



**HAL**  
open science

# Estimation de la biomasse forestière du Brésil à partir des données LiDAR

Vivien Scholkopf

► **To cite this version:**

Vivien Scholkopf. Estimation de la biomasse forestière du Brésil à partir des données LiDAR. Sciences de l'ingénieur [physics]. 2019. dumas-02959002

**HAL Id: dumas-02959002**

**<https://dumas.ccsd.cnrs.fr/dumas-02959002>**

Submitted on 6 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**CONSERVATOIRE NATIONAL DES ARTS ET METIERS**

**ECOLE SUPERIEURE DES GEOMETRES ET TOPOGRAPHES**

---

**MEMOIRE**

**présenté en vue d'obtenir**

**le DIPLOME D'INGENIEUR CNAM**

**SPECIALITE : Géomètre et Topographe**

**par**

**Vivien SCHOLKOPF**

---

**Estimation de la biomasse forestière du Brésil à partir des données LiDAR**

**Soutenu le Jeudi 5 septembre 2019**

---

**JURY**

Monsieur Laurent MOREL	Président du jury
Madame Prénom2 NOM2	Examineur
Monsieur Martin BELAND	Maître de stage
Monsieur Laurent POLIDORI	Enseignant référent

## Remerciements

Tout d'abord, j'adresse mes remerciements à M. Martin BELAND, professeur adjoint du département des sciences géomatiques de l'université Laval. En tant que directeur de mémoire, il m'a guidé dans la démarche du projet et m'a été d'une aide précieuse au cours de mon stage.

Je remercie également M. Laurent POLIDORI, en tant que professeur référent, qui m'a donné des conseils avisés dans mes travaux.

Je tiens à adresser mes remerciements au corps professoral et administratif de l'Université Laval pour la qualité de leur enseignement et leur aide administrative.

Je désire remercier également au corps professoral et à l'administration de l'ESGT, spécialement M. Stéphane DURAND et Mme. Hélène MASSOT qui m'ont permis d'effectuer un stage à l'étranger particulièrement enrichissant.

Je tiens à témoigner mes remerciements pour M. Jean Romain ROUSSEL qui m'a fait part de ses connaissances sur le travail qu'il a entrepris.

Pour finir, je voudrais exprimer mes reconnaissances envers mes camarades étudiants du département sciences géomatiques, spécialement M. Olivier MATTE qui m'a partagé ses connaissances. Je voudrais remercier également Mlle. Valérie COTTON et M. Tristan MAURIN pour leur soutien et leur lecture.

## Liste des abréviations

AGB : Above-Ground Biomass

CRAN: Comprehensive R Archive Network

FWF : Full WaveForm

INPE: Instituto Nacional de Pesquisas Espaciais

LiDAR: Light Detection And Ranging

MHC : Modèle de Hauteur de Canopée ou MHC : Canopy Height Model

MNS : Modèle Numérique de Surface ou DSM : Digital Surface Model

MNT : Modèle numérique de terrain ou DEM : Digital Elevation Model

REDD : Reduced Emissions from Deforestation and Forest Degradation

RMSE : Root-Mean-Square Error

TIN : Triangulated Irregular Network

## Glossaire

Angle de balayage : valeur en degrés entre  $-90^\circ$  et  $90^\circ$ . A  $0^\circ$ , l'impulsion laser est directement en dessous de l'avion (au nadir). A  $-90^\circ$ , l'impulsion laser est à la gauche de l'avion, à  $+90^\circ$ , l'impulsion laser est à la droite de l'appareil dans la direction du vol. La plupart des systèmes LiDAR captent le signal à l'intérieur de  $+ ou - 30^\circ$ .

Biais : erreur statistique

Biomasse aérienne : toute biomasse vivante au-dessus du sol, y compris les tiges, les souches, les branches, l'écorce, les graines et le feuillage.

Canopée : surface de la cime des arbres

Conifères : arbres dont le fruit est en forme de cônes.

CRAN : réseau de serveurs web dans le monde entier qui stocke des versions identiques et à jour du code et de la documentation pour le langage de programmation R.

Densité de points : nombre de points par unité de surface.

Feuillus : arbres qui ont beaucoup de feuilles.

Format las et laz : format binaire standard permettant de stocker les données de LiDAR. Laz est un format binaire standard pour stocker les données LiDAR mais est compressé.

Histogramme : diagramme servant à représenter par des bandes juxtaposées une distribution d'un caractère statistique quantitatif continu sur un échantillon donné.

Pixel : c'est le plus petit élément constitutif d'une image produite. Il représente la cellule d'une grille raster.

Raster : image matricielle : où chaque cellule de taille prédéterminée attribue des valeurs spécifiques. La taille de ces cellules s'appelle la résolution spatiale, ou l'échelle d'observation.

# Table des matières

Remerciements .....	2
Liste des abréviations .....	3
Glossaire .....	4
Introduction .....	7
I Etude de l'ampleur du biais de la hauteur de la canopée en fonction de la densité d'impulsions .....	9
<b>I.1 INTRODUCTION</b>	<b>9</b>
I.1.1 DEFINITION DE LA HAUTEUR DE CANOPEE	9
I.1.2 APPROCHES POUR DECRIRE LA STRUCTURE DE LA FORET	10
<b>I.2 CADRE THEORIQUE DU MODELE</b>	<b>12</b>
I.2.1 NOTATION	12
I.2.2 UNE CANOPEE PARFAITEMENT PLATE	13
I.2.3 UNE CANOPEE PLATE AVEC UNE SINGULARITE	14
I.2.4 UNE CANOPEE PLATE AVEC DEUX SINGULARITES	15
I.2.5 UNE CANOPEE DE FORME CONTINUE	16
I.2.6 QUANTIFICATIONS DU BIAIS A L'AIDE D'HISTOGRAMMES STANDARDISES	17
I.2.7 EFFET DE LA TAILLE DE L'EMPREINTE	18
<b>I.3 MATERIELS ET METHODES</b>	<b>19</b>
I.3.1 SITE D'ETUDE	19
I.3.2 DONNEES :	21
I.3.3 PRE-TRAITEMENT	23
I.3.4 TRAITEMENT DU MODELE DE ROUSSEL ET AL.	30
<b>I.4 RESULTATS</b>	<b>30</b>
<b>I.5 DISCUSSION</b>	<b>35</b>
II Effet de l'estimation de la hauteur sur la biomasse forestière du Brésil .....	36
<b>II.1 CADRE THEORIQUE :</b>	<b>36</b>
II.1.1 DEFINITION DE LA BIOMASSE	36
II.1.2 SYSTEMES DE TELEDETECTION POUR QUANTIFIER LA BIOMASSE	37
II.1.3 DIFFERENTES APPROCHES POUR MESURER LA BIOMASSE :	39
II.1.4 ÉQUATIONS ALLOMETRIQUES.	40
II.1.5 ÉQUATIONS ALLOMETRIQUES SUR LE TERRITOIRE DU BRESIL :	42
II.1.6 EXEMPLE DE DEUX EQUATIONS APPLIQUEES EN AMAZONIE	44
II.1.7 EXEMPLE DE DEUX EQUATIONS APPLIQUEES EN FORET ATLANTIQUE DU BRESIL	45
<b>II.2 MATERIELS ET METHODES :</b>	<b>45</b>
II.2.1 SITE D'ETUDE ET DONNEES	45
II.2.2 PRE-TRAITEMENT	45
II.2.3 TRAITEMENT	46
<b>II.3 RESULTATS</b>	<b>47</b>
II.3.1 ESTIMATION DE LA BIOMASSE AVEC DES DONNEES A FAIBLE DENSITE : 1, 3 POINTS / M <sup>2</sup>	47

I.1	II.3.2	ESTIMATION DE LA BIOMASSE AVEC DES DONNEES A MOYENNE DENSITE : 7 POINTS / M <sup>2</sup>	48
I.2	II.3.3	ESTIMATION DE LA BIOMASSE AVEC DES DONNEES A HAUTE DENSITE : 21 POINTS / M <sup>2</sup>	48
I.3	II.3.4	COMPARAISON DE L'ESTIMATION DE LA BIOMASSE SELON LES DIFFERENTES DONNEES ECHANTILLONNEES	49
II	<b>II.4</b>	<b>DISCUSSION</b>	<b>50</b>
		Conclusion : .....	52

## Introduction

Les forêts tropicales sont d'importants réservoirs de carbone et de biodiversité qu'il est important de quantifier. La télédétection par LiDAR aéroporté est un outil puissant pour déterminer la biomasse forestière et d'évaluer la dynamique du cycle de carbone au fil du temps. Des estimations précises de la biomasse aérienne (toute biomasse vivante au-dessus du sol) et de la variation des stocks de carbone provenant d'activités humaines sont nécessaires pour l'atténuation du changement climatique en cours. Le LiDAR permet donc la cartographie des stocks de carbone à haute résolution sur de vastes échelles spatiales et un large éventail d'écosystèmes avec un potentiel pour la surveillance des stocks de carbones et les efforts d'atténuation du changement climatique.

Une entente entre l'université Laval de Québec et l'Institut National de Recherche Spatiale du Brésil (INPE) a été mise en place en vue de la création d'un projet de cartographie de la biomasse forestière du Brésil. Le partenaire brésilien a souligné l'incertitude de l'estimation de la biomasse par LiDAR liée à l'estimation de la hauteur. Aussi, ce projet s'inscrit dans le cadre des efforts de REDD+ (Reduced Emissions from Deforestation and Forest Degradation), c'est une initiative internationale qui vise à lutter contre le réchauffement climatique provoqué par les émissions de dioxyde de carbone induites par la déforestation.

Ainsi ce TFE (Travail de Fin d'Etudes) s'inscrit dans le projet de l'estimation de la biomasse forestière brésilienne qui constitue un enjeu planétaire. Estimer la biomasse à l'aide de l'outil LiDAR implique un coût de production important et du temps, c'est pourquoi limiter le temps d'utilisation et acquérir les données sur une plus grande surface serait intéressant. Pour cela, il est possible d'acquérir des points à plus faible densité donc obtenir la possibilité d'une hauteur de vol plus grande et par conséquent couvrir une plus grande surface, tout en gardant une bonne précision.

Un modèle mathématique permet d'évaluer le biais de l'estimation de la hauteur de la canopée en fonction de la densité d'impulsions. Jean-Romain ROUSSEL ancien élève doctorant de Martin BELAND, est l'auteur de ce modèle. Il a publié un article en 2017 sur la suppression du biais des estimations de la hauteur de la canopée à partir du LiDAR aérien. (Roussel et al. (2017)). En 2015, il a commencé à développer un logiciel de programmation statistique gratuit et libre accès. Il s'avère que cet outil qu'il a nommé « lidR » a eu un fort succès dans l'application des inventaires forestiers par LiDAR et est rendu disponible sur le

CRAN<sup>1</sup> (Roussel et al. (2019)). « lidR » fournit de nombreuses fonctions permettant de traiter et visualiser les données LiDAR aériens pour les applications forestières. Jean-Romain ROUSSEL poursuit actuellement le développement de « lidR » en y intégrant de nouvelles fonctions.

L'objectif de ce mémoire est d'adapter son modèle aux jeux de données LiDAR de la forêt brésilienne, puis d'estimer la biomasse à l'aide d'équations allométriques.

Nous allons tout d'abord vérifier l'application du modèle développée par Roussel et al. (2017) qui permet de quantifier théoriquement le biais sur la hauteur de la canopée en fonction de la densité d'impulsions, sur un jeu de donnée provenant de la forêt tropicale du Brésil, comprenant de grands arbres avec une couverture forestière importante. Ensuite, le but est de montrer l'effet de l'estimation de la hauteur sur le calcul de la biomasse, en appliquant des équations allométriques avec plusieurs jeux de données de densités différentes à faible et haute densité. Cela permettrait d'en déduire s'il y a un écart important sur le calcul de biomasse entre les deux jeux de données. S'il y a un écart important dans les résultats, cela est probablement dû à un biais dans l'estimation de la hauteur, à l'inverse, si l'écart est faible, c'est que le biais n'est probablement pas important. Si nous trouvons une sous-estimation moyenne de biomasse inférieure à 50 Mg / ha c'est que le biais n'est pas très élevé, en revanche, si nous obtenons une sous-estimation de plus de 100 Mg / ha, c'est que le biais de l'estimation de la hauteur commence à être important.

En première partie, nous étudierons l'ampleur du biais de la hauteur de la canopée en fonction de la densité d'impulsions, et en deuxième partie, nous montrerons l'effet de l'estimation de la hauteur sur la biomasse forestière du Brésil.

---

<sup>1</sup> CRAN : Comprehensive R Archive Network

# I Etude de l'ampleur du biais de la hauteur de la canopée en fonction de la densité d'impulsions

## I.1 Introduction

### I.1.1 Définition de la hauteur de canopée

En biologie, la canopée est la partie aérienne d'un sommet de culture formé par les couronnes végétales. Pour les forêts, la canopée est la couche supérieure formée par la cime des arbres matures y compris d'autres organismes biologiques.

En botanique, la canopée est définie par la strate supérieure d'une forêt fermée abritant une grande partie des organismes de la forêt. Parfois, le terme couvert forestier est un synonyme du mot canopée pour indiquer l'étendue de la couche externe de feuilles d'arbres. Les arbres de la canopée interceptent la lumière du soleil et sont efficaces pour la photosynthèse permettant la réduction du dioxyde de carbone. Elle protège contre les vents forts et tempêtes tout en interceptant la lumière. La canopée de la forêt tropicale intercepte environ 95 % de la lumière du soleil et présente une couche de 10 m d'épaisseur. Elle est en dessous de la couche émergente (les arbres qui dépassent les autres en hauteur). Mais l'absorption de la lumière empêche la croissance des arbres du sous-étage à la canopée.



Figure 1: canopée en forêt tropicale d'Amérique du Sud, source : <https://www.aquaportail.com/synonyme-11461-canopee.html>

Dans notre cas, la canopée représente la surface de la cime des arbres. La hauteur de la canopée est donc la distance entre le sol et le sommet des arbres.

### ***1.1.2 Approches pour décrire la structure de la forêt***

On suppose facilement que la hauteur estimée par LiDAR aérien est hautement sensible à la densité de points quand on considère le problème sous forme probabiliste. Il existe donc une probabilité de ne pas toucher le sommet de l'arbre dans un espace prédéfini. Par conséquent, plus il y a de points dans l'espace donné, plus il y aura de probabilité de toucher ce sommet, mais à l'inverse, moins il y aura de points plus la probabilité est faible d'atteindre le haut et donc de sous-estimer la métrique hauteur maximale. Donc si la densité tend vers l'infini alors la probabilité d'atteindre le sommet tend vers 1. Le but du modèle est de quantifier l'estimation de la hauteur en se basant sur la théorie des probabilités, il permettrait d'obtenir de façon statistique la hauteur maximum d'une parcelle comme si le jeu de données était acquis avec une densité infinie.

De façon générale, l'application du LiDAR dans le secteur de la foresterie est soit une approche zonale soit une approche par arbres individuels.

L'approche qui reste la plus pratique pour décrire la structure de la forêt est l'approche zonale (Area-based approach). Le but est de lier une grandeur biophysique  $Q$  à des grandeurs  $X_i$  extraits du nuage de points appelé communément « métrique ». Cette relation se fait par modélisation statistique. Par exemple, voici une modélisation simple :

$$Q = aX$$

- $Q$  est une variable dépendante qui représente une grandeur non observable comme le volume marchand, la biomasse
- $X$  correspond à une variable explicative qui est une métrique dépendante du nuage de points, calculée à partir de la distribution verticale de points tels que la hauteur moyenne, maximale, écart-type, quantiles... Ils sont de nature statistique et tirent parti d'une seule dimension en général. La modélisation statistique permet de lier les métriques  $X_i$  à  $Q$  par des équations dont les paramètres sont ajustés automatiquement à des données d'inventaire de terrains utilisées comme référence. Les attributs d'intérêts forestiers tels que la hauteur de l'arbre et le diamètre à hauteur de poitrine sont les mesures relatives aux arbres mesurés au sein d'une placette de terrain.

Tout d'abord, pour utiliser cette approche, les données par LiDAR sont acquises pour toute la région étudiée et mesurées sur le terrain des placettes témoins. Ensuite, il faut prévoir la création de modèles prédictifs en ayant découpé le nuage de points en fonction de la surface

de chaque pixel. Les variables explicatives (hauteur moyenne, écart-type, percentiles de hauteur, densité du couvert...) sont ensuite calculées à partir du nuage de points, découpées et normalisées. Une équipe au sol mesure les attributs d'intérêts hauteur et diamètre. Ces calculs et mesures permettent ainsi de construire des modèles prédictifs. Puis, les modèles prédictifs sont appliqués sur toute la région étudiée à partir de l'ensemble des paramètres dans le but d'obtenir l'ensemble des estimations et des cartes d'attributs d'inventaire forestier.

L'avantage de l'approche par zone par rapport aux inventaires forestiers classiques est que l'échelle des peuplements comprend la connaissance spatiale complète, des prédictions précises des variables dépendantes et la capacité de calculer des intervalles de confiance pour les estimations. (White et al. 2013)

L'approche alternative est une approche par arbre individuel qui revient à délimiter et mesurer les couronnes d'arbres individuels.

La première étape consiste à la reconnaissance et segmentation individuelle de chaque arbre à partir du nuage de points et la deuxième étape est d'extraire pour chaque arbre les métriques  $X_i$  comme la hauteur de l'arbre ou le diamètre de la couronne. La dernière étape est de lier par modélisation statistique  $Q$  à  $X_i$  avec des équations allométriques.

Cette approche gagne rapidement en importance, mais est moins utilisée malgré la baisse des coûts d'acquisition des données et l'augmentation constante de la puissance de calcul. L'approche par zone est plus utilisée, car elle nécessite une densité de points inférieure et donc moins chère. Cependant, un des inconvénients de l'approche par zone est que les modèles statistiques ne peuvent pas être généralisés dans toutes les configurations. Par exemple, deux mesures  $X$  et  $Y$  sont associées à une quantité d'intérêt  $Q$  selon l'équation  $Q = aX^bY^c$ , ce modèle n'est pas spécifique au type de forêt échantillonnée, car  $a$ ,  $b$  et  $c$  sont estimés à partir d'un inventaire local, et  $X$  et  $Y$  pourraient être spécifiques à la configuration des paramètres LiDAR. Une solution pour pallier à ce problème consiste à développer des modèles à l'aide de métriques qui restent stables lorsque les paramètres d'acquisition changent. Mais ces considérations sont rarement présentées dans la littérature, bien que Nasset (2004) ait signalé que la hauteur des premiers retours ne variait pas de manière significative avec l'altitude de vol ou le diamètre de l'empreinte, les derniers retours étaient quant à eux plus sensibles aux variations du diamètre de l'empreinte. Ce qui a davantage été retenu dans la littérature pour résoudre ce problème est d'examiner l'effet des paramètres

d'acquisition sur la structure du nuage de points et donc sur les métriques et les prédictions du modèle. En particulier, l'effet de la densité d'impulsions sur les prédictions du modèle (Gobakken et Næsset 2008; Jakubowski, Guo, et Kelly 2013). Ces études ont la plupart conclu que la densité des impulsions n'avait que peu ou pas d'effet sur les prévisions parce que les métriques restent stables lorsque la densité d'impulsions est réduite artificiellement. En revanche, les métriques telles que la hauteur maximale et ses dérivées ne sont pas stables parce qu'elles ne sont pas des statistiques. Les modèles qui ont des métriques instables peuvent générer des prédictions comprenant des biais pour de faibles densités d'impulsions. Les études antérieures utilisent généralement une approche empirique pour vérifier si les paramètres d'acquisition ont un effet mesurable sur des métriques en particulier. Les efforts visant à corriger le biais ont été principalement limités à la normalisation de l'intensité du signal. (Höfle et Pfeifer 2007). L'approche de Roussel et al., consiste à recalculer les métriques LiDAR comme si elles avaient été obtenues à partir d'un « périphérique standard » (Roussel et al. 2017). Cette méthode de normalisation donne les mêmes mesures que celles obtenues avec une densité d'impulsions infinie, une taille d'empreinte nulle, et un angle de balayage constant au nadir.

## **I.2 Cadre théorique du modèle**

Dans cette partie théorique, nous allons nous concentrer sur la métrique hauteur maximale. Un modèle mathématique est créé pour comprendre comment varie le biais de cette métrique en fonction de la densité d'impulsions, de la structure de la forêt et de l'échelle d'observation à laquelle il est calculé. Nous allons expliquer comment ce modèle mathématique est construit, à l'aide de diagrammes pour illustrer la manière dont le biais de l'estimation de la hauteur maximale peut être évalué de façon probabiliste. La nature probabiliste du phénomène de sous-estimation devrait être plus claire en examinant les quatre formes de canopée ci-dessous.

### **I.2.1 Notation**

$h_{max}$  : hauteur maximale réelle pour une zone donnée

$\hat{h}_{max}$  : hauteur maximale observée pour une zone donnée

$\hat{h}$  : hauteur maximale attendue, la plus probable pour une zone donnée

$P(E)$  : Probabilité d'un évènement  $E$ .  $X$  : variable aléatoire

### 1.2.2 Une canopée parfaitement plate

Dans le cas où la canopée est plane, la hauteur maximale observée ne peut prendre qu'une valeur, les Figure 2 (a) et (b) l'expliquent simplement : une impulsion ne peut atteindre la canopée qu'avec une hauteur  $h_0$  suivant une probabilité  $p_0$ . Donc la hauteur maximale observée sera toujours la hauteur maximale réelle quel que soit le nombre d'impulsions  $n_0$ . Ainsi,  $h_{max} = \hat{h}_{max} = h_0$ . La valeur attendue de  $\hat{h}_{max}$  exprimée en fonction de  $n$  impulsion s'écrit alors :

$$\bar{h}(n) = p_0^n \times h_0 = h_0 \quad (1)$$

Puisque  $p_0^n = 1$ .

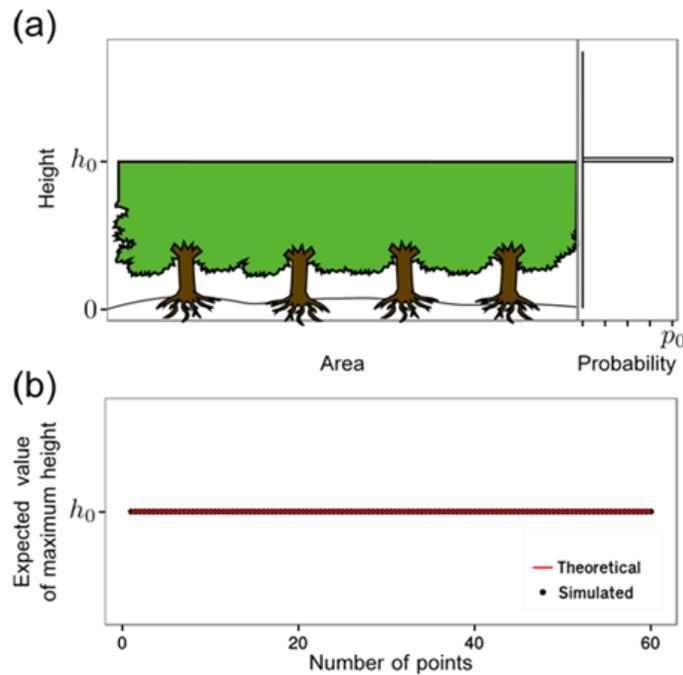


Figure 2: (a) probabilité qu'une impulsion retourne une hauteur selon une canopée toute plate. Une impulsion peut forcément toucher la canopée avec la hauteur  $h_0$  suivant une probabilité  $p_0$ , dont  $p_0 = 1$ . (b)  $h_0$  est la valeur attendue de  $h_{max}$ , peu importe le nombre de points échantillonnés.

### 1.2.3 Une canopée plate avec une singularité

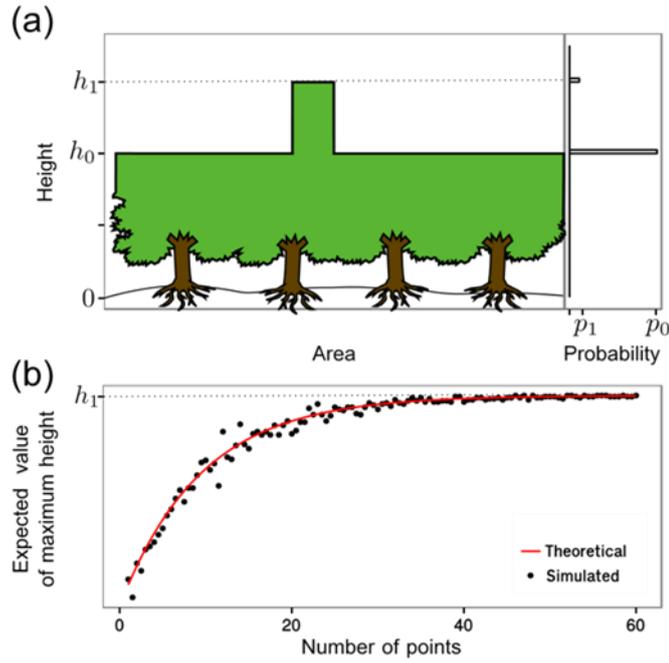


Figure 3: (a) Probabilité qu'une impulsion revienne à une hauteur  $h$  selon une canopée plate avec une singularité. Une impulsion peut revenir à une hauteur  $h_0$  ou  $h_1$  suivant les probabilités  $p_0$  et  $p_1$ . (b) La  $h_{max}$  a été simulée par Roussel et al (2017) en fonction du nombre de points utilisés  $n$  et la valeur  $h$  attendue a été calculée en utilisant l'équation (4).

S'il y a une singularité à la surface, la hauteur maximale observée peut prendre deux valeurs : soit  $h_0$  ou  $h_1$ .

La probabilité d'observer la hauteur maximum à  $h_0$  serait de  $p_0$  et donc la probabilité d'observer le maximum à  $h_1$  serait de  $p_1 = 1 - p_0$ . La probabilité d'observer la valeur attendue de la hauteur maximale pour une seule impulsion est de :

$$\bar{h}(1) = P(X = 1)h_1 + P(X = 0)h_0 = p_1h_1 + (1 - p_1)h_0 \quad (2)$$

Lorsque échantillonné de manière aléatoire avec  $n$  impulsions indépendantes, un seul doit revenir à  $h_1$  pour trouver la hauteur maximale réelle. Ce processus correspond à  $n$  itérations indépendantes et suit une distribution binomiale. Soit  $X$  une variable aléatoire qui compte le nombre de fois où  $h_1$  est omis.  $X$  suit une distribution binomiale  $X \sim B(n, p_0)$  de sorte que la probabilité de  $h_1$  soit la hauteur maximale observée quand au moins une des  $n$  impulsions revient à la hauteur  $h_1$ .

$$P(X < n) = 1 - P(X = n) = 1 - (1 - p_1)^n \quad (3)$$

La valeur attendue de la hauteur maximale observée en fonction du nombre de points  $n$  est donc :

$$\bar{h}(1) = P(X < n)h_1 + P(X = n)h_0 = (1 - p_0^n)h_1 + p_0^n h_0 \quad (4)$$

Parfois, une ou plusieurs impulsions reviennent à la hauteur maximale, mais parfois non, on constate donc un biais.

La Figure 3 (b) montre que la hauteur maximale observée augmente d'abord avec la densité d'impulsions puis se rapproche du maximum réel  $h_1$  de manière asymptotique. Ainsi, 40 impulsions sont nécessaires pour observer la hauteur maximale réelle avec une probabilité élevée. Lorsque la densité est faible, il existe en moyenne un biais, car de nombreuses impulsions reviendront à  $h_0$ . L'objectif est de quantifier le biais indiqué en fonction de la densité et de la forme de la canopée. Mais il faut une équation plus générique qui permettrait de prendre en compte les canopées à plus d'une singularité. Nous avons donc besoin de réécrire les notations de l'équation (4). Comme  $p'_1 = p_1 = \frac{p_1}{p_1+p_0} + p_1$ , car  $p_0 + p_1 = 1$  alors de façon générale, nous avons :

$$P_k^n = 1 - \left(1 - \frac{p_k}{\sum_{i=0}^k p_i}\right)^n \quad (5)$$

Ce qui donne :

$$P_1^n = 1 - \left(1 - \frac{p_1}{p_1 + p_0}\right)^n = 1 - (1 - p_1)^n \quad (6)$$

Substituons  $p_1$  par  $P_1^n$  :

$$\bar{h}(n) = P_1^n h_1 + (1 - P_1^n)h_0 \quad (7)$$

Cette équation peut également être calculée pour quantifier la valeur maximum attendue quand la canopée est échantillonnée avec deux hauteurs. C'est ce que nous allons voir.

#### **1.2.4 Une canopée plate avec deux singularités**

En suivant le même processus d'avant, si nous ajoutons deux singularités à une surface plate, la hauteur maximale observée peut prendre trois valeurs. Dans ce cas, l'histogramme comprendra une troisième hauteur  $h_2$  suivant une probabilité  $p_2$ , qui est la hauteur maximale dans cet exemple. La probabilité de rater la hauteur maximale est donc  $1 - p_2$ . Si  $h_2$  est omis, alors il y a deux possibilités : soit toucher  $h_1$  ou soit  $h_0$ . La probabilité de trouver  $h_1$  et  $h_0$  devient donc  $p'_1 = \frac{p_1}{p_1+p_0}$  et  $p'_0 = 1 - \frac{p_1}{p_1+p_0}$ . D'ailleurs  $p'_2 = p_2 = \frac{p_2}{p_1+p_0+p_2}$ .

Comme l'échantillonnage avec plus d'une impulsion est un processus binomial, la valeur attendue de  $\hat{h}_{max}$  peut être calculée en substituant chacune des probabilités  $p'_i$ , avec  $P_i^n$  comme démontré avec l'équation ( 7 ). Selon deux singularités, l'équation finale est :

$$\bar{h}(n) = p_2^n h_2 + (1 - p_2^n)(p_1^n h_1 + (1 - p_1^n)h_0) \quad (8)$$

Les valeurs attendues dans la Figure 4 (b) montrent à nouveau que la hauteur maximale observée augmente asymptotiquement avec la densité d'impulsions, mais que 20 impulsions seulement sont nécessaires pour observer le maximum réel avec une probabilité élevée. La courbe est plus raide dans ce cas parce que la singularité numéro 2 est plus large et que la singularité numéro 1.

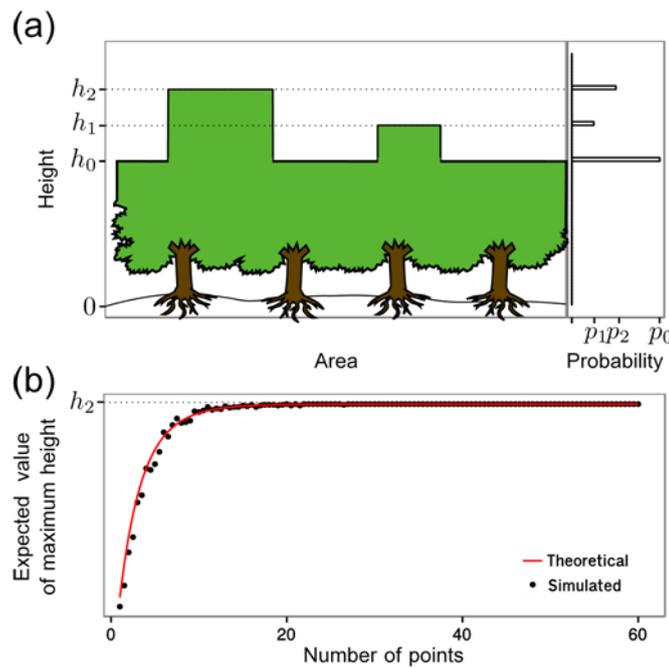


Figure 4: (a) Probabilité d'une impulsion retourne une hauteur selon une canopée à deux singularités. Une impulsion peut retourner trois valeurs de hauteur :  $h_0$ ,  $h_1$ ,  $h_2$  avec une probabilité respective de  $p_0$ ,  $p_1$ ,  $p_2$ . (b) Graphique de Roussel et al (2017) qui représente la hauteur maximale observée qui a été simulée,  $\hat{h}$  a été calculée avec l'équation ( 8 ).

### 1.2.5 Une canopée de forme continue

La Figure 5 nous montre qu'une canopée peut-être discrétiser en utilisant un histogramme avec  $k$  singularités, un pour chaque hauteur  $h_i$  et probabilité  $p_i$  où est  $i$ , compris entre 0 et  $k$ . Lorsque la surface est échantillonnée avec  $n$  impulsions, la valeur attendue de la hauteur maximale est :

$$\bar{h}(n) = P_k^n h_k + (1 - P_k^n)[P_{k-1}^n h_{k-1} + (1 - P_{k-1}^n)(P_{k-2}^n h_{k-2} + \dots)] \quad (9)$$

Cette équation peut être simplifiée sous sa forme récursive. Soit  $H$  l'ensemble des couples hauteur/probabilité :  $H = (h_i, p_i)$  où  $i \in (0, k)$ . Nous pouvons écrire :

$$H_i^n(H) = \begin{cases} h_0 & \text{if } i = 0 \\ P_i^n h_i + (1 - P_i^n) H_{i-1}^n & \text{else} \end{cases} \quad (10)$$

Donc  $\bar{h}(n) = H_k^n(H)$

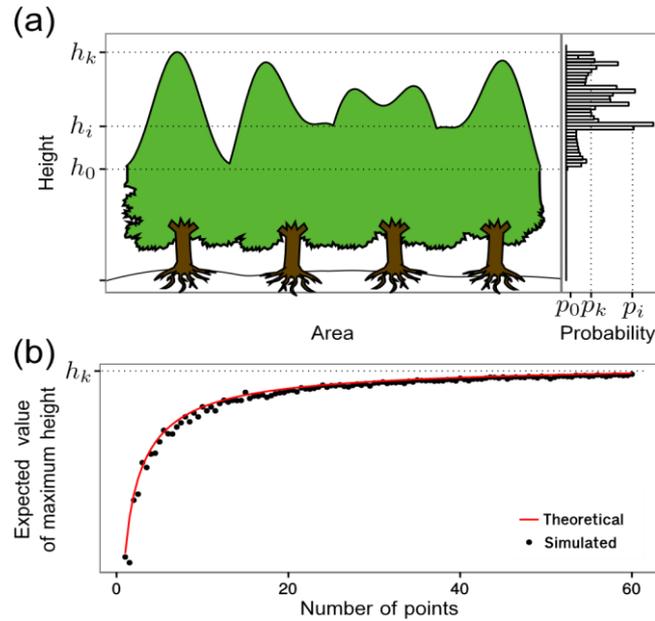


Figure 5: (a) Probabilité d'une impulsion retourne une hauteur  $h$  selon une canopée continue (b) la hauteur attendue a été calculée par Roussel et al (2017) avec l'équation ( 10 ).

### 1.2.6 Quantifications du biais à l'aide d'histogrammes standardisés

La comparaison des valeurs attendues et simulées de Roussel et al. (2017) a montré que la hauteur attendue peut être calculée pour toute forme de canopée et que celle-ci varie en fonction de la forme de la canopée et de la densité d'impulsions.

$$e_r(n) = H_k^n(H_r) \quad (11)$$

$h_{max}$  étant le point de référence, le biais  $e$  est toujours négatif et est calculé à l'aide des histogrammes standardisés.  $H_r$  est un histogramme normalisé avec une résolution  $r$ .

Des histogrammes de la canopée ont été créés pour offrir une visualisation quantitative de la variation verticale de la hauteur. Le nombre de retours dans chaque case reflète la probabilité qu'une impulsion revienne à une hauteur donnée. La forme de l'histogramme reflète donc la distribution verticale des hauteurs des retours. Le point de référence n'est pas le sol, mais la hauteur maximale réelle dans une zone de pixels donnée. Ainsi, les histogrammes ont été

normalisés en soustrayant le maximum local de la hauteur de chaque retour, de sorte que le maximum local soit égal à zéro et que tous les autres retours soient négatifs.

### 1.2.7 Effet de la taille de l'empreinte

La densité des impulsions n'est certainement pas la seule source de biais, Hirata (2004) a montré que la taille de l'empreinte affectait les estimations de la hauteur de la canopée. Il existe donc certainement un second terme d'erreur dans l'estimation de la hauteur maximale. On note  $\delta$  la taille de l'empreinte au sol pour l'équation globale du modèle :

$$h_{max} = \hat{h}_{max} + e_r(n) + \delta \quad (12)$$

D'après Roussel et al (2017), les grandes empreintes sous-estiment la hauteur maximale. La Figure 6 illustre la cause de cette sous-estimation.

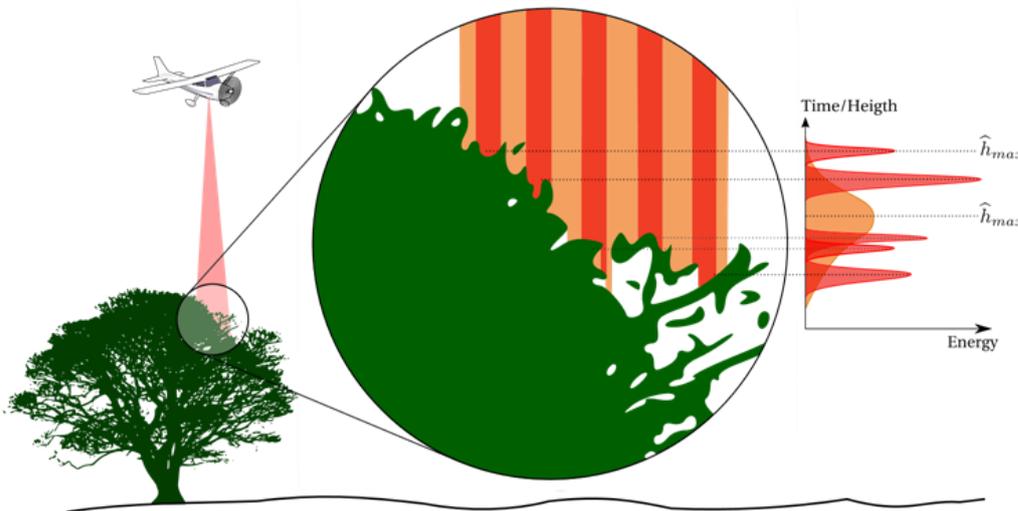


Figure 6: effet de la taille de l'empreinte sur la hauteur maximale observée  $h_{max}$ . La couleur orange en arrière-plan appartient à une empreinte dix fois plus grande que les petites empreintes. Les colonnes rouges appartiennent aux impulsions de faible empreinte. L'empreinte la plus grande a une forme d'onde plus large car l'intensité est intégrée sur une plus grande surface. Par conséquent, la grande empreinte sous-estime la hauteur maximale enregistrée par les plus petites empreintes, cela signifie que la hauteur du pic de la forme d'onde orange est inférieure au pic rouge le plus élevé.

Les colonnes rouges représentent les impulsions de faible empreinte et la colonne orange en arrière-plan représente les impulsions de grande empreinte. La forme d'onde des empreintes est gaussienne. La hauteur renvoyée par les impulsions de petite empreinte est assez précise puisque l'écart-type des ondes est faible. En revanche, les impulsions d'empreinte large induisent une moins bonne précision sur la hauteur maximale car la forme de l'onde est plus

large. Par conséquent, l’empreinte plus large sous-estime la hauteur du maximum local par rapport aux hauteurs maximums renvoyées par des impulsions de faible empreinte.

La méthode consiste à soustraire la hauteur du maximum local renvoyée par les impulsions de faible empreinte d’une hauteur équivalente estimée de l’empreinte importante. En d’autres termes, cela signifie que la distance moyenne entre le pic le plus élevé de la forme d’onde rouge et d’onde orange est estimée.

Roussel et al (2017) suppose que pour l’ensemble des données à haute densité l’empreinte est suffisamment petite pour avoir un effet négligeable sur le biais de la hauteur maximale.  $\delta$  est donc fixé à 0 pour les données à haute densité. Cependant,  $\delta$  est à estimer pour l’ensemble des données à basse et moyenne densité. En supposant que la réflectivité de la canopée ne change pas entre les relevés à basse et à haute densité, la hauteur du pic orange est estimée comme étant le point auquel les plus petites empreintes ont renvoyé 50 % de leur intensité totale. Cette méthode ne fournit probablement pas la valeur correcte dans tous les cas, mais en la calculant sur un grand nombre de pixels ponctuels (0.14 m<sup>2</sup> par exemple), on s’attend à ce qu’elle fournisse une bonne estimation de la hauteur équivalente moyenne.

## **I.3 Matériels et méthodes**

### ***I.3.1 Site d’étude***

Le modèle de Roussel et al. (2017) a été appliqué sur une zone d’étude au Canada dans la réserve forestière et faunique de Haliburton, située dans la région forestière des Grands Lacs et du Saint-Laurent au centre de l’Ontario. La région est plutôt plate puisque l’altitude varie entre 400 et 500 mètres au-dessus du niveau de la mer. La forêt est composée d’un mélange de feuillus et de conifères typiques du nord du Canada. Il y avait deux jeux de données distincts. Le premier ensemble couvrait la totalité de la forêt de Haliburton (320 km<sup>2</sup>) acquis avec une densité standard de faible à moyenne densité, puis le deuxième ensemble était une petite parcelle de 68 ha située dans la forêt de Haliburton acquis avec densité d’impulsions élevée (26 impulsions/m<sup>2</sup> en moyenne).

Son modèle n’a pas encore été appliqué dans une forêt tropicale c’est pourquoi il était intéressant de faire cette expérience sur une forêt comprenant une canopée de grande envergure et plate.

Nous avons choisi cette zone d’étude située dans le parc national de Serra do Mar à Sao Paulo dans le sud-est du Brésil, parce que tout d’abord ce site faisait l’objet de recherche par

les chercheurs et les organismes suivants : l'INPE, la NASA, toutes deux en relation avec mon superviseur de stage dans le cadre de la REDD+ (Leitold et al. 2015). Ce site retient particulièrement l'attention de ces organismes, car il est situé dans le parc de la forêt atlantique (« mata atlântica »). En effet, le parc fait partie de la forêt atlantique du Brésil et de la forêt tropicale humide du Brésil, qui joue un rôle important sur le cycle du carbone mondial. Mais aussi, cette forêt détient une biodiversité extrêmement dense avec un nombre important d'espèces endémiques à la région. Le problème majeur est qu'elle est actuellement menacée par la déforestation du Brésil qui reste très préoccupante à l'échelle mondiale. Le parc aide à modérer le climat et maintenir les pentes, mais vise aussi à protéger les espèces d'oiseaux et animaux en général ainsi que les sources d'eau. Il est donc considéré comme un point d'attention pour la communauté scientifique, les ONG, les gouvernements, les entreprises privées et d'autres secteurs de la société en raison de son souci de la préservation de la forêt atlantique.

De plus, pour appliquer son modèle correctement, il nous fallait un jeu de données à haute densité, en effet une densité de points importante nous donne une bonne représentation de la surface de la canopée. Pareillement, le modèle de Roussel et al. (2017) avait été appliqué sur une petite parcelle de données à haute densité. Ensuite, les chercheurs avaient corrigé le biais qu'ils avaient obtenu sur la totalité de la forêt en émettant comme hypothèse que la structure de la forêt ne changerait pas puisque la petite parcelle était commune à la forêt de Haliburton. Pour notre part, nous allons appliquer ce modèle sur la totalité des données obtenues avec une haute densité.

La zone du levé LiDAR couvre une aire de 998 ha d'après les données de Leitold et al. (2015). Elle se caractérise par un terrain complexe avec un fort dénivelé de 78 à 1200 m. La forêt tropicale humide de cette zone est divisée en plusieurs types de végétation en fonction de l'altitude : forêts de plaine, vallée, montagne. Suivant une distance de 4000 à 6000 m le long d'un transect, on observe une forte pente comprenant une zone forestière de pente. Cette zone de pente raide de 37 degrés environ représente 37 % de la zone d'étude. Les 63 % restants de la zone d'étude sont constitués de forêts de plaine relativement plates qu'on retrouve en région montagneuse sur le plateau et dans la vallée. L'étude inclut neuf parcelles permanentes d'inventaire forestier établies le long d'un transect. Une parcelle est située à une altitude de 100 m, quatre parcelles à environ 1000 m d'altitude. Les parcelles ont chacune une superficie projetée de 1 ha.

### Lidar Data of Serra do Mar Park

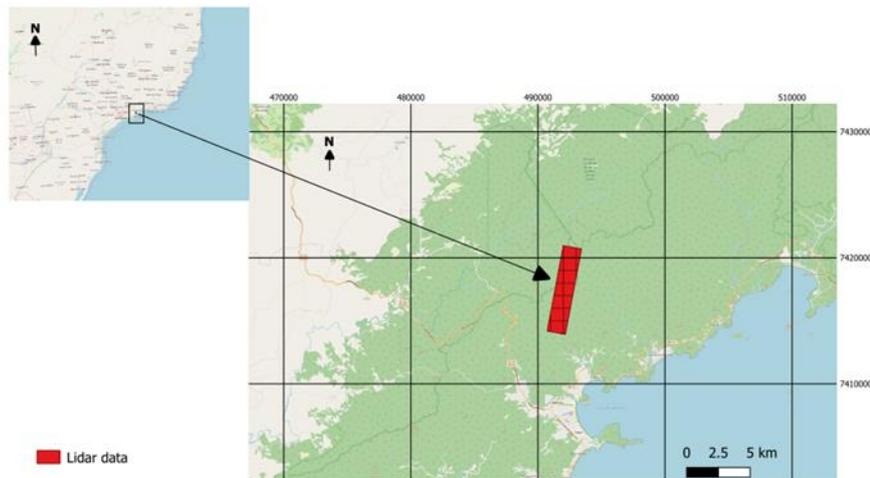


Figure 7: emplacement des données LiDAR

### 1.3.2 Données :

Les données LiDAR du parc national de Serra do Mar se rapportent à une étude dans le cadre du projet Développement Durable, soutenu par l'Agence des Etats-Unis pour le développement international, et le Département d'État des Etats-Unis. Le service forestier des Etats-Unis en collaboration avec l'entreprise brésilienne de recherche agricole (EMBRAPA) a permis de fournir des données LiDAR de haute précision visant à mettre au point de nouvelles méthodes et à générer des connaissances sur le terrain. Le site internet nous donne accès librement avec permission de l'administration du site, au nuage de points et au modèle de terrain numérique acquis en 2012, à la carte de la zone de couverture LiDAR et aux limites des tuiles. Le nuage de points est donné au format LAS classifié. Il est possible de télécharger pour la même zone le nuage de points et le modèle de terrain numérique acquis en 2017. Les coordonnées terrain de l'emprise sont dans le système de coordonnées UTM zone 23S, l'aire d'étude est située à 23°34'S et 45°02'W ; 23°17'S et 45.11 'W. Le jeu de données comprend 21 tuiles donc 21 fichiers laz. La densité de points en moyenne est de 22 points / m<sup>2</sup>, la densité des premiers retours est de 15.09 points / m<sup>2</sup>. Les données LiDAR ont été acquises avec le scanner laser Optech ALTM 3100 d'une longueur d'onde de 1064 nm. La zone d'étude a été survolée avec une hauteur de vol moyenne de 1600 m (au-dessus du niveau de la mer) couvrant une bande rectangulaire de surface d'environ 1.5 km de large sur 7 km de longueur soit une superficie d'environ 1000 ha.

Compte tenu de la différence de niveau d'altitude entre les régions de montagne et de vallée, la distance moyenne entre le capteur et la surface du sol varie entre 660 m en montagne au sommet du plateau et 1320 m dans la vallée. Comme la distance entre la surface du sol et le capteur est deux fois plus grande dans la vallée, on s'attend à ce que la taille de l'empreinte au sol double dans la vallée pour une divergence de faisceau constante. D'après les paramètres du système laser, la divergence du faisceau laser est de 0.25 mrad donc la taille de l'empreinte au sol varie entre environ 16 et 33 cm. Les impulsions laser ont été émises avec un angle de champ de  $11^\circ$ , avec une fréquence de balayage de 25 Hz, puis une fréquence de GNSS de 5 Hz et IMU de 100 kHz. L'incertitude verticale est inférieure ou égale à 15 cm et l'incertitude planimétrique est inférieure ou égale à 15 cm. Le recouvrement entre les lignes de vol est de 30 %.

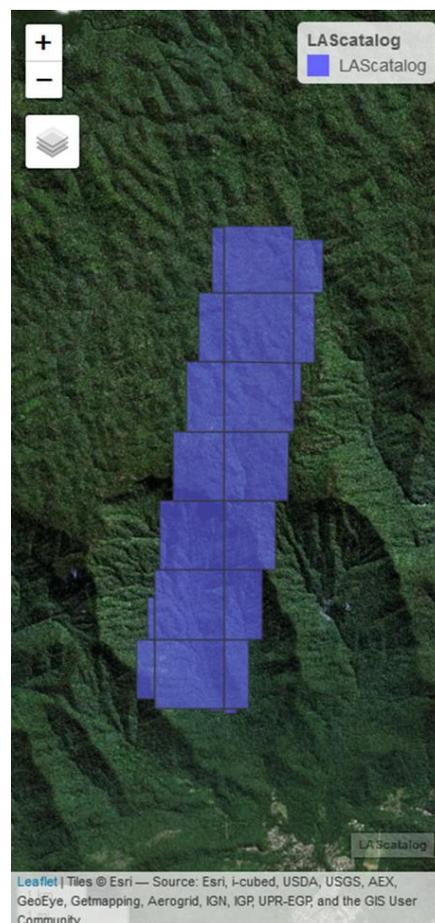


Figure 8: affichage sur le logiciel R de l'ensemble des données LiDAR.

### **I.3.3 Pré-traitement**

#### **I.3.3.1 Classification des points au sol :**

C'est la première étape avant de générer un modèle numérique de terrain. La classification au sol consiste à classer le nuage de points en deux catégories :

- Les points appartenant au sol
- Les points n'appartenant pas au sol

Comme la quantité de données est importante, ce processus de classification doit être établi sur des algorithmes automatisés. La classification est souvent effectuée par les fournisseurs de données comme dans notre cas-ci, les utilisateurs finaux n'ont pas forcément accès aux informations des algorithmes appliqués.

#### **I.3.3.2 Modèle Numérique de Terrain :**

La génération du MNT<sup>2</sup> suit généralement la classification des points au sol comme étant la deuxième grande étape de l'analyse de données LiDAR. Un MNT est décrit comme une image du sol, il représente donc la surface du sol. Il est utilisé dans de nombreuses applications telles que la détermination des bassins versants de rétention d'eau, de débit, la génération de la normalisation de nuage de points. En effet, il permet aux utilisateurs de soustraire le terrain de l'élévation des points pour manipuler le nuage de points comme s'il avait été acquis sur une surface plane.

La construction du MNT repose simplement sur une interpolation spatiale des points au sol à des emplacements non échantillonnés. La précision du MNT est importante, car des erreurs dans le MNT provoqueront des erreurs d'estimation de la hauteur des arbres (Leitold et al. 2015). Plusieurs méthodes de génération du MNT ont été étudiées pour diverses situations de terrain et ils sont tous applicables pour les données LiDAR aérien, mais ne sont pas tous forcément disponibles dans les logiciels dédiés. Chen, Gao, et Devereux (2017) critiquent les différentes possibilités de génération de MNT. Dans cette étude, nous n'avons peu d'informations concernant la construction du MNT, mais nous savons quels logiciels a traité le nuage de points donnés par le fournisseur : PosPac, DashMap, TerraScan et logiciels CAO.

---

<sup>2</sup> MNT : Modèle Numérique de Terrain

### **I.3.3.3 Filtre**

Le filtrage permet de retirer les artefacts des données laser comme les points hauts (nuages, oiseaux) et bas.

Le fournisseur de données a dû en effet utiliser un premier filtrage avec ses paramètres sans que l'on sache lesquels, mais les données brutes comportaient tout de même des points aberrants qui ne représentaient ni le sol ni la végétation. Des points très hauts qui étaient probablement dus à des retours d'impulsions laser qui ont frappé des oiseaux, car comme on l'a vu ce parc regorge d'oiseaux typiques de la forêt tropicale.

Le filtrage est une opération délicate, car nous devons arriver à garder les points utiles et à supprimer les points qui empêchent d'obtenir de bons résultats.

L'élimination des artefacts au-dessus de la canopée peut se faire en fonction de plusieurs métriques statistiques telles que hauteur maximale, hauteur moyenne, percentiles de hauteur, écart-type...

La suppression des points bas peut s'appliquer avant ou après normalisation. Après normalisation, il suffit de supprimer tous les points en dessous de l'altitude 0.

Par communication personnelle avec Roussel et al nous avons choisi d'appliquer un algorithme de filtrage de points qui permet de supprimer les points au-dessus de la canopée en fonction de la hauteur du 95<sup>ème</sup> percentile de chaque arbre afin de résoudre le problème évoqué précédemment. Pour cela, nous avons utilisé des fonctions du package « lidR » développées par Roussel et al. (2019) sur le logiciel de programmation R, (tel que « grid\_metrics », « lasmergespatial » et « lasfilter »). Avec ces fonctions, nous avons créé une fonction qui retourne un nuage de points sans artefacts.

Nous avons supprimé également les points en dessous du sol après la normalisation, processus qui va être expliqué par la suite.

### **I.3.3.4 Normalisation**

La troisième étape de l'analyse classique de données LiDAR est la normalisation des données LiDAR. Il consiste à enlever le terrain à partir du nuage de points permettant ainsi la représentation d'un sol plat. Avoir un sol plat facilite l'analyse et donc le traitement du nuage de points. Cette étape permet donc de conduire directement à l'analyse statistique de la distribution des hauteurs.

Deux méthodes peuvent être utilisées pour effectuer cette tâche :

- Utiliser la représentation raster du MNT
- Interpoler entre chaque point

#### ***1.3.3.4.1 Normalisation basée sur un MNT raster***

C'est la méthode de normalisation la plus courante, par exemple utilisée par Jakubowski, Guo, et Kelly (2013). Elle a pour objectif de soustraire le MNT raster sur tous les points du nuage. Elle est simple et rapide à mettre en œuvre, pour chaque point du jeu de données, l'algorithme recherche la valeur du pixel du MNT correspondant, et soustrait cette valeur de la valeur de l'altitude du point. En revanche, l'inconvénient de cette méthode est qu'elle provoque des inexactitudes à cause de la nature discrète du MNT. Tous les points correspondant au sol ne sont pas individuellement normalisés à 0, car le MNT a été interpolé à l'aide de points régulièrement espacés qui ne correspondent pas forcément à l'emplacement des points au sol du jeu de données. Ce qui a pour conséquence des points qui sont positionnés à plus de 0 ou moins de 0. Le format raster reste un moyen simple de stocker, visualiser et partager un MNT, mais amène des imprécisions.

#### ***1.3.3.4.2 Normalisation par interpolation entre points :***

Cette méthode est basée sur l'interpolation de tous les points, chaque point classé au sol est interpolé à sa position exacte. Le MNT n'est plus un raster mais un nuage de points ce qui a pour conséquence de supprimer les inexactitudes relatives à la représentation du terrain. Cette méthode permet d'assurer que chaque point utilisé comme référence à 0 est bien à normaliser à 0 ce qui permet d'éviter considérablement les points négatifs. Cependant, certains points peuvent tout de même encore apparaître en dessous de 0 à cause des imprécisions de la méthode d'interpolation ou de classification. Le logiciel « Lastools » qui est très utilisé en foresterie utilise cette méthode, ainsi que Roussel et al. Quant à nous, nous avons utilisé également ce processus avec la méthode d'interpolation par TIN<sup>3</sup>. Les points qui figuraient encore en dessous du sol normalisé ont été supprimé automatiquement.

---

<sup>3</sup> TIN : Triangular Irregular Network

### 1.3.3.4.3 Limite de la normalisation des données

L'inconvénient de la normalisation est qu'elle implique une distorsion du nuage de points et donc des objets échantillonnés comme les arbres ou bâtiments. Khosravipour et al. (2015) et Vega et al. (2014) nous mentionnent que cet effet est accru pour des objets de grandes tailles horizontales et lorsque le terrain est très incliné (Figure 9). Il n'y a pas de solution dans la littérature pour résoudre ce problème. Nous rencontrons aussi ce même problème lorsque nous avons effectué la normalisation, car les arbres sont dotés de grande couverture végétale horizontale et que le terrain est très pentu.

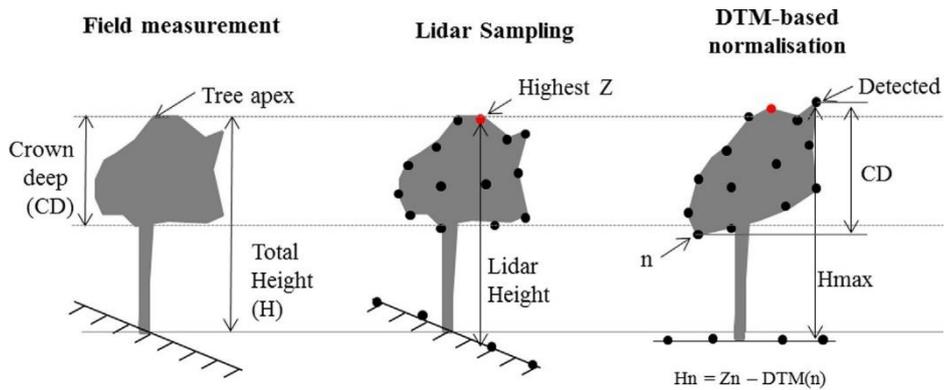


Figure 9: Illustration de Véga et al (2014) de l'effet de la normalisation sur la géométrie du nuage de points dans le cas de fortes pentes.

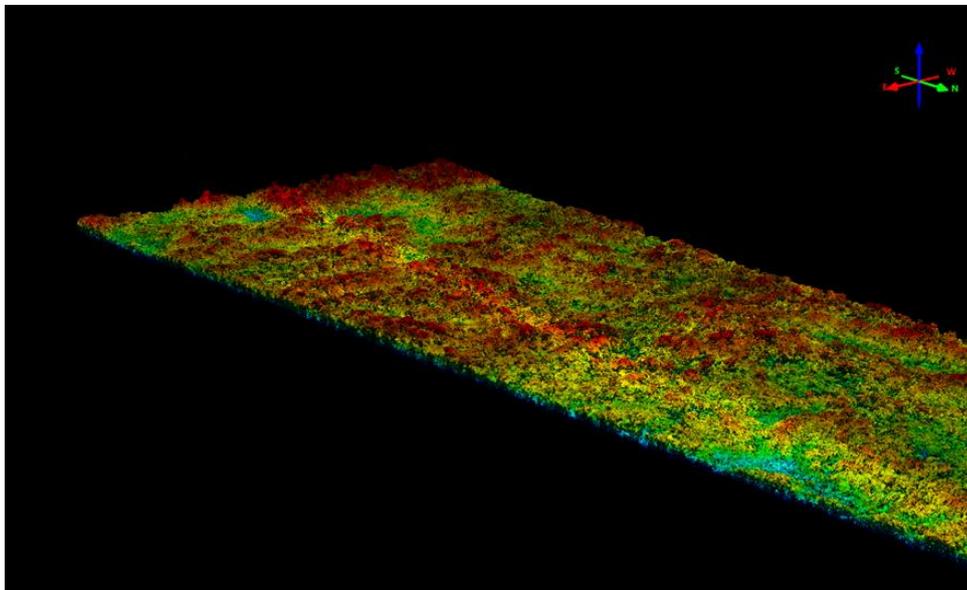


Figure 10 : résultat de la normalisation d'une partie des données affichées sur le logiciel Quick Terrain Reader

### I.3.3.5 Modèle de Hauteur de Canopée :

Le modèle de hauteur de canopée est une surface numérique représentant le sommet du couvert forestier. Il représente ainsi la hauteur des arbres. Il ne s'agit pas d'une valeur d'altitude, mais bien de la hauteur ou de la distance entre le sol et le sommet des arbres.

Voici une illustration qui montre le principe :

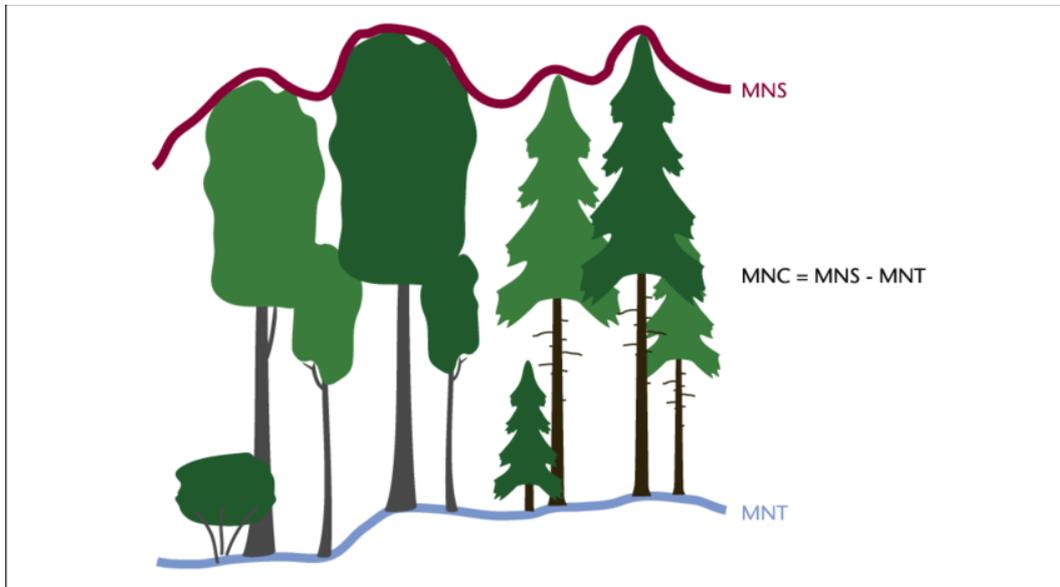


Figure 11 illustration qui montre le principe du modèle de hauteur de canopée, source: [https://www.researchgate.net/figure/Illustration-du-principe-de-creation-du-MNC-a-partir-des-MNS-et-MNT\\_fig1\\_301228831](https://www.researchgate.net/figure/Illustration-du-principe-de-creation-du-MNC-a-partir-des-MNS-et-MNT_fig1_301228831)

Le modèle de hauteur de canopée désigne donc la différence du MNS<sup>4</sup> et le MNT. Il informe directement sur la hauteur de la canopée c'est-à-dire la hauteur des éléments constitutifs des formations végétales par rapport à la surface du sol.

Il désigne la surface normalisée (Ruiz et al. 2014) alors que le MNS désigne la même surface non normalisée (Ruiz et al. 2014). Il existe deux grandes variantes d'algorithmes pour calculer un MHC<sup>5</sup> :

- Algorithmes de points vers raster
- Algorithmes basés sur la triangulation

---

<sup>4</sup> MNS : Modèle Numérique de Surface

<sup>5</sup> MHC : Modèle de Hauteur de Canopée

#### ***1.3.3.5.1 Algorithme de point vers raster***

Les algorithmes de points vers raster consistent à construire une grille de pixel à une résolution donnée et à attribuer à chaque pixel l'altitude du point le plus élevé de ce pixel. En termes de temps calcul l'algorithme est rapide et simple à mettre en place c'est pourquoi il a souvent été cité dans la littérature. (Alexander, Korstjens, et Hill 2018 ; Roussel et al. 2019; Véga et al. 2016). C'est l'algorithme implémenté par défaut dans la plupart des logiciels comme FUSION, Lastools, Arcgis. L'inconvénient de cette méthode est que certains pixels peuvent être vides si la résolution de la grille raster est trop fine pour la densité de points disponibles. Il se peut donc que certains pixels se retrouvent sans valeurs, car ils ne contiennent aucun point. Ce qui implique une deuxième étape de post-traitement pour résoudre ce problème en utilisant une méthode d'interpolation.

#### ***1.3.3.5.2 Algorithmes basés sur la triangulation***

Le TIN<sup>6</sup> a été développé à l'origine au début des années 70. Il s'agissait simplement de construire une surface à partir d'un ensemble de points répartis dans l'espace de façon irrégulière. Dans un TIN, les points sont connectés par des segments de lignes pour former des triangles, ils garantissent la continuité de la surface. Pour construire les triangles, il existe plusieurs méthodes comme le classement selon la distance ou la triangulation Delaunay. L'algorithme par le classement selon la distance calcule la distance entre toutes les paires de points, puis est trié de la plus petite à la plus grande. Les paires de points les plus rapprochés sont ensuite reliées. Puis c'est au tour de la seconde paire de points les plus rapprochés et ainsi de suite jusqu'à ce qu'on ne puisse plus former de ligne sans croiser de ligne existante. Cette méthode n'est pas utilisée en pratique, car elle peut produire des triangles aigus. Pour éviter ce problème, la triangulation de Delaunay qui produit des triangles plus obtus est utilisée. Cet algorithme repose par définition sur un cercle circonscrit qui passe par trois points sans en contenir d'autres.

Une fois triangulée, une interpolation des premiers retours dans chaque triangle est opérée pour calculer l'altitude de chaque pixel de la grille raster. Isenburg et al. (2006) nous montre l'utilité de cet algorithme lors d'un nombre important de points acquis par relevé LiDAR

---

<sup>6</sup> TIN : Triangular Irregular Network

L'approche par triangulation présente l'avantage surtout de ne pas laisser des pixels vides, quelle que soit la résolution du raster en sortie, tandis que la méthode de point vers raster provoque des espaces vides et des creux qui sont dus aux nombreuses impulsions laser qui parviennent à pénétrer profondément dans le couvert avant de produire le premier retour. La profondeur et la distribution des creux dans une norme MHC dépendent de la structure de la couronne de l'arbre et du diamètre du faisceau laser ainsi que la sensibilité de la forme de l'onde de retour (Gaveau et Hill 2003). Au lieu de frapper le point le plus élevé de la canopée, les impulsions laser peuvent produire leur premier retour lorsqu'elles touchent une branche inférieure ou même après avoir pénétré dans la cime jusqu'au sol. (Ben-Arie et al. 2009; Khosravipour et al. 2014).

Pour éviter ce problème, le MNT est souvent lissé afin de produire une surface plus réaliste avec moins de creux. D'autres algorithmes pour éviter ces bruits ont été créés pour éviter cette étape de post-traitement. (Khosravipour et al. 2014) ont développé un algorithme "pit-free". Il comprend deux étapes : la première étape est de construire une triangulation et un raster qui représentent le MHC standard de tous les premiers retours et la deuxième de construire une série de MHC partiel des premiers retours qui correspondent aux impacts des impulsions laser de végétation plus élevée.

Comme la profondeur des différents creux de la canopée varie considérablement, l'utilisation d'un seuil fixe pour définir les MHC partiels devient impossible. C'est pourquoi un ensemble de MHC partiel a été calculé en excluant tous les premiers retours en dessous d'une certaine hauteur. Pour chaque seuil de hauteur, le réseau triangulaire est nettoyé des triangles trop larges, puis rasterisé. En dernière étape, les rasters partiels sont combinés et seuls les pixels les plus élevés de chaque raster sont conservés. La sortie est un MHC qui est nativement sans creux et sans utiliser de post-traitement. Une étude faite par Khosravipour, Skidmore, et Isenburg (2016) décrit le développement d'un algorithme « spike-free » qui consiste à construire un TIN en utilisant tous les retours sans se préoccuper des points responsables de « creux » dans le maillage du TIN. Une triangulation de Delaunay est utilisée de manière sélective pour tous les retours. Une autre approche est prise en compte par Roussel et al. (2019) pour améliorer ces algorithmes décrits ci-dessus, elle consiste à remplacer chaque retour par un petit disque avec un diamètre prédéfini. Comme le rayon laser a un diamètre il est normal de considérer que c'est un cercle au lieu de point avec une surface nulle. Cet ajustement densifie le nuage de points et réduit le nombre de pixels vides ou creux en lissant naturellement le MHC sans avoir besoin de post-traitement. Pour

conclure, il y a donc plusieurs méthodes disponibles pour calculer un MHC, le choix de différents algorithmes influence grandement la qualité des résultats et l'analyse ultérieurs, par exemple, la segmentation des arbres.

Dans notre étude, nous avons employé la méthode basée sur la triangulation de Delaunay pour créer le MHC. Pour éviter les pixels creux et vides, nous avons appliqué l'algorithme « pit-free » de (Khosravipour et al. 2014) avec l'ajustement des « cercles » de Roussel et al. (2019) pour obtenir des résultats favorables concernant notre modèle de hauteur de canopée<sup>7</sup>.

### ***1.3.4 Traitement du modèle de Roussel et al.***

Nous avons utilisé l'équation ( 10 ) de la partie « cadre théorique du biais » ci-dessus ainsi qu'une fonction histogramme standardisée pour appliquer le modèle de biais sur le jeu de données à haute densité au Brésil. Ces équations et ces fonctions ont été écrites en langage R sous le logiciel Rstudio.<sup>8</sup>

Après avoir rasterisé l'ensemble des données normalisées de cette étude à une résolution de 10 x 10 m, nous avons appliqué la fonction histogramme qui permet de retourner la valeur de chaque hauteur attendue dans chaque pixel, en fonction d'un seuil défini et d'un pas. Nous avons défini le pas à 10 cm. Nous avons calculé la probabilité qu'une impulsion revienne à une hauteur donnée pour pouvoir afficher l'histogramme de la canopée.

Nous avons considéré le biais associé à la taille d'empreinte ( $\delta$ ) sur la hauteur maximale à 0 comme Roussel et al (2017), car les tailles des empreintes au sol sont petites (16 à 33 cm) donc son effet est considéré comme négligeable.

## **1.4 Résultats**

Après application du modèle de Roussel et al. (2019) sur le jeu de données à haute densité de la forêt atlantique du Brésil, nous obtenons les résultats suivants :

---

<sup>7</sup> Voir annexe 2

<sup>8</sup> Voir annexe 1

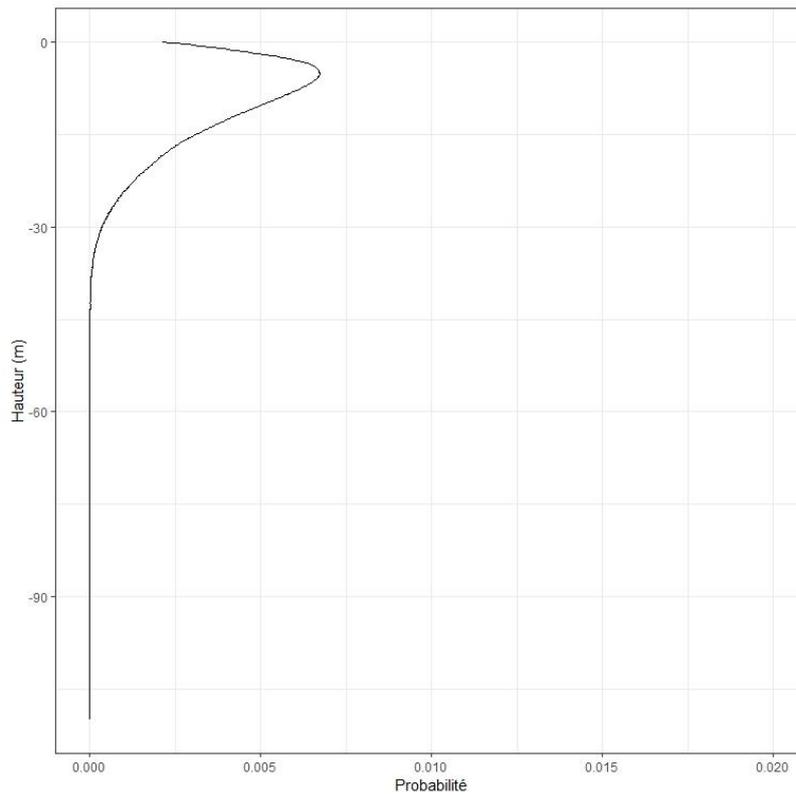


Figure 12: graphique des hauteurs de chaque impulsion en fonction des probabilités de chaque retour associé à une hauteur donnée suivant une canopée de forêt tropicale au Brésil, calculé avec une résolution de pixel de 400 m<sup>2</sup>

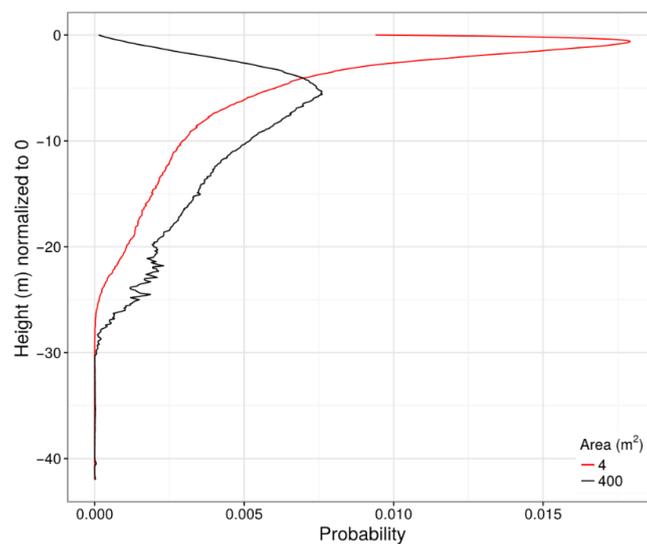


Figure 13: résultats de Roussel et al (2017) : graphique des hauteurs de chaque impulsion en fonction des probabilités de chaque retour associé à une hauteur donnée suivant une canopée de forêt typique d'Amérique du Nord, calculé avec une résolution de pixel de 400 m<sup>2</sup>.

D'après le graphique obtenu de la Figure 12, nous pouvons observer la forme de la canopée sachant que la hauteur maximale est considérée à zéro. Avec une échelle d'observation de

400 m<sup>2</sup>, nous obtenons une probabilité qui augmente à partir de -30 m de hauteur jusqu'à atteindre une probabilité maximale associée à une hauteur de -5 m environ de la hauteur maximale. Les hauteurs attendues les plus probables sont donc à -5 m de la hauteur maximale. Cela signifie que la forme de la canopée a la plus grande probabilité de se concentrer à -5 m de la hauteur maximale de la canopée. Par cet histogramme, nous pouvons ainsi identifier la représentation de la forme de la canopée.

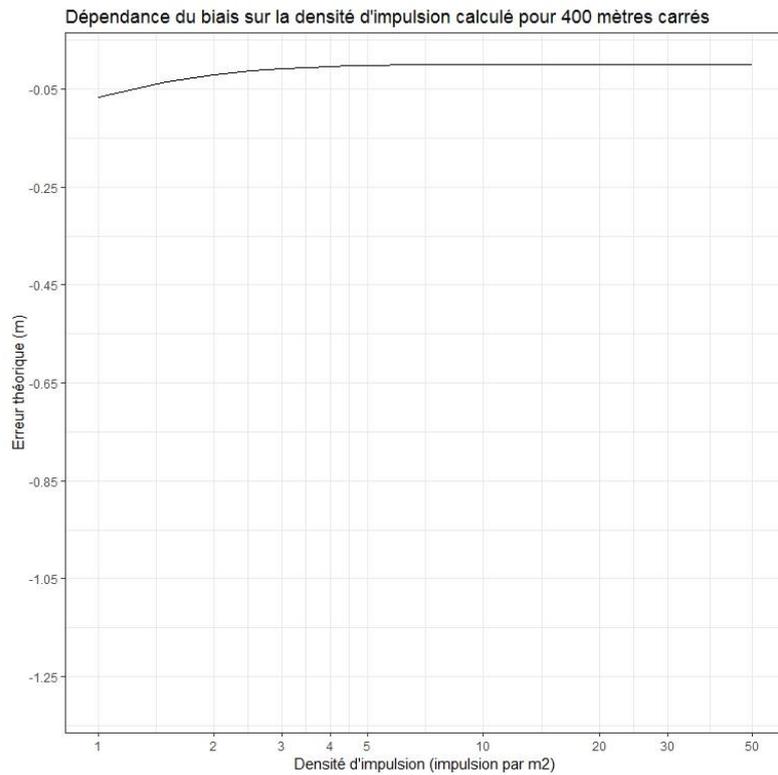


Figure 14: évaluation du biais de l'estimation de la hauteur maximale avec une résolution de 400 m<sup>2</sup> en fonction de la densité d'impulsions

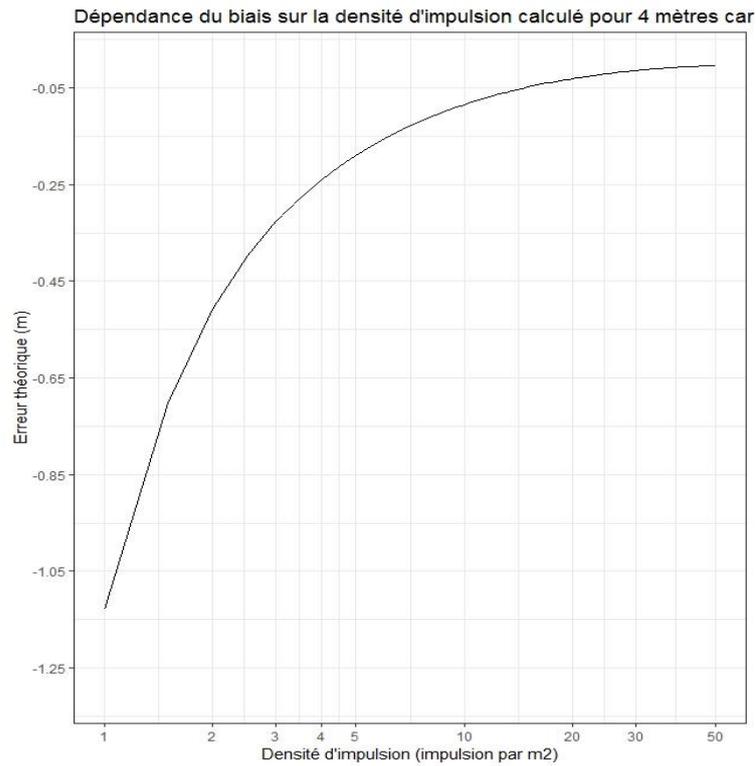


Figure 15: évaluation du biais de l'estimation de la hauteur maximale avec une résolution de 4 m<sup>2</sup> en fonction de la densité d'impulsions

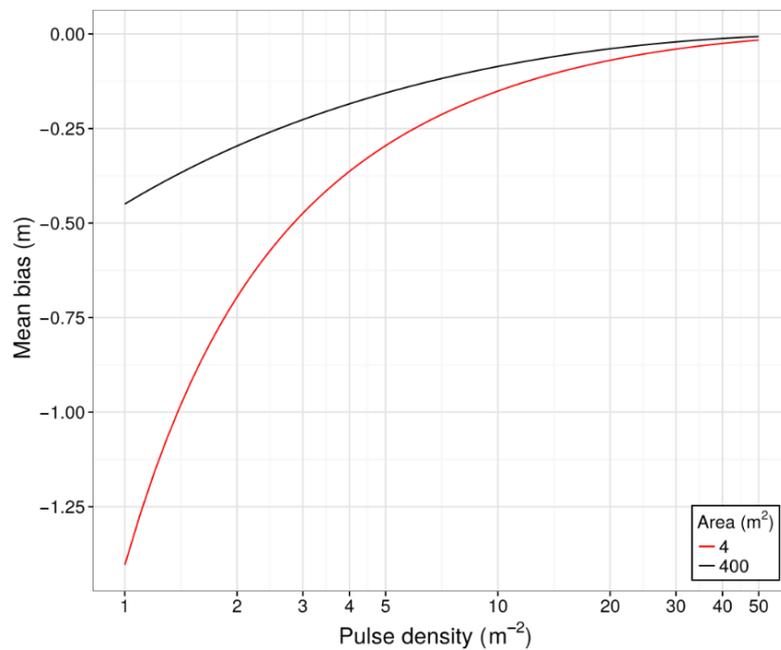


Figure 16 : résultats de Roussel et al (2017): évaluation du biais de l'estimation de la hauteur maximale en fonction de la densité d'impulsions

Avant d'obtenir ces résultats, nous attendions à ce que les valeurs de biais soient plus faibles sur cette zone d'étude que celles obtenues par Roussel et al, (2017). A partir d'une densité

moyenne de points : 4 à 7 impulsions/m<sup>2</sup> nous pensions que le biais allait par la suite fortement diminuer, puisque le site d'étude présente des feuillus avec une canopée tropicale de grande envergure et de couverture forestier plutôt plate dans l'ensemble, par rapport au site d'étude de Roussel et al (2017) qui présente une forêt mixte constituée de conifères avec une forme géométrique conique où la probabilité de l'impulsion d'obtenir la hauteur maximale est plus faible. Par conséquent, le biais attendu globalement devait être moins important que le biais obtenu de Roussel et al (2017), ainsi la courbe représentant le biais en fonction de la densité devrait tendre vers 0 rapidement selon une échelle d'observation haute (400 m<sup>2</sup>).

La Figure 14 nous indique justement cette évolution du biais moyen de l'estimation de la hauteur maximale en fonction de la densité d'impulsions. Avec une échelle d'observation de 400 m<sup>2</sup>, nous observons que le biais moyen commence à se stabiliser à partir de 4 à 5 impulsions par m<sup>2</sup>. Après 5 impulsions par m<sup>2</sup>, nous voyons la courbe tendre vers la hauteur maximale. Le biais commence à environ 7 cm de la hauteur maximale à partir de 1 impulsion par m<sup>2</sup> puis tend vers 0 passée après 5 impulsions par m<sup>2</sup>.

Ce biais n'est vraiment pas important même à partir 1 impulsion par m<sup>2</sup>, quand il est observé avec une échelle d'observation de 400 m<sup>2</sup>. En revanche avec une échelle d'observation de 4 m<sup>2</sup> (Figure 15), l'évolution de ce biais présente un autre comportement : c'est à partir de 20 à 30 impulsions par m<sup>2</sup> que la courbe tend vers 0. La courbe reste un peu près semblable à celle obtenue par Roussel et al, (2017) (Figure 16) à une échelle d'observation de 4 m<sup>2</sup> exceptée que la courbe que nous avons eue démarre avec une plus faible erreur : -1.10 m au lieu d'environ -1.40 m.

Quand on compare les résultats obtenus avec ceux de Roussel et al. (2017) (Figure 16) avec la même échelle d'observation 400 m<sup>2</sup>, nous remarquons que le biais est moins important que Roussel et al. (2017), à partir de 1 impulsion par m<sup>2</sup>. En effet, il obtient un biais de 40 cm environ à partir de 1 impulsion par m<sup>2</sup>, il faut qu'il atteigne 10 impulsions par m<sup>2</sup> pour obtenir un biais moyen inférieur à 10 cm. Tandis que le modèle appliqué sur notre site d'étude indique un biais moyen inférieur à 10 cm à partir de 1 impulsion par m<sup>2</sup> et tend rapidement vers un biais nul.

## I.5 Discussion

Avec une échelle d'observation de 400 m<sup>2</sup>, le biais moyen des résultats de Roussel et al. (2017) se stabilisait à 10 impulsions par m<sup>2</sup> pour estimer une hauteur maximale avec une précision acceptable (biais moyen < 10 cm). Or nous obtenons un biais qui se stabilise à 4 à 5 impulsions par m<sup>2</sup> avec une précision du biais moyen inférieur à 5 cm. Pourquoi ? Nous supposons que ce phénomène est certainement dû à la forme de la canopée de la forêt tropicale qui diffère par rapport à la structure de forêt du jeu de donnée de Roussel et al. (2017). En effet, le couvert de la forêt tropicale du Brésil est relativement grand et plat, ce qui expliquerait ce biais si peu élevé sur l'estimation de la hauteur en fonction de la densité de points. Cependant, le biais reste grand avec une échelle d'observation de 4 m<sup>2</sup> mais n'est pas représentatif de la réalité selon les communications personnelles de Roussel et al (2017), car il y a besoin d'une représentation globale de la canopée pour évaluer le biais, un pixel de surface de 4 m<sup>2</sup> n'est pas suffisant pour l'évaluation du biais de la hauteur maximale de la canopée.

Roussel et al. (2017) a dû corriger ces données à faible densité en supprimant le biais moyen sur la hauteur maximale. Il a corrigé pour chaque pixel en fonction de leur densité. Dans notre cas, ce processus de correction n'est évidemment pas nécessaire puisque nous détenons des données à haute densité, mais si nous étions dans le même cas que Roussel et al. (2017) c'est-à-dire avoir des données de faible densité sur la majorité du territoire et détenir une petite parcelle de haute densité, nous n'aurions pas nécessairement besoin de corriger le biais de l'estimation de la hauteur maximale.

Néanmoins, le modèle comporte des limites, il faut avoir des données à haute densité pour pouvoir avoir une représentation suffisante de la canopée et pour ainsi appliquer son modèle théorique. Dans son cas, il avait des données à haute densité sur une petite parcelle, et il a appliqué son modèle théorique sur tout le site d'étude. L'ensemble du territoire comportait les mêmes essences d'arbres globalement, avec une structure de forêt qui différait peu. Il a pu donc corriger les erreurs sur le territoire en entier. La forêt canadienne d'Amérique du Nord comporte des types d'arbres différents tels que les feuillus et conifères, et leurs formes géométriques diffèrent beaucoup. Les feuillus présentent une canopée plutôt plate tandis que les conifères présentent des formes triangulaires. La réception du signal LiDAR se comporte différemment sur des conifères ou feuillus, ce phénomène influence donc grandement l'estimation de la hauteur maximale. Ce qui expliquerait pourquoi les résultats obtenