seulement pour comprendre ses effets sur le cycle global du carbone, mais pour atténuer les effets du réchauffement global par l'intermédiaire de la conservation des stocks et des puits de carbone. Actuellement, les méthodes d'estimation d'AGB existantes à partir des données de télédétection sont limitées soit en termes de leur sensibilité à des niveaux d'AGB élevés (saturation du signal de télédétection à un faible niveau de biomasse en utilisant principalement les données radar et optiques), soit en termes de couverture spatiale (couverture horizontale limitée en utilisant les données LiDAR).

Compte tenu de ces limitations, les études utilisent généralement les relations allométriques pour relier les caractéristiques d'une forêt (la hauteur de l'arbre, le diamètre à hauteur de poitrine, et la densité du bois) à sa biomasse (par exemple [39]; [40]; [42]; [133]). Cependant, l'une des variables les plus importantes dans les relations allométriques et qui peut être estimée à partir des techniques de télédétection est la hauteur de la canopée. En effet, plusieurs allométries reposent uniquement sur la hauteur de la canopée pour

l’estimation de la biomasse ([39]; [40]). En outre, des études ont montrées que l'utilisation de la hauteur de la canopée augmente la précision de l'estimation de la biomasse (par exemple [42]; [41]; [40]; [44]; [24]).

En plus de l'importance de la hauteur des arbres dans l’estimation d’AGB, la hauteur des arbres est également intéressante elle-même pour répondre aux questions écologiques sur les déterminants de la plante, la structure de la forêt, et la dynamique forestière. La hauteur des arbres est importante dans les décisions d'aménagement forestière et d'évaluer les ressources en bois. En outre, la hauteur des arbres peut avoir des effets directs sur les modèles et processus microclimatiques ([49]). En effet, le microclimat est modifié non seulement par les conditions météorologiques locales, mais aussi par la végétation elle-même. La hauteur de la forêt contrôle la qualité, la quantité, la distribution spatiale et temporelle de la lumière et de l'énergie atteignant les sous-étages de la canopée et le sol. Elle influence également les précipitations locales et le mouvement de l'air. Ces facteurs combinés déterminent jusqu'à un certain point l'humidité de l’air et la température.

Actuellement, la meilleure technique pour l'estimation de la hauteur de la canopée par télédétection est l'utilisation de la technologie LiDAR. Comme l'utilisation d'autres technologies telles que le radar ne donne pas des résultats satisfaisants (Tableau 1). De nombreuses études utilisent le LiDAR aérien ou satellitaire pour l'estimation de la hauteur de la canopée (par exemple [24]; [44]; [45]; [46]; [65]; [72]). Tandis que l’estimation de la hauteur de la canopée à partir des données de LiDAR aéroportés peut être très précises (erreur quadratique moyenne EQM mieux que 2 m [134]), le LiDAR satellitaire a une précision inférieure sur l'estimation de la hauteur de la canopée comprise entre 2 m et 10 m variant en fonction des caractéristiques de la forêt (par exemple, [44]; [62]; [63]; [64]; [65]; [72]). Cependant, le LiDAR aéroporté est limité dans le domaine horizontal (couverture spatiale limitée), alors que le LiDAR satellitaire offre une couverture mondiale, mais avec une densité relativement faible (environ 0,51 points/ km² sur la Guyane française par exemple) avec un échantillonnage spatial inhomogène (lignes d'échantillonnage le long des orbites du satellite). Par conséquent, la fusion des données LiDAR (spatiaux et/ou aéroportés) avec une autre source de données est essentielle afin de créer des carte de hauteurs avec une couverture complète et une bonne précision (par exemple, [47]; [65]; [66]).

Hudak *et al.* [66] ont testé différents modèles non-spatiales et spatiales (linéaire, krigeage, et co-krigeage) et différentes stratégies d'échantillonnage pour cartographier les hauteurs de la canopée à partir d'estimations de hauteurs issues de données LiDAR aéroporté et de données optiques LANDSAT Enhanced Thematic Mapper (ETM +) sur un site d'étude de 200 km^2 dans l'ouest de l'Oregon (USA). Dans leur étude, les modèles de régression ont maintenu la structure de la hauteur des arbres, mais ont montré un biais pour les arbres les plus petits et les plus grands. En revanche, le krigeage et les modèles de co-krigeage présentent de meilleurs résultats en termes de précision par rapport au modèle de régression. Néanmoins, la meilleure méthode pour l'estimation de la hauteur de la canopée est la méthode appelée régression-krigeage (Hengl *et al.* [98]). Cette méthode conserve la structure de la hauteur des arbres tout-en améliorant la précision. Enfin, la stratégie d'échantillonnage joue un rôle majeur sur la précision de l'estimation puisque l'écart-type sur l'estimation de la hauteur de la canopée varie de 5,5 m à 10,0 m respectivement pour un échantillonnage de 250 et 2000 m.

Lefsky *et al.* [47] ont créé une carte de la hauteur de la canopée estimée à partir du radiomètre spectral de moyenne résolution MODIS et du LiDAR satellitaire GLAS. Leur technique a montré une estimation de la hauteur de la canopée avec une erreur quadratique moyenne (EQM) sur l'estimation des hauteurs de la canopée de 5,9 m et un coefficient de corrélation (R^2) de 0,67. Enfin, une étude plus récente menée par Simard *et al.* [65] a amélioré les résultats de Lefsky *et al.* [47] pour la cartographie mondiale de la hauteur de la canopée en utilisant d'autres données auxiliaires tels que la moyenne de la précipitation annuelle, les précipitations saisonnières, la moyenne annuelle de la température, la température saisonnières, les données à partir d'un modèle numérique de terrain (MNT) et le pourcentage du couvert forestier à partir de MODIS. Leur carte mondiale de la hauteur de la canopée validée par des mesures *in-situ* a montré une précision sur l'estimation de la hauteur avec une EQM de 6,1 m (R^2 de 0,5).

Jusqu'à présent, les approches utilisant les méthodes d'interpolation et les données LiDAR (satellitaires et aéroportées) pour la spatialisation de la hauteur de la canopée à l'échelle régionale n'ont pas été pleinement étudiées. Etant donné que les cartes de la hauteur de la canopée existantes ont besoin d'être plus précises pour être plus bénéfiques pour les écologues forestiers et les gestionnaires des ressources de la forêt.

La plupart des études ont utilisé une technique d'interpolation unique pour la production de leur carte de hauteurs de la canopée. Par conséquent, la première question est: serait-il avantageux de mélanger différentes techniques d'interpolation, avec les deux aspects non-spatiaux et spatiaux pour améliorer la précision de la cartographie de la hauteur de la canopée?

Deuxièmement, la question du coût d'acquisition des données LiDAR aéroporté sur une échelle régionale, et la disponibilité globale des données LiDAR satellitaire avec une densité inférieure posent une nouvelle question. Quel est l'impact de la densité d'échantillonnage spatial (en particulier pour le LiDAR satellitaire) sur la précision des cartes de hauteurs de la canopée?

Dans la présente étude, la hauteur de la canopée à partir du LiDAR aéroporté et satellitaire en combinaison avec des données auxiliaires ont été utilisées pour créer une carte de hauteur de la canopée avec une résolution de 250 m couvrant toute la Guyane française (superficie de 83.534 km^2). Les données utilisées sont les estimations de hauteur de la canopée à partir du LiDAR aéroporté couvrant les 4/5 de la Guyane française, ainsi que les estimations de la hauteur de la canopée à partir du LiDAR satellitaire GLAS/ICESat. Les données auxiliaires, disponibles au niveau mondial comme l'indice d'activité photosynthétique de végétation (données EVI – Enhanced Vegetation Index du capteur MODIS), les précipitations, la topographie (calculés à partir du Modèle Numérique de Terrain - MNT du Shuttle Radar Topographic Mission), la géologie et les différents types de paysage forestier, ont été utilisées. En outre, comme en témoigne Hudak *et al.* [66] sur la corrélation entre la stratégie d'échantillonnage et la précision sur les hauteurs de la canopée, les effets de l'échantillonnage spatial des ensembles de données de référence LiDAR utilisés dans cette étude ont été analysés. Enfin, toutes les cartes de hauteur créées ont été validées à l'aide des estimations de la hauteur de la canopée obtenues à partir d'une base LiDAR aéroporté indépendante.

En plus du rôle de la hauteur des arbres dans l'estimation de la biomasse, la classification des paysages forestiers jouent un rôle important dans les méthodes d'estimation de la biomasse. En effet, de nombreuses études ont montré que les modèles pour l'estimation de

la biomasse sont plus pertinents si l'on inclut les types de forêts ([24]; [42]; [67]; [68]; [69]).

Zheng *et al.* [67] ont constaté que le couplage des métriques forestières acquises à partir des mesures terrain et divers indices dérivés des données Landsat 7 ETM a sensiblement amélioré l'estimation de la biomasse lors de la séparation des forêts hardwood et les forêts de pins. Chave *et al.* [42] ont testé plusieurs modèles pour l'estimation de la biomasse dans les forêts anciennes, sèches, humides, montagneuses, et les forêts de mangroves. Leurs résultats indiquent que l'un des facteurs les plus importants pour l'estimation de la biomasse est le type de forêt. Les résultats ont également indiqués que les meilleurs modèles prédictifs sont aussi dépendants du type de forêt. Ni-Meister *et al.* [68] ont développé un modèle pour l'estimation de la biomasse qui utilise une fusion de données des capteurs LiDAR et optique, et intégrant le type de forêts conifères/feuillus et décidues/sempervirentes. Leurs résultats indiquent que les modèles qui intègrent le type de forêt ont fourni une meilleure estimation de la biomasse par rapport aux modèles qui n'utilisent pas cette variable.

Une description des ensembles des données LiDAR aéroporté et satellitaire utilisées dans cette étude est donnée dans la section 1.2. La section 1.3 présente les méthodes d'estimation de la hauteur des arbres en utilisant le lidar GLAS. La méthodologie utilisée pour la création d'une carte des hauteurs sur toute la Guyane française est décrite dans la section 1.4. La classification des paysages forestiers est présentée dans la section 1.5. Enfin les conclusions et perspectives sont présentées dans la section 1.6.

7.2 Description des jeux de données

7.2.1 Site d'étude

La Guyane française est située sur la côte nord du continent sud-américain; face à l'Océan Atlantique. Elle est bordée par le Brésil à l'est et par le Suriname à l'ouest (cf. Figure 2.1). La superficie de la Guyane française est de 83.534 km^2. La forêt occupe environ 97%, soit environ 80.820 km^2 de sa superficie totale. Le terrain est en grande partie de faible altitude, s'élevant parfois en petites collines et montagnes, avec une altitude maximale de 851 m.

Soixante-huit pourcent (68%) des pentes sont inférieures à cinq degrés, 24% entre cinq et dix degrés et 8,2% plus de dix degrés (dérivés des élévations MNT SRTM à 90 m). Une forêt tropicale dense domine la Guyane en dehors de la plaine côtière et couvre plus des quatre cinquième de la superficie totale. D'autres types de végétation existent aussi comme les savanes, les marais et l'agriculture (pâturages, manioc, maraîchage). La Guyane française a un climat équatorial avec deux saisons principales, la saison sèche, de août à décembre, et la saison des pluies ou humide de décembre à juin.

7.2.2 Base de données LiDAR satellitaire

Les formes d'ondes LiDAR provenant du système GLAS/ICESat acquises dans le proche infrarouge à 1064 nm, de 2003 à 2009, sur la Guyane française ont été utilisées dans cette étude. Au cours de ses années de fonctionnement (2003-2009), GLAS a fonctionné avec des cycles orbitaux qui se répétaient tous les 57 à 197 jours pour un total de 18 missions. La surface mesurées par GLAS (l'empreinte) a un diamètre compris entre 50 et 100 m (moyenne de 70 m), et les formes d'onde ont été acquises tous les 175 m le long de la trace. Les formes d'onde GLAS ont été numérisées en 544 ou 1000 échantillons avec une résolution verticale de 1 ns (15 cm). Par conséquent, elles peuvent mesurer des structures verticales jusqu'à 81,6 m et 150 m, respectivement. La précision verticale de GLAS a été estimée à environ 3,2 cm en moyenne sur les zones plates [70].

En Guyane française, GLAS a acquis plus de 100 000 formes d'onde entre 2003 et 2009 (cf. Figure 2.1). Cependant, toutes ces formes d'onde ne sont pas adaptées pour l'estimation de la hauteur des arbres. Certaines d'entre elles sont perturbées par la diffusion atmosphérique, les nuages, etc. Pour supprimer ces formes d'onde, plusieurs filtres ont été appliqués [72]. Après le filtrage, 47 348 formes d'onde ont été conservées pour la suite de l'étude

7.2.3 Données du radiomètre spectral à moyenne résolution MODIS

Le capteur MODIS embarqué sur les satellites Terra et Aqua possède un total de 36 bandes spectrales dont sept prévues spécifiquement pour les applications terrestres avec des résolutions spatiales qui vont de 250 m à 1 km. Le jeu de données MODIS utilisé dans cette étude comprend dix années (du 1er janvier 2003 au 31 décembre 2012) d'indice de végétation amélioré (EVI, MOD13A1). Les données EVI caractérisent efficacement les

états et processus biophysiques et biochimiques des surfaces végétalisées. Une période de 10 ans a été utilisée pour une meilleure synchronisation avec les données GLAS (2003-2009).

En utilisant les séries temporelles EVI, six cartes ont été préparées: les valeurs minimales, moyennes et maximales des données EVI (MIN_EVI, MEAN_EVI et MAX_EVI respectivement) (cf. Figure 2.3), ainsi que les trois premières composantes issues de l'analyse en composantes principales des données EVI (PC1, PC2 et PC3). L'indice de végétation normalisé (NDVI) et l'indice de végétation amélioré (EVI) sont souvent utilisés dans de nombreuses applications qui étudient les paramètres biophysiques. Cependant, la relation entre eux varie en fonction des caractéristiques de la zone d'étude. Une forte corrélation entre les indices de végétation et de la biomasse verte a été rapportée par Hardisky *et al.* [135]; en revanche, d'autres études ont rapportées peu de corrélation entre les deux ([136]). Enfin, Freitas *et al.* [137] et Pascual *et al.* [138] ont constaté une forte relation entre la hauteur de la canopée et les indices de végétation. C'est pourquoi, il pourrait être intéressant d'inclure l'EVI dans les modèles d'estimation de hauteur de la canopée.

7.2.4 Données issues du Modèle Numérique de Terrain MNT SRTM

Le modèle numérique de terrain (MNT) acquis par la NASA (SRTM) a également été utilisé dans cette étude. C'est le produit avec une résolution de 90 m à l'échelle du Globe qui a été utilisée. Bourgine et Baghdadi [73] ont trouvé que la précision des données d'élévation SRTM a été évaluée à environ 10 m (écart type d'erreur) en Guyane française. Le jeu de données MNT SRTM et ses cartes dérivées ont été analysés puisque la topographie locale et le drainage sont importants pour l'ancrage des arbres et la dynamique de la forêt.

D'après le MNT SRTM, trois produits ont été dérivés:

(1) Une carte de pente (Slope) (cf. Figure 2.4a), qui est calculée en utilisant la variation maximale d'élévation de la distance entre chaque cellule du MNT et ses huit voisins dans une fenêtre de 3x3. La pente a été largement utilisée dans de nombreuses études pour la correction des estimations de la hauteur des arbres à partir des données LiDAR aéroporté et satellitaire ([44]; [63]; [71]). L'intégration de la pente du sol dans les modèles

d'estimation de la canopée augmente la précision des estimations car elle élimine une partie de la hauteur ajoutée par la pente ([63]).

(2) Une carte de la rugosité de la surface (Roughness) (cf. Figure 2.4b). La rugosité correspond à l'écart type des élévations dans une fenêtre 3x3. Les zones ayant un faible écart type représentent des zones des hauteurs d'arbres plus homogènes, tandis qu'un écart-type fort représente des hauteurs d'arbres plus hétérogènes. L'indice de rugosité a été utilisé dans les études de Carabajal et Harding [70] et Fayad *et al.* [105] et les résultats montraient une bonne corrélation avec la métrique LiDAR Wext des formes d'onde GLAS (mesure représentant une estimation de la hauteur des arbres).
(3) Enfin, une carte de la zone de drainage en échelle logarithmique (ln_drain) (cf. Figure 2.4c). La zone drainée mesure la surface du bassin hydraulique qui s'écoule à travers une cellule en utilisant le modèle D8. Une valeur faible indique des cellules situées sur une crête en amont, tandis que les valeurs les plus élevées indiquent des cellules situées en aval dans les vallées.

7.2.5 Carte géologique

La géologie est un déterminant important de la formation du sol, conditionnant ses propriétés chimiques et physiques, qui affecte la croissance des arbres et d'autres paramètres forestiers. Une carte de substrat géologique (GEOL) produite par le BRGM ([76]) a été donc utilisée dans cette étude (cf. Figure 2.5a). La carte a été simplifiée pour retenir seulement les cinq plus grandes formations de roche: sédiments récents, roche sédimentaire volcanique, granites, gabbros et gneiss.

7.2.6 Carte des types de paysage forestier

Une carte des types de paysage forestier développée par Gond *et al.* [77] avec une résolution d'un km a été utilisée (cf. Figure 2.5b). Dans cette carte, 33 types de paysage forestier (LTs) utilisant des images SPOT-VEGETATION ont été interprétés. Cinq classes du total des 33 classes ont été utilisées dans cette étude puisqu'elles occupent environ 78% de la forêt dans cette zone. Les LTs peuvent être décrites comme suit:
(1) LT8 représente une forêt dense avec un couvert fermé La forêt y est composée de petites couronnes et de petites ouvertures mélangées dans un couvert régulier avec des arbres ayants sensiblement la même hauteur.

(2) LT9 est une forêt avec un couvert fermé d'arbres de même hauteur dominée par des émergents bien développées sans grandes ouvertures.
(3) LT10 est une forêt irrégulière où les arbres ont des hauteurs et des diamètres de couronne différents avec de grandes ouvertures. Cette forêt irrégulière est mélangée avec une forêt ayant un couvert fermé et dominé par des couronnes bien développées sans grandes ouvertures. LT10 est aussi mélangé avec des forêts de liane.
(4) LT11 est semblable à LT10 avec plus de forêts de liane et des couverts non-forestier.
(5) LT12 est une forêt ouverte associée aux marécages et la forêt de bambou.
L'ensemble de données de LTs a été choisi pour sa corrélation avec la hauteur des arbres.

7.2.7 Carte de précipitation

Les données de précipitations de la mission de la NASA pour la mesure des précipitations tropicales (TRMM) ont été utilisées (cf. Figure 2.5c). Cette mission a été lancée en 1997 pour la mesure et le suivi des pluies tropicales. Les données TRMM utilisées dans cette étude représentent la précipitation moyenne quotidienne au cours des 10 dernières années (2003 à 2013) avec une résolution de 8 km (Rain). Des études récentes suggèrent une relation entre la hauteur de la canopée maximale dans les vieux peuplements et le cumul des précipitations annuelles ([139]). Lorsque le rapport des précipitations augmente, la hauteur maximale des arbres augmente aussi.

7.3 Estimation de la hauteur des arbres à partir des données GLAS

Les modèles d'estimation de la hauteur des arbres basés sur les données de formes d'ondes LiDAR GLAS peuvent être divisés en deux catégories: la méthode directe et les modèles statistiques. La méthode directe permet d'estimer la hauteur de la canopée dans les zones avec un faible relief en utilisant la différence d'altitude entre le début du signal GLAS et le pic du sol. Cependant, sur les zones en pente, la méthode directe surestime la hauteur des arbres en raison de la hauteur supplémentaire introduite par la pente. Pour supprimer les effets de la pente, des modèles statistiques ont été développés en utilisant des métriques extraient de GLAS et d'un MNT. Tandis que les métriques développées dans des études précédentes ont réussi à augmenter la précision des modèles d'estimation de la hauteur des arbres (par exemple, [44]; [62]; [64]; [71]), ces modèles avaient leurs propres problèmes.

En effet, afin d'utiliser ces métriques pour une meilleure estimation de la hauteur des arbres, la position exacte du pic du sommet de la canopée et le pic du sol est souvent nécessaire. Sur les zones avec une végétation dense comme les forêts tropicales, l'extraction de la position exacte du pic du sommet de la canopée et le pic du sol est particulièrement difficile en utilisant un processus automatisé car les formes d'onde LiDAR ne présentent pas souvent des pics distinctifs [62]. L'extraction de ces métriques manuellement est toujours possible, mais devient inefficace lorsqu'il s'agit d'un grand nombre de formes d'onde GLAS.

Les objectifs de cette section sont de tester plusieurs modèles couramment utilisés pour l'estimation de la hauteur des arbres, qui utilisent des métriques dérivées des formes d'onde GLAS et du MNT-SRTM, et de tester deux techniques nouvelles dans le domaine LiDAR appliqué à la forêt: l'analyse en composantes principales (ACP) des formes d'onde et la technique Random Forest (RF) qui va servir comme modèle de régression pour estimer la hauteur des arbres. Le but d'utiliser l'approche ACP est d'éliminer la nécessité d'utiliser les métriques extraites de GLAS dans les modèles d'estimation de la hauteur des arbres. En effet, l'extraction de certaines métriques peut présenter des erreurs, en particulier dans les forêts très denses, comme dans le Guyane française à cause de la difficulté à identifier la position du sol. Pour la régression RF, les mêmes métriques que celles utilisées dans les modèles de régression GLAS seront tout d'abord utilisées. Les résultats de tous les modèles testés dans cette section seront validés par rapport aux estimations de la hauteur des arbres obtenus à partir d'un ensemble de données de LiDAR aéroporté indépendant.

7.3.1 Contexte de l'estimation de la hauteur des arbres en utilisant GLAS

L'estimation de la hauteur des arbres en utilisant la méthode directe est effectuée en utilisant tout simplement la différence entre le début du signal de la forme d'onde (Hb) et le pic du sol (Hg). La méthode directe estime la hauteur des arbres avec une bonne précision sur les zones plates (par exemple [72]). Cependant, sur les zones en pente, à la fois le pic du sol et celui de la végétation s'élargissent et diminuent en intensité. Le pic identifié comme le pic du sol ne représentera plus seulement le sol, mais un mélange d'objets et du terrain ([71]; [70]). En fait, sur un terrain en pente l'étendu de la forme d'onde 'Wext' va augmenter en fonction de la pente et de la taille de l'empreinte [10]. Cette augmentation

se traduira par une détection plus précoce du début du signal et cela provoquera une surestimation de la hauteur des arbres. Pour corriger l'effet de la pente sur le signal GLAS, plusieurs études ont développés des modèles qui utilisent des paramètres issus des signaux GLAS (l'étendue de la forme d'onde "Wext", l'étendue du bord antérieur "Lead" et l'étendue du bord postérieur "Trail") (cf. Figure 2.2) afin de mieux estimer la hauteur des arbres (par exemple : [44]; [62]; [71]; [70]). Finalement, afin d'analyser la précision de l'estimation de la hauteur des arbres à l'aide du modèle 'Random Forest', plusieurs configurations de métriques ont été testées. Dans cette étude, les principaux modèles testés sont résumés dans le tableau (cf. Tableau 3.1).

La comparaison entre les estimations de la hauteur des arbres à partir des formes d'onde GLAS en utilisant la méthode directe et les estimations de la hauteur des arbres de l'ensemble de données du LiDAR aéroporté a montré une forte erreur quadratique moyenne (EQM) sur l'estimation de la hauteur des arbres (7,9 m) et un faible R^2 de l'ordre de 0,50 (cf. modèle Id 1, Tableau 3.1, Figure 3.5a). Ce résultat peut être expliqué par le fait que la plupart des empreintes GLAS tombent sur une zone avec une pente entre 5° et 10°. En outre, les résultats des modèles de régression avec validation croisée ont montré que les modèles de régression qui utilisent le Trail (cf. modèle Id 4-8, Tableau 3.1, Figure 3.5b) fournissent une estimation de la hauteur des arbres légèrement meilleure en comparaison aux modèles qui utilisent soit le Lead, l'indice de terrain à partir du MNT-SRTM (TI) ou le Lead et TI (EQM compris entre 3,7 et 4,0 m, R^2 entre 0,79 et 0,81). La contribution du Lead semble être faible par rapport au Trail lorsqu'on estime la hauteur maximale des arbres. En effet, le modèle Id 7 qui utilise le Trail avait de meilleurs résultats par rapport aux modèles Id 3 qui utilisent le Lead (cf. Tableau 3.1). En outre, l'utilisation du 'Terrain Index' (TI) calculé à partir du MNT dans les modèles de régression produit une précision sur l'estimation de la hauteur des arbres plus basse (EQM = 4,9 m et R^2 = 0,72).

Les résultats de l'estimation de la hauteur des arbres en utilisant les métriques à partir de GLAS et le RF, ont montrés que la meilleure configuration pour estimer la hauteur de la canopée est la configuration qui utilise toutes les métriques: Wext, Lead, Trail, et TI (cf. modèle Id 13, tableau 3.1). La différence entre les estimations de la hauteur des arbres de GLAS et les hauteurs obtenues à partir des données LiDAR aéroporté (données de référence) pour la configuration utilisant toutes les métriques, a montré une EQM de 3,4

m, et un coefficient de détermination R^2 de 0,82. En outre, le test de l'importance des variables a montré que la hauteur des arbres provenant de GLAS est mieux expliquée en utilisant la Wext avec un facteur d'importance presque trois fois plus élevé en comparaison avec les trois autres métriques (Trail, Lead, TI), qui ont presque la même importance. D'autres modèles de régression qui utilisaient Wext, Lead, et TI ou Wext et Lead ou Wext et TI (cf. modèles Id 14, 15, 16, Tableau 3.1) ont montré des précisions semblables sur l'estimation de la hauteur des arbres (EQM environ 3,6 m). L'estimation de la hauteur d'arbres provenant de GLAS et utilisant uniquement la Wext a montré une EQM de 4,4 m avec un R^2 de 0,73. Figue 3.6 montre des exemples de la comparaison entre les estimations de la hauteur des arbres à partir de GLAS et les estimations de référence de la hauteur obtenues à partir de données LiDAR aéroporté.

7.3.2 Techniques proposées pour l'estimation de la hauteur des arbres

La section précédente a présenté un certain nombre de modèles de régression élaborés dans plusieurs études pour l'estimation de la hauteur des arbres. Cependant, ces modèles exigent plusieurs métriques dérivées des empreintes GLAS, comme le pic du sol, pic du sommet, Lead, Trail, et les métriques dérivées du MNT SRTM, tels que l'indice de terrain (TI). En outre, l'extraction de certaines métriques à partir des formes d'onde GLAS, tels que la position du pic du sol, peut être erronée, en particulier dans les forêts denses. En fait, la pénétration de la forme d'onde dans la canopée dans les zones de végétation dense est parfois insuffisante pour atteindre le sol. Ainsi, soit le signal LiDAR n'a pas attient le sol ou bien il a atteint le sol mais le retour est trop faible pour une détection fiable. Ces difficultés dans la détection du pic sol affectent l'estimation du Wext, et par conséquent, l'estimation de la hauteur des arbres. Ainsi, un modèle statistique pour l'estimation de la hauteur basé uniquement sur les valeurs de la forme d'onde peut être une alternative intéressante. Une analyse en composantes principales des formes d'onde GLAS a été menée. Ensuite des modèles de régression linéaire et la technique 'Random Forest' ont étés construits pour l'estimation de la hauteur de la canopée avec les composantes principales (PC).

L'estimation de la hauteur des arbres en utilisant le RF avec les PCs a été effectuée en utilisant différentes configurations des PCs utilisées. En utilisant les 13 premiers PCs avec la technique 'Random Forest' a abouti à une meilleure précision de l'estimation de la

hauteur des arbres (EQM = 4,7 m, R^2 = 0,7) par rapport au modèle de régression linéaire qui a utilisé les 13 premiers PCs (EQM = 5,9 m, R^2 = 0,52). Le test de l'importance des variables a montré que la hauteur des arbres GLAS est mieux expliquée en utilisant PC1, PC2, PC4 et PC11 (variance 62,4%). La seule utilisation de ces quatre PCs dans RF a montré un résultat similaire (EQM = 4,8 m, R^2 = 0,69). Ensuite, l'incorporation de la Wext avec les 13 premières composantes principales a considérablement amélioré la précision de l'estimation de la hauteur des arbres (EQM = 3,6 m, R^2 = 0,83) en comparaison avec les régressions RF sans Wext. En plus, ce résultat est du même ordre que celui obtenu en utilisant une régression linéaire avec les 13 premières PCs et Wext (EQM = 3,8 m). L'utilisation des variables les plus importantes (Wext, PC1, PC2, PC4 et PC11) dans la régression RF, a donné des résultats similaires, avec une erreur quadratique moyenne de 3,6 m et un R^2 de 0,82. Enfin, en remplaçant la Wext par la classe de Wext (WC, avec WC1 pour Wext < 20m ; WC2 pour Wext entre 20 et 40 m et WC3 pour Wext > 40 m) avec les 13 premières PCs dans RF, l'estimation de la hauteur des arbres a montré des résultats similaires (EQM = 3,7 m, R^2 = 0,81). Des résultats similaires ont été observés lorsque l'on ne retient que les variables les plus importantes avec la classe de Wext (WC, PC1, PC2, PC4 et PC11), avec une erreur quadratique moyenne de 3,7 m et un R^2 de 0,81. La Figure 3.9 montre la comparaison entre les hauteurs de la canopée à partir de GLAS en utilisant la technique de RF sur les PCs, Wext et WC et les hauteurs de référence obtenues à partir des données LiDAR aéroporté.

7.4 La spatialisation de la hauteur des arbres LiDAR

Afin d'estimer la hauteur des arbres sur une grille régulière de 250 m x 250 m et non plus au niveau des formes d'onde LiDAR (GLAS ou LiDAR aéroporté) une procédure en cinq étapes a été menée, en se basant sur la relation statistique et spatiale entre les estimations de la hauteur des arbres à partir du LiDAR et des variables auxiliaires (GEOL, LT, Rain, Slope ...) en utilisant des méthodes d'estimation empiriques largement utilisées: régressions par Random Forest, krigeage ordinaire, et régression krigeage (par exemple [65]; [66]; [96]; [97]; [98]). Les cartes produites ont une résolution de 250 m x 250 m correspondant à la résolution de la majorité des variables auxiliaires utilisées.

7.4.1 Contexte sur la technique régression-krigeage

Nous avons cartographié la hauteur des arbres à partir de données LiDAR à une résolution de 250 m en utilisant la technique de régression krigeage (RK). RK est une technique spatiale de prédiction qui combine la valeur de régression des prédicteurs (variables auxiliaires) et le krigeage des résidus de régression (hauteurs de référence de la canopée - hauteurs de la canopée estimée par RF) ([98]). Cette technique a été développée principalement pour tenir compte de la corrélation entre les variables environnementales et la qualité du fitting insatisfaisante du modèle de variance spatiale de l'ensemble de données ([99]) empêchant ainsi la stationnarité des variables auto corrélées étudiés. Pour le modèle de régression, nous allons utiliser la technique RF, et pour la régression des résidus nous allons utiliser le krigeage ordinaire.

7.4.2 La cartographie de la hauteur des arbres en utilisant la régression krigeage

Pour créer les cartes de la hauteur des arbres, les variables prédictives qui expliquent le mieux la hauteur des arbres ont d'abord été choisies. Deux ensembles de données de référence pour la hauteur des arbres à partir du LiDAR ont été utilisés séparément: l'ensemble de données GLAS, et l'ensemble des données LiDAR aéroportés (AD). Les résultats ont montré que les prédicteurs qui expliquent le mieux la hauteur des arbres sont les mêmes pour les deux jeux de données de référence. Les prédicteurs sont les suivants: La rugosité du terrain (Rug), la valeur moyenne de l'EVI (EVI_AVG), la carte géologique (Geo), la valeur moyenne des précipitations (Rain), et la pente du terrain (Slope). Ensuite, la technique de régression RF a été utilisée pour modéliser la hauteur des arbres à l'aide des variables prédictives les plus pertinentes. Un modèle de régression RF a été développé pour chaque ensemble de données LiDAR afin de créer des cartes spatialisées pour la hauteur de la canopée pour toute la Guyane. Les résultats ont montré que toutes les cartes de la hauteur des arbres ont présentées une précision semblable quelle que soit la précision des hauteurs d'arbres de référence utilisées (cf. Figure 4.2, EQM environ 6 m). Afin d'améliorer la précision des cartes obtenues pour la hauteur des arbres, les résidus (la hauteur de référence des arbres – la hauteur estimée des arbres) obtenus à partir de chaque ensemble de données LiDAR ont été krigés. Afin de kriger les résidus, le semivariogramme des résidus des deux jeux de données (GLAS et LiDAR aéroporté) a été ajusté. Le palier, la portée et la pépite obtenus ont été similaires. Ensuite, chaque carte de résidu krigé a été

ajoutée à la carte de hauteur des arbres correspondante obtenue à partir du modèle de régression RF. Les résultats ont montré une amélioration de la précision des cartes de la hauteur des arbres. Cependant, cette amélioration de la précision varie entre la carte de hauteur de la canopée utilisant le résidu de hauteur provenant des données GLAS (EQM de 4,2 m avec RK, cf. Figure 4.5a) et celle obtenue en utilisant l'ensemble de données AD (EQM de 1,8 m avec RK) (cf. Figure 4.5b). Cette différence de précision entre GLAS et AD est due à la densité des points, et l'espacement des lignes de vol pour chaque jeu de données. En effet, pour l'ensemble de données AD, les estimations de la hauteur des arbres sont réparties sur des lignes de vol avec une distance moyenne d'environ 500 m, tandis que les estimations de la hauteur des arbres pour l'ensemble de données GLAS sont répartis sur les lignes de vol avec une distance moyenne de 20 km

7.4.3 Relation entre l'espacement des lignes de vol LiDAR et la précision de la hauteur des arbres krigée

Dans cette section la précision des cartes de la hauteur des arbres obtenues par krigeage des données LiDAR à différentes densités de points LiDAR a été évaluée. Le but est d'analyser l'impact de l'espacement des lignes de vol LiDAR sur la précision de la carte des hauteurs des arbres, en créant à partir du jeu de données d'origine AD plusieurs sous-ensembles de données LiDAR avec différents espacements de lignes de vol (espacement de 5, 10, 20, 30, 40, et 50 km). Afin de créer les cartes pour la hauteur des arbres en utilisant les sous-ensembles de la base AD, les meilleures variables prédictives ont tout d'abord été sélectionnées pour être utilisées dans les régressions RF. Les résultats indiquent que pour les sous-ensembles AD, les variables prédictives qui expliquent le mieux la hauteur des arbres étaient les mêmes que celles utilisées avec les jeux de données GLAS et AD. En plus, les résultats ont également montré que la précision des cartes produites pour la hauteur des arbres à l'aide des régressions RF avec les sous-ensembles AD était dans le même ordre de grandeur que les cartes de hauteur de la canopée obtenues avec les deux ensembles de données de référence (GLAS et AD) (cf. Tableau 4.1). Pour ces sous-ensembles, l'erreur quadratique moyenne sur l'estimation de la hauteur des arbres se situe entre 5,7 et 6,2 m (R^2 entre 0,60 et 0,65). Afin d'ajouter les résidus krigés aux cartes de la hauteur des arbres par RF, le semivariogramme des résidus pour chaque sous-ensemble AD a été ajusté (cf. Figure 4.7). Le palier, la portée et la pépite obtenus ont été similaires à ceux obtenus avec les jeux de données GLAS et AD. En y ajoutant les résidus krigés correspondant à chacun

des sous-ensembles AD, la précision sur les cartes de la hauteur des arbres a augmentée comme prévu (cf. Tableau 4.1). Cette amélioration de la précision semble être corrélée négativement avec l'espacement des lignes de vol des sous-ensembles AD. Pour les sous-ensembles AD_5 et AD_10 (espacement de 5 et 10 km respectivement), la précision sur l'estimation de la hauteur des arbres était similaire aux résultats obtenus avec l'ensemble de données AD (EQM = 1,8 m, R^2 = 0,94). Toutefois, pour les sous-ensembles AD_20, AD_30, AD_40 et Ad_50, la précision sur l'estimation de la hauteur des arbres a diminué d'une erreur EQM = 3,3 m pour LD_20 à une erreur de 4,8 m pour LD_50.

7.5 Le potentiel du couplage GLAS et SRTM pour la discrimination des types de paysage forestier

Dans cette section, les formes d'ondes acquises par GLAS ont été utilisées pour analyser la différence entre les élévations GLAS et celles du MNT SRTM, dans le but de discriminer les cinq principaux types du paysage forestiers (LTs) de la Guyane. Les LTs ont été délimités par Gond *et al.* [77] en utilisant un processus d'experts mélangeant des analyses terrain et d'images à basse résolution [77]. Tout d'abord, deux différences ont été calculées: (1) la profondeur de pénétration définie comme la différence entre l'élévation du pic du sommet à partir d'une forme d'onde GLAS et l'élévation du MNT, et (2) la différence entre l'élévation du centroïde d'une forme d'onde GLAS et l'élévation du MNT. La classification des cinq types de paysage forestiers (LTs) en utilisant les données GLAS et MNT a été évaluée en utilisant la technique RF. Cette classification a été réalisée selon la profondeur de pénétration, la différence entre l'élévation du centroïde d'une forme d'onde GLAS et l'élévation du MNT, et la hauteur des arbres à partir de GLAS (Hc) (différence entre le pic du sol et le début du signal), et l'indice de rugosité issu du MNT (R). Enfin, en utilisant les changements qui se produisent au signal GLAS à différentes saisons, le potentiel de discrimination des LTs en utilisant ces changements a été étudié.

7.5.1 Classifications des empreintes GLAS

Une classification des empreintes GLAS basées sur la technique Random Forest (RF) a été réalisée en utilisant la profondeur de pénétration (la différence entre l'élévation du pic du sommet à partir de GLAS et l'élévation du MNT), la différence entre l'élévation du centroïde de la forme d'onde GLAS et l'élévation du MNT, la hauteur de la canopée (H_c)

et l'indice de rugosité (R). La technique 'Random Forest' est connue pour s'adapter à des problèmes de classification ayant des caractéristiques spatiaux complexes et est de plus en plus utilisée par la communauté de la télédétection pour la classification de l'occupation des sols [119]. Les résultats de la discrimination ont montrés que tous les LTs ont été bien classés avec une précision entre 78,4% à 97,5% et la discrimination a été réalisée avec une précision globale de 81,3% (coefficient Kappa de 0,75) (cf. Tableau 5.2). Le test d'importance des variables a montré que la différence entre l'élévation du pic du sommet à partir de GLAS et l'élévation du MNT était la variable la plus importante, suivie par la hauteur des arbres (H_c) et l'indice de rugosité (R).

7.5.2 Les effets de la saison sur les acquisitions GLAS

Le potentiel de discrimination des différents LTs en utilisant les formes d'onde GLAS ayants des empreintes GLAS coïncidentes, avec une forme d'onde pendant la saison humide et une forme d'onde pendant la saison sèche, a été analysé. Premièrement, les paires des formes d'onde ont été séparées en deux catégories. La première appelée leaf-on et correspond à des formes d'onde sans modifications des formes d'onde entre les deux saisons (arbres avec des feuilles dans les deux saisons). La seconde catégorie appelée leaf-off représente les arbres qui perdent leurs feuilles pendant la saison sèche. Les arbres leaf-off ont été identifiées à partir des paires de formes d'onde GLAS qui ont montrés des changements dans l'amplitude des pics du sommet de la canopée et du sol d'une saison à l'autre.

Les résultats montrent que la forêt ouverte (LT12), qui est connue pour être décidue, a présenté un rapport des énergies des sommets de canopées dans les saisons sèches et humides (Top_{wet} / Top_{dry}) de 2,4 vs ~1,1 pour LT8, LT09 et LT10, et 1,2 pour la leaf-on des LT12 (cf. Tableau 5.4). En outre, le rapport des énergies du pic du sol des formes d'ondes dans les saisons sèches et humides ($Ground_{wet}$ / $Ground_{dry}$) est similaire pour LT8, LT10 et les leaf-on des LT12 d'une valeur d'environ 1,1. Un rapport inférieur a été obtenu pour la leaf-off de LT12 (0,7) (cf. Tableau 5.4). La différence entre les élévations du centroïde de la forme d'onde GLAS et les élévations SRTM est similaire pour les données dans les deux saisons sèches et humides avec des valeurs comprises entre -4,3 et -6.3 m pour la saison humide et entre -4,7 et -5,8 m pour la saison sèche. Inversement, cette

différence augmente pour les paires leaf-off de LT12 de -5.8 m dans la saison humide jusqu'à -9.6 m pendant la saison sèche (cf. Tableau 5.4). La Figure 5.6 montre les emplacements des paires leaf-off et leaf-on en Guyane française, révélant que la majorité des paires leaf-off sont situés dans la même zone dans le sud de la Guyane française.

7.6 Conclusions et perspectives

7.6.1 Conclusions

Dans cette étude, la hauteur des arbres à partir des données GLAS en Guyane française a été estimée à l'aide de multiples modèles de régression linéaire et de la technique 'Random Forest' (RF). Cette analyse est basée soit sur les métriques extraites des formes d'ondes GLAS et des informations sur le terrain issu du MNT SRTM ou sur l'analyse en composantes principales (PCA) des formes d'onde GLAS. Les régressions linéaires et les régressions RF ont fournies des estimations de la hauteur des arbres avec une précision similaire en utilisant soit les métriques GLAS ou les composantes principales (EQM ~ 3.6 m). Toutefois, un modèle de régression (régression linéaire ou RF) basé sur l'ACP et les échantillons de la forme d'onde avec une information sur la Wext est une alternative intéressante pour estimer la hauteur des arbres car il ne nécessite pas d'autres métriques difficiles à obtenir à partir des formes d'onde GLAS dans les forêts denses, telles qu'en Guyane française.

Sachant que les données acquises à partir de GLAS ont une densité d'acquisition faible, mais une couverture géographique mondiale, il est donc utile de proposer une méthode pour cartographier la hauteur des arbres avec une bonne précision et à une résolution spatiale élevée. Dans cette étude, la hauteur des arbres extraites des deux capteurs LiDAR aéroporté et satellitaire, a été estimée à partir des données environnementales disponibles (par exemple, la géologie, de la pente, des indices de végétation, etc.) et utilisant la technique de régression-krigeage (krigeage des résidus de la régression du RF). Les cartes de la hauteur des arbres estimées en utilisant la régression-krigeage ont montrées une EQM sur l'estimation de la hauteur des arbres de 4,2 m en utilisant les données de calibration GLAS et de 1,8 en utilisant les donnes LiDAR aéroporté. Enfin, l'impact de l'échantillonnage spatial sur la précision des estimations de la hauteur des arbres a été

étudié. Les résultats indiquent qu'en utilisant l'approche de régression-krigeage, la précision sur la carte des hauteurs d'arbres était de 1,8 m avec un espacement des lignes de vol de 5 km et a évoluée pour atteindre un EQM de 4,8 m avec la configuration d'un espacement des lignes de vol de 50 km.

Finalement, dans cette étude, les formes d'ondes acquises par GLAS ont été utilisées pour analyser la différence entre les élévations de GLAS et du SRTM, dans le but de discriminer les cinq types de paysage forestier (LTs) en Guyane française. Les résultats ont montrés que la discrimination des cinq LTs en Guyane française est désormais possible avec un taux de classification de 81,3% et un coefficient kappa de 0,75. Finalement, les différences entre les formes d'onde GLAS coïncidentes (une en saison humide et l'autre en saison sèche) ont montrées que les arbres de LT12 pouvaient être facilement distingués des autres LTs qui conservent leurs feuilles en utilisant seulement trois critères: (1) la différence entre l'élévation du centroïde de la forme d'onde GLAS et l'élévation du SRTM, (2) Le ratio de l'énergie du sommet de la forme d'onde de la saison humides à l'énergie du sommet de la forme d'onde de la saison sèche, et (3) le rapport de l'énergie du sol de la forme d'onde de la saison humide à l'énergie du sol de la forme d'onde de la saison sèche.

7.6.2 Perspectives

Cette recherche ouvre un certain nombre de perspectives. Elles ouvrent de nombreux sujets sur l'amélioration des modèles d'estimations de la hauteur des canopées afin d'améliorer l'estimation de la biomasse.

7.6.2.1 La spatialisation de la hauteur des arbres à partir du LiDAR

Dans la section 1.4, la spatialisation de la hauteur de la canopée a été réalisée avec la régression-krigeage et l'utilisation des ensembles de données LiDAR aéroportés et satellitaire. Tandis que les résultats de l'estimation de la hauteur de la canopée ont été satisfaisants, plusieurs aspects doivent être poursuivis. Pour la méthode de cartographie de la hauteur de la canopée non-spatiale, nous avons utilisé la régression RF. Cependant, plusieurs autres modèles de régression tels que les modèles de régression inversés ([125]) ou, les modèles de régression à axe majeur réduit (AMR) ([126]; [125]) pourraient être utilisés à la place de RF, et leurs performances analysées. En ce qui concerne les méthodes de spatialisation, des modèles de krigeage anisotrope ayant une composante directionnel

doivent être testés. Enfin, les cartes créées dans cette étude pour l'estimation de la hauteur de la canopée ont une résolution de 250x250 m. Par conséquent d'autres cartes avec différentes résolutions pourraient être réalisées et analysées.

7.6.2.2 L'estimation de la biomasse

L'objectif principal de cette thèse était de spatialiser la hauteur de la canopée sur toute la Guyane française. Cet objectif a été atteint avec des résultats intéressants. Cependant, la hauteur de la canopée peut être utilisée pour estimer d'autres ressources forestières importantes, telles que les stocks de carbone (ACD) ou la biomasse (AGB), soit directement ou indirectement. Dans ce cadre, il serait intéressant d'étudier la relation entre la hauteur de la canopée à partir du LiDAR à différents niveaux d'énergie (RH100, RH75, et RH50) et leur corrélation à la biomasse dans notre zone d'étude. Des paramètres sur la topographie de la surface, le pourcentage du couvert forestier, les types de forêts, et l'âge du peuplement, pourraient également être ajoutés au modèle d'estimation de la biomasse dans une tentative d'améliorer la précision de l'estimation. Idéalement, une approche similaire à celle utilisée dans section 1.4 devrait être abordée. Finalement, l'approche présentée par Asner *et al.* [40]; pour l'estimation de l'ACD ou l'AGB en utilisant les estimations de la hauteur de la canopée à partir du LiDAR doit être analysée. Asner *et al.* [40]; ont proposé une allométrie des parcelles agrégées pour l'estimation de l'ACD ou l'AGB en utilisant des estimations de hauteur de la canopée par LiDAR. Dans leur étude, ils ont émis l'hypothèse que si la structure de la forêt et de l'organisation de la biomasse suivent des modèles de graduation constants, les variables au niveau de la parcelle pourraient capturer les mêmes informations sur AGB par rapport aux inventaires de terrain. Ensuite, puisque les paramètres tels que DBH et WD ne sont pas directement estimés par LiDAR, des relations ont été trouvées entre les hauteurs de la canopée par LiDAR et ces paramètres. Par conséquent, l'AGB peut maintenant être estimée à l'aide d'une allométrie des parcelles agrégées et en utilisant seulement les estimations de la hauteur de la canopée par LiDAR.

REFERENCES

[1] P. Falkowski, R. J. Scholes, E. E. A. Boyle, J. Canadell, D. Canfield, J. Elser and S. Steffen, "The global carbon cycle: a test of our knowledge of earth as a system," *Science,* vol. 290, p. he global carbon cycle: a test of our knowledge of earth as a system, 2000.

[2] T. J. Crowley, "Causes of climate change over the past 1000 years," *Science,* vol. 289, pp. 270-277, 2000.

[3] C. L. Sabine, R. A. Feely, N. Gruber and *et al.*, "The oceanic sink for anthropogenic CO2," *Science,* vol. 305, pp. 367-371, 2004.

[4] I. C. Prentice, G. D. Farquhar, F. Fasham, M. L. Goulden, M. Heimann, V. J. Jaramillo and *et al.*, "The carbon cycle and atmospheric carbon dioxide," *Cambridge University Press,* pp. 185-237, 2001.

[5] R. A. Houghton, K. T. Lawrence, J. L. Hackler and S. Brown, "The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates," *Global Change Biology,* vol. 7, pp. 731-746, 2001.

[6] Y. Pan, R. A. Birdsey, R. Houghton, P. E. Kauppi, W. A. Kurz and D. Hayes, "A large and persistent carbon sink in the world's forests.," *Science,* vol. 333, pp. 988-993, 2011.

[7] J. Martin, "Glacial-interglacial CO 2 change: The Iron Hypothesis," *Paleoceanography,* vol. 5, pp. 1-13, 1990.

[8] K. Kumar, M. Hoerling and B. Rajagopalan, "Advancing dynamical prediction of Indian monsoon rainfall," *Geophys. Res. Lett.,* vol. 32, pp. 1-4, 2005.

[9] Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.), "IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," 2014.

[10] D. L. Hartmann, A. M. G. Klein Tank, M. Rusicucci, L. V. Alexander, B. Broenniman, Y. Charabi and *et al.*, "Observations: atmosphere and surface," *Cambridge University Press,* pp. 159-254, 2013.

[11] FAO. 2010, "Global Forest Resources Assessment 2010 Tech. rept.".

[12] C. Beer, R. Markus, T. Enrico, C. Philippe, J. Martin, C. Nuno and R. Christian, "Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate," *Science,* vol. 329, pp. 834-838, 2010.

[13] FAO. 2000, "Global Forest Resources Assessment 2000. Tech. rept.".

[14] T. Feldpausch, M. Rondon, E. Fernandes and S. Riha, "Carbon and nutrient accumulation in secondary forests regenerating on patures in central Amazonia," *Ecological Appllications,* vol. 14, pp. 164-176, 2004.

[15] R. Hughes, . J. Kauffman and . V. Jaramillo-Luque, "Ecosystem-scale impacts of deforestation and land use in a humid tropical region of Mexico," *Ecological Applications,* vol. 10, pp. 515-527, 2000.

[16] G. Matthews, "The carbon content of trees.," *UK Forestry Commission, Edinburgh, UK,* 1993.

[17] S. Lamlom and R. Savidge, "A reassessment of carbon content in wood: variation within and between 41 North American species," *Biomass and Bioenergy,* vol. 25, pp. 381-388, 2003.

[18] J. Chambers, J. Schimel and A. Nobre, " Respiration from coarse wood litter in Central Amazon Forests," *Biogeochemistry,* vol. 52, pp. 115-131, 2001.

[19] Y. Malhi and J. Grace, "Tropical forests and atmospheric carbon," *Trends in Ecology &amp ; Evolution,* vol. 15, pp. 332-337, 2000.

[20] R. A. Houghton, "Aboveground Forest Biomass and the Global Carbon," *Global Change Biology,* vol. 11, pp. 945-958, 2005.

[21] G. M. Foody, G. Palubinskas, R. M. Lucas, P. J. Curran and M. Honzak, "Identifying terrestrial carbon sinks : Classication of successional stages in regenerating tropical forest from Landsat TM data.," *Remote Sensing of Environment,* vol. 55, pp. 205-216, 1996.

[22] O. L. Phillips, Y. Malhi, N. Higuchi, W. F. Laurance, P. V. Núñez, R. M. Vásquez and *et al.*, "Changes in the Carbon Balance of Tropical Forests : Evidence from Long-Term Plots.," *Science,* vol. 282, pp. 439-442, 1998.

[23] B. Hérault, J. Beauchêne, F. Muller, C. Baraloto, L. Blanc and J. Martin, "Modeling decay rates of dead wood in a neotropical forest," *Oecologia,* vol. 164, pp. 243-251, 2010.

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

[25] G. Sandberg, L. M. Ulander, J. E. S. Fransson, J. Holmgren and T. Le Toan, "L-and P-band backscatter intensity for biomass retrieval in hemiboreal forest," *Remote Sensing of Environment,* vol. 115, pp. 2874-2886, 2011.

[26] T. Le Toan, S. Quegan, M. W. J. Davidson, H. Balzter, P. Paillou, K. Papathanassiou and S. Plummer, "The BIOMASS mission: Mapping global forest biomass to better

understand the terrestrial carbon cycle," *Remote sensing of environment,* vol. 115, pp. 2850-2860, 2011.

[27] P. Ploton, R. Pélissier, N. Barbier, C. Proisy, B. R. Ramesh and P. Couteron, "Canopy texture analysis for large-scale assessments of tropical forest stand structure and biomass," *Treetops at Risk,* pp. 237-245, 2013.

[28] D. Lu, Q. Chen, G. Wang, E. Moran, M. Batistella, M. Zhang, G. V. Laurin and D. Saah, "Aboveground forest biomass estimation with Landsat and lidar data and uncertainty analysis of the estimates," *International Journal of Forestry Research,* pp. 1-16, 2012.

[29] P. Couteron, R. Pelissier, E.-A. Nicolini and D. Paget, " Predicting tropical forest stand structure parameters from Fourier transform of very high-resolution remotely sensed canopy images," *Journal of Applied Ecology,* vol. 42, pp. 1121-1128, 2005.

[30] C. Proisy, P. Couteron and F. Fromard, "Predicting and mapping mangrove biomass from canopy grain analysis using Fourier-based textural ordination of IKONOS images," *Remote Sensing of Environment,* vol. 109, pp. 379-392, 2007.

[31] N. Barbier, P. Couteron, C. Proisy, Y. Malhi and J. Gastellu-Etchegorry, "The variation of apparent crown size and canopy heterogeneity across lowland Amazonian forests," *Global Ecology and Biogeography,* vol. 19, pp. 72-84, 2010.

[32] N. Baghdadi, G. Le Maire, K. Osé, Y. Nouvellon, Z. Mehrez, C. Lemos and R. Hakamada, "Evaluation of ALOS/PALSAR L-band data for the estimation of Eucalyptus plantations aboveground biomass in Brazil," *IEEE, JSTARS,* p. in press, 2014.

[33] S.-T. Wu, "Potential application of multipolarization SAR for pine-plantation biomass estimation," *IEEE Geoscience and Remote Sensing,* Vols. 403-409, p. 3, 1987.

[34] M. C. Dobson, F. T. Ulaby, T. LeToan, A. Beaudoin, E. S. Kasischke and N. Christensen, "Dependence of radar backscatter on coniferous forest biomass," *IEEE Geoscience and Remote Sensing,* vol. 30, pp. 412-415, 1992.

[35] M. L. Imhoff, "Radar backscatter and biomass saturation: ramifications for global biomass inventory," *IEEE Geoscience and Remote Sensing,* vol. 33, pp. 511-518, 1995.

[36] A. Luckman, J. Baker, T. M. Kuplich, F. Y. C. da Costa and A. C. Frery, "A study of the relationship between radar backscatter and regenerating tropical forest biomass for spaceborne SAR instruments," *Remote Sensing of Environment,* vol. 60, pp. 1-13, 1997.

[37] A. Luckman, J. Baker, M. Honzák and R. Lucas, "Tropical forest biomass density estimation using JERS-1 SAR: Seasonal variation, confidence limits, and

application to image mosaics," *Remote Sensing of Environment,* vol. 63, pp. 126-139, 1998.

[38] V. Nizalapur, C. Sekhar Jha and R. Madugundu, "Estimation of above ground biomass in Indian tropical forested area using multifrequency DLRESAR data," *International Journal of Geomatics and Geosciences,* vol. 1, pp. 167-178, 2010.

[39] J. B. Drake, K. Robert G., D. Ralph O., C. David B., C. Richard, B. J. Bryan and H. Michelle, "Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: Factors affecting the generality of relationships," *Global Ecology and Biogeography,* vol. 12, pp. 147-159, 2003.

[40] G. P. Asner and J. Mascaro, "Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric," *Remote Sensing of Environment,* vol. 140, pp. 614-624, 2014.

[41] T. R. Feldpaush, J. Lloyd, S. L. Lewis, R. J. W. Brienen, M. Gloor and *et al.*, "Tree height integrated into pantropical forest biomass estimates," *Biogeosciences,* vol. 9, pp. 3381-3403, 2012.

[42] J. Chave, C. Andalo, S. Brown, M. A. Cairns, J. Q. Chambers, D. Eamus and H. Fölster, "Tree allometry and improved estimation of carbon stocks and balance in tropical forests," *Oecologia,* vol. 145, pp. 87-99, 2005.

[43] G. Vieilledent, R. Vaudry, S. Andriamanohisoa, O. Rakotonarivo, H. Randrianasolo, H. Razafindrabe and *et al.*, "A universal approach to estimate biomass and carbon stock in tropical forests using generic allometric models," *Ecological Applications,* vol. 22, pp. 572-583, 2012.

[44] M. A. Lefsky, H. David J., K. Michael, C. Warren B., C. Claudia C., E.-S. Fernando Del Bom, H. Maria O. and d. O. Raimundo, "Estimates of forest canopy height and aboveground biomass using ICESat," *Geophysical Research Letters,* vol. 32, p. L22S02, 2005.

[45] J. Boudreau, N. Ross F., M. Hank A., B. André, G. Luc and K. Daniel S., "Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec," *Remote Sensing of Environment ,* vol. 112, pp. 3876-3890, 2008.

[46] S. S. Saatchi, H. Nancy L., B. Sandra, L. Michael, M. Edward TA, S. William and Z. Brian R., "Benchmark map of forest carbon stocks in tropical regions across three continents," *Proceedings of the National Academy of Sciences,* vol. 108, pp. 9899-9904, 2011.

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

[48] G. P. Asner, J. Mascaro, H. C. Muller-Landau, G. Vieilledent, R. Vaudry, M. Rasamoelina, J. S. Hall and M. van Breugel, "A universal airborne LiDAR approach for tropical forest carbon mapping," *Oecologia,* vol. 168, pp. 1147-1160, 2012.

[49] N. Brokaw and R. Lent, Maintaining Biodiversity in Forest Ecosystems., Cambridge, UK: Cambridge University Press, 1999.

[50] T. E. Martin, "Habitat and area effects on forest bird assemblages: Is nest predation an influence?," *Ecology,* vol. 69, pp. 74-84, 1988.

[51] B. A. Maurer and R. C. Whitmore, "Foraging of five bird species in two forests with different vegetation structure," *The Wilson Bulletin,* vol. 93, pp. 478-490, 1981.

[52] S. K. Robinson and R. T. Holmes, "Foraging behavior of forest birds: the relationships among search tactics, diet, and habitat structure," *Ecology,* vol. 63, pp. 1918-1931, 1982.

[53] K. O'Hara, P. Latham and N. Valappil, "Parameters for describing stand structure.," in *Recent Advances in Forest Mensuration and Growth and Yield Research*, J. Skovsgaard and H. Burkhart, Eds., Hørsholm, Denmark, Ministry of Environment and Energy, Danish Forest and Landscape Research Institute, 1995, pp. 134-145.

[54] J. Praks, M. Hallikainen, O. Antropov and D. Molina, "Boreal forest tree height estimation from interferometric TanDEM-X images," *IEEE, IGARSS,* pp. 1262-1265, 2012.

[55] T. Castel, A. Beaudoin and G. Trouche, "Analysis of SAR interferometry for tree height," *Agricultura,* pp. 15-23, 2002.

[56] A. Reigber and A. Moreira, "First Demonstration of Airborne SAR Tomography using Multibaseline L-band Data," *IEEE, GRS,* vol. 38, 2000.

[57] F. Garestier, P. Dubois-Fernandez and I. Champion, "Forest Height Inversion Using High-Resolution P-Band Pol-InSAR Data," *IEEE, TGRS,* vol. 46, pp. 3544-3559, 2008.

[58] M. Neumann, S. Hensley, M. Lavalle and R. Ahmed, "Forest Structure Characterization Using Jpl'S Uavsar Multi-Baseline," *in The 4th Asia-Pacific Conference on Synthetic Aperture Radar,* 2013.

[59] Y. Huang, L. Ferro-Famil and A. Reigber, "Under-Foliage Object Imaging Using SAR Tomography and Polarimetric Spectral Estimators," *IEEE, TGRS,* vol. 50, pp. 2213-2225, 2011.

[60] B. Mercer, Q. Zhang, M. Schwaebisch, M. Denbina and S. Cloude, "Forest height and ground topography at L-band from and experimental single-pass airborne PolInSAR system," *proceedings of the PolInSAR 2009 workshop,* pp. 26-30, 2009.

[61] M. A. Wulder and D. Seeman, "Forest inventory height update through the integration of LIDAR data with segmented Landsat imagery," *Can. J. Remote Sens.,* vol. 29, p. 536–543, 2003.

[62] C. Hilbert and C. Schmullius, "Influence of surface topography on ICESat/GLAS forest height estimation and waveform shape," *Remote Sensing ,* vol. 4, pp. 2210-2235, 2012.

[63] S. Lee, W. Ni-Meister, W. Yang and Q. Chen, "Physically based vertical vegetation structure retrieval from ICESat data: Validation using LVIS in White Mountain National Forest, New Hampshire, USA," *Remote Sensing of Environment,* vol. 115, pp. 2776-2785, 2011.

[64] Y. Pang, L. Michael, A. Hans-Erik, M. Mary Ellen and S. Kirk, "Validation of the ICEsat vegetation product using crown-area-weighted mean height derived using crown delineation with discrete return lidar data," *Canadian journal of remote sensing ,* vol. 34, pp. S471-S484, 2008.

[65] M. Simard, P. Naiara, F. J. B and B. Alessandro, "Mapping forest canopy height globally with spaceborne lidar," *Journal Of Geophysical Research,* vol. 116, p. G04021, 2011.

[66] A. Hudak, M. Lefsky, W. B. Cohen and M. Berterretche, "Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height," *Remote Sens. Environ.,* vol. 82, pp. 397-416, 2002.

[67] D. Zheng, J. Rademacher, J. Chen, T. Crow, M. Bresee, J. Le Moine and S. R. Ryu, "Estimating aboveground biomass using Landsat 7 ETM± data across a managed landscape in northern Wisconsin, USA," *Remote Sensing of Environment,* vol. 93, pp. 402-411, 2004.

[68] W. Ni-Meister, S. Lee, A. H. Strahler, C. E. Woodcock, C. Schaaf, T. Yao, K. J. Ranson, G. Sun and J. B. Blair, "Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing," *J. Geophys. Res.,* vol. 115, p. G00E11, 2010.

[69] P. Addo-Fordjour and Z. Rahmad, "Mixed Species Allometric Models for Estimating above-Ground Liana Biomass in Tropical Primary and Secondary Forests, Ghana," *ISRN Forestry,* 2013.

[70] C. C. Carabajal and D. J. Harding, "SRTM C-band and ICESat laser altimetry elevation comparisons as a function of tree cover and relief," *Photogrammetric Engineering and Remote Sensing,* vol. 72, pp. 287-298, 2006.

[71] Q. Chen, "Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry," *Remote Sensing of Environment,* vol. 114, pp. 1610-1627, 2010.

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

[73] B. Bourgine and N. Baghdadi, "Assessment of C-band SRTM DEM in a dense equatorial forest zone," *Comptes Rendus Geoscience,* vol. 337, pp. 1225-1234, 2005.

[74] B. Bourgine, N. Baghdadi, S. Hosford and P. Daniels, "Generation of a ground-level DEM in a dense equatorial forest zone by merging airborne laser data and a top-of-canopy DEM," *Canadian journal of remote sensing,* vol. 30, pp. 913-926, 2004.

[75] G. Vincent, F. Caron, D. Sabatier and L. Blanc, "LiDAR shows that higher forests have more slender trees," *Bois Forêts Tropiques,* vol. 314, pp. 51-56, 2012.

[76] C. Delor, D. Lahondère, E. Egal and P. Marteau, "Carte géologique de la France à 1/500 000. Département de la Guyane. 2e édition," 2001.

[77] V. Gond, V. Freycon, J.-F. Molino, O. Brunaux, F. Ingrassia, P. Joubert, J.-F. Pekel and M.-F. Prévost, "Broad-scale spatial pattern of forest landscape types in the Guiana Shield," *International Journal of Applied Earth Observation and Geoinformation,* vol. 13, pp. 357-367, 2011.

[78] M. A. Lefsky, M. Keller, Y. Pang, P. B. De Camargo and M. O. Hunter, "Revised method for forest canopy height estimation from Geoscience Laser Altimeter System waveforms," *Journal of Applied Remote Sensing,* vol. 1, pp. 013537-013537, 2007.

[79] N. Baghdadi, S. Cavelier, J.-P. Chiles, B. Bourgine, T. Toutin, C. King, P. Daniels, J. Perrin and C. Truffert, "Merging of airborne elevation data and Radarsat data to develop a Digital Elevation Model," *International Journal of Remote Sensing,* vol. 26, pp. 141-166, 2005.

[80] G. Vincent, D. Sabatier, L. Blanc, J. Chave, E. Weissenbacher, R. Pélissier, E. Fonty, J.-F. Molino and P. Couteron, "Accuracy of small footprint airborne LiDAR in its predictions of tropical moist foreststand structure," *Remote Sensing of Environment,* vol. 125, pp. 23-33, 2012.

[81] J. A. B. Rosette, P. R. J. North and J. C. Suarez, "Vegetation height estimates for a mixed temperate forest using satellite laser altimetry," *International Journal of Remote Sensing,* vol. 29, pp. 1475-1493, 2008.

[82] G. Sun, K. J. Ranson, D. S. Kimes, J. B. Blair and K. Kovacs, "Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data," *Remote Sensing of Environment,* vol. 112, pp. 107-117, 2008.

[83] H. Duong, R. Lindenbergh, N. Pfeifer and G. Vosselman, "ICESat full-waveform altimetry compared to airborne LASER scanning altimetry over the Netherlands," *IEEE Geoscience and Remote Sensing,* vol. 47, pp. 3365-3378, 2009.

[84] T. Allouis, J.-S. Bailly, Y. Pastol and C. Le Roux, "Comparison of LiDAR waveform processing methods for very shallow water bathymetry using Raman, near-infrared and green signals," *Earth Surface Processes and Landforms,* vol. 35, pp. 640-650, 2010.

[85] D. Karlis, G. Saporta and A. Spinakis, "A simple rule for the selection of principal components," *Communications in Statistics-Theory and Methods,* vol. 32, pp. 643-666, 2003.

[86] A. L. Neuenschwander, L. A. Magruder and M. Tyler, "Landcover classification of small-footprint, full-waveform lidar data," *Journal of applied remote sensing,* vol. 3, pp. 033544-033544, 2009.

[87] L. I. Duncanson, K. O. Niemann and M. A. Wulder, "Estimating forest canopy height and terrain relief from GLAS waveform metrics," *Remote Sensing of Environment,* vol. 114, pp. 138-154, 2010.

[88] H. Akaike, "Information theory and an extension of the maximum likelihood principle," *Breakthroughs in statistics,* pp. 610-624, 1992.

[89] R. A. Houghton, D. Butman, A. G. Bunn, O. N. Krankina, P. Schlesinger and T. A. Stone, "Mapping Russian forest biomass with data from satellites and forest inventories," *Environmental Research Letters,* vol. 2, p. 045032, 2007.

[90] V. Kankare, M. Vastaranta, M. Holopainen, M. Räty, X. Yu, J. Hyyppä, H. Hyyppä, P. Alho and R. Viitala, "Retrieval of forest aboveground biomass and stem volume with airborne scanning LiDAR," *Remote Sensing,* vol. 5, pp. 2257-2274, 2013.

[91] G. le Maire, C. Marsden, Y. Nouvellon, C. Grinand, R. Hakamada, J.-L. Stape and J.-P. Laclau, "MODIS NDVI time-series allow the monitoring of Eucalyptus plantation biomass," *Remote Sensing of Environment,* vol. 115, pp. 2613-2625, 2011.

[92] O. Mutanga, E. Adam and M. Azong Cho, "High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm," *International Journal of Applied Earth Observation and Geoinformation,* vol. 18, pp. 399-406, 2012.

[93] L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.

[94] R. A. Houghton, F. Hall and S. J. Goetz, "Importance of biomass in the global carbon cycle," *Journal of Geophysical Research,* vol. 114, p. G00E0, 2009.

[95] F. G. Hall, K. Bergen, J. B. Blair, R. Dubayah, R. Houghton, G. Hurtt, J. Kellndorfer, M. Lefsky, J. Ranson, S. Saatchi, H. Shugart and D. Wickland,

"Characterizing 3D vegetation structure from space: Mission requirements," *Remote Sensing of Environment,* vol. 115, p. 2753–2775, 2011.

[96] A. Baccini, N. Laporte, S. Goetz, M. Sun and H. Dong, "A first map of tropical Africa's above-ground biomass derived from satellite imagery," *Environmental Research Letters,* vol. 3, 2008.

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

[98] T. Hengl, G. B. Heuvelink and A. Stein, "A generic framework for spatial prediction of soil variables based on regression-kriging," *Geoderma,* vol. 120, p. 75–93, 2004.

[99] W. Sun, B. Minasny and A. McBratney, "Analysis and prediction of soil properties using local regression-kriging," *Geoderma,* vol. 171, pp. 16-23, 2012.

[100] I. O. A. Odeha, A. B. McBratney and D. J. Chittleborough, "Spatial prediction of soil properties from landform attributes derived from a digital elevation model," *Geoderma,* vol. 63, pp. 197-214, 63.

[101] I. O. Odeh, A. B. McBratney and D. J. Chittleborough, "Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging," *Geoderma,* vol. 67, pp. 215-226, 1995.

[102] T. Ota, O. Ahmed, S. Franklin, M. Wulder, T. Kajisa and *et al.*, "Estimation of Airborne Lidar-Derived Tropical Forest Canopy," *Remote Sensing,* vol. 6, pp. 10750-10772, 2014.

[103] R. Genuer, J.-M. Poggi and C. Tuleau-Malot, "Variable selection using Random Forests," *Pattern Recognition Letters,* vol. 31, pp. 2225-2236, 2010.

[104] P. Goovaerts, Geostatistics for Natural Resources Evaluation, New York: Oxford University Press, 1997.

[105] I. Fayad, N. Baghdadi, J.-S. Bailly, N. Barbier, V. Gond, M. El Hajj, F. Fabre and B. Bourgine, "Canopy Height Estimation in French Guiana with LiDAR ICESat/GLAS Data Using Principal Component Analysis and Random Forest Regressions," *Remote Sens.,* vol. 6, 2014.

[106] D. J. Harding and C. C. Carabajal, "ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure," *Geophysical research letters,* vol. 32, p. L21S10, 2005.

[107] P. Mayaux, E. Bartholome, S. Fritz and A. Belward, "New land-cover map of Africa for the year 2000," *Journal of Biogeography,* vol. 31, p. 861–877, 2004.

[108] R. Dubayah and J. Drake, "LIDAR remote sensing for forestry," *Journal of Forestry,* vol. 98, pp. 44-46, 2000.

[109] A. T. Hudak, v. Crookston, J. S. Evans, D. E. Hall and M. J. Falkowski, "Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data," *Remote Sensing of Environment,* vol. 112, p. 2232–2245, 2008.

[110] V. Duong, R. Lindenbergh, N. Pfeifer and G. Vosselman, "Single and two epoch analysis of ICESat full waveform data over forested areas," *International Journal of Remote Sensing,* vol. 29, pp. 1453-1473, 2008.

[111] S. Ali, P. Dare and S. Jones, "Fusion of remotely sensed multispectralimagery and Lidar data for forest structure assessment at the tree level," *ISPRS Proceedings,* vol. XXXVII, 2008.

[112] J. T. Mundt, D. R. Streutker and N. F. Glenn, "Mapping sagebrush distribution using fusion of hyperspectral and lidar classifications," *Photogrammetric Engineering & Remote Sensing,* vol. 72, pp. 47-54, 2006.

[113] M. Dalponte, L. Bruzzone and D. Gianelle, "Fusion of Hyperspectral and LiDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geoscience and Remote Sensing,* vol. 46, pp. 1416-1427, 2008.

[114] v. Bhang, F. Schwartz and A. Braun, "Verification of the vertical error in C-band SRTM DEM using ICESat and Landsat-7, Otter Tail County, MN," *IEEE-GRSE,* vol. 45, pp. 36-44, 2007.

[115] E. Rodríguez, C. Morris and J. Belz, "A Global Assessment of the SRTM Performance," *Photogrammetric Engineering & Remote Sensing,* vol. 72, pp. 249-260, 2006.

[116] M. Hofton, R. Dubayah, J. Blair and D. Rabine, "Validation of SRTM elevations over vegetated and non-vegetated terrain using medium footprint LiDAR," *Photogrammetric Engineering & Remote Sensing,* vol. 72, pp. 279-285, 2006.

[117] K. Becek, "Investigation of elevation bias of the SRTM C- and X-Band digital elevation models," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* pp. 105-110, 2008.

[118] A. Liaw and M. Wiener, "Classification and Regression by randomForest," *R news,* vol. 2, pp. 18-22, 2002.

[119] V. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo and J. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 67, pp. 93-104, 2012.

[120] M. Immitzer, C. Atzberger and T. Koukal, "Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data," *Remote Sensing,* vol. 4, pp. 2661-2693, 2012.

[121] M. Hofton and J. B. Blair, "Laser altimeter return pulse correlation: A method for detecting surface topographic change," *Journal of Geodynamics,* vol. 34, pp. 477-489, 2002.

[122] Y. Huang and S. Du, "Weighted support vector machine for classification with uneven training class sizes," *Proceedings of 2005 International Conference on Machine Learning and Cybernetics,* vol. 7, pp. 4365-4369, 2005.

[123] S. Kotsiantis and P. Pintelas, "Mixture of Expert Agents for Handling Imbalanced Data Sets," *Annals of Mathematics Computing & TeleInformatics,* vol. 1, pp. 45-55, 2003.

[124] X. Liu, J. Wu and Z. Zhou, "Exploratory Undersampling for Class-Imbalance Learning," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 2, pp. 539-550, 2009.

[125] P. Curran, Hay and A.M., "The importance of measurement error for certain procedure in remote sensing of optical wavelengths," *Photogrammetric Engineering and Remote Sensing,* vol. 52, pp. 229-241, 1986.

[126] W. B. Cohen, T. K. Maiersperger, S. T. Gower and D. P. Turner, "An improved strategy for regression of biophysical variables and Landsat ETM+ data," *Remote Sensing of Environment,* vol. 84, pp. 561-571, 2003.

[127] A. Journel and M. Rossi, "When do we need a trend model in kriging?," *Mathematical Geology,* vol. 21, pp. 715-739, 1989.

[128] A. Stein and L. Corsten, "Universal kriging and cokriging as a regression procedure," *Biometrics,* vol. 47, pp. 575-587, 1991.

[129] M. Berterretche, "Comparison of regression and geostatistical methods to develop LAI surfaces for NPP modeling," *Master's thesis (169 pp.). Corvallis, OR: Oregon State University.,* 2001.

[130] Q. Ketterings, R. Coe, M. van Noordwijk and C. Palm, "Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests," *Forest Ecology and management,* vol. 146, pp. 199-209, 2001.

[131] R. Pilli, T. Anfodillo and M. Carrer, "Towards a functional and simplified allometry for estimating forest biomass," *Forest Ecology and Management,* vol. 237, pp. 583-593, 2006.

[132] D. Zianis and M. Mencuccini, "On simplifying allometric analyses of forest biomass," *Forest Ecology and Management,* vol. 187, pp. 377-332, 2004.

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

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

[135] M. Hardisky, F. Daiber, C. Roman and V. Klemas, "Remote sensing of biomass and annual net aerial primary productivity of a salt marsh," *Remote Sensing of Environment,* vol. 16, pp. 91-106, 1984.

[136] G. Anderson and J. Hanson, "Evaluating hand-held radiometer derived vegetation indices for estimating above ground biomass," *Geocarto International,* vol. 7, pp. 71-78, 1992.

[137] S. R. Freitas, M. C. Mello and C. B. Cruz, " Relationships between forest structure and vegetationindices in Atlantic Rainforest," *Forest Ecology and Management ,* vol. 218, pp. 353-362, 2005.

[138] C. Pascual, A. Garcia-Abril, W. Cohen and S. Martin-Fernandez, "Relationship between LiDAR-derived forest canopy height and Landsat images," *nternational Journal of Remote Sensing,* vol. 31, pp. 1261-1280, 2010.

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

[140] S. Vieira, P. B. de Camarago, S. Diogo, R. da Silva, L. Hutyra, J. Q. Chambers, I. Foster Brown, N. Higuchi, J. dos Santos, S. C. Wofsy, S. E. Trumbore and L. A. Martinelli, "Forest structure and carbon dynamics in Amazonian tropical rain forests," *Oecologia,* vol. 140, pp. 468-479, 2004.

[141] M. Santoro, J. Askne and P. B. G. Dammert, "Tree height estimation from multi-temporal ers sar interferometric phase omography," *Proc. of FRINGE 2003 Workshop, Frascati, Italy,* 2003.

[142] C. Proisy, E. Mougin, F. Fromard and M. A. Karam, "Interpretation of polarimetric radar signatures of mangrove forests," *Remote Sensing of Environment,* vol. 71, pp. 56-66, 2000.

[143] C. Proisy, N. Barbier, M. Gueroult, R. Pelissier, J.-P. Gastellu-Etchegorry, E. Garu and P. Couteron, "Biomass prediction in tropical forests: the canopy grain approach," *Fatoyinbo, T. E. (Ed.) / Book 1. Intech Open Access Publisher,* pp. 59-76, 2012.

[144] Y. Pan, R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz and O. L. Phillips, "A large and persistent carbon sink in the world's forests," *Science,* vol. 333, pp. 988-993, 2011.

[145] T. Maia Araújo, N. Higuchi and J. A. de Carvalho Júnior, "Comparison of formulae for biomass content determination in a tropical rain forest site in the state of Pará, Brazil," *Forest Ecol. Manag.,* vol. 117, pp. 43-52, 1999.

[146] A. J. N. Lima, R. Suwa, G. H. P. de Mello Ribeiro, T. Kajimoto, J. dos Santos, R. P. da Silva, C. A. S. de Souza and e. al., "Allometric models for estimating above- and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil," *Forest Ecol. Manag.,* vol. 2012, pp. 163-172, 2012.

[147] C. O. Justice, E. Vermote, J. R. Townshend, R. Defries, D. P. Roy, D. K. Hall, V. V. Salomonson and e. al., "The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research," *IEEE Transactions on Geoscience and Remote Sensing ,* vol. 36, pp. 1228-1249, 1998.

[148] S. Guillaso and A. Reigber, "Scatterer Characterisation Using Polarimetric SAR Tomography," *Proc. of IEEE 2005 Int. Geosci. and Remote Sens-ing Symp,* 2005.

[149] H. T. M. Dinh, T. Le Toan, F. Rocca, S. Tebaldini, d. M. Mariotti and L. Villard, "Relating P-band synthetic aperture radar tomography to tropical forest biomass," *IEEE Geoscience and Remote Sensing,* vol. 52, pp. 967 - 979 , 2014.

[150] X. Huang, H. Xie, T. Liang and D. Yi, "Estimating vertical error of SRTM and map-based DEMs using ICESat altimetry data in the eastern Tibetan Plateau.," *International Journal of Remote Sensing,* vol. 32, pp. 5177-5196, 2011.

[151] M. Immitzer, C. Atzberger and T. Koukal, "Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data," *Remote Sensing,* vol. 4, pp. 2661-2693, 2012.

[152] V. Kankare, M. Vastaranta, M. Holopainen, M. Räty, X. Yu, J. Hyyppä, H. Hyyppä, P. Alho and R. Viitala, "Retrieval of Forest Aboveground Biomass and Stem Volume with Airborne Scanning LiDAR," *Remote Sensing,* vol. 5, pp. 2257-2274, 2013.

[153] P. Pennec, V. Gond and D. Sabatier, "Characterization of tropical forests phenology in French Guiana using MODIS time-series," *Remote Sensing Letters,* vol. 2, p. 337–345, 2011.

[154] S. Sylvander, P. Henry, C. Bastien-Thiry, F. Meunier and D. Fuster, "VEGETATION geometrical image quality," *VEGETATION-2000 Symposium,* pp. 33-44, 2000.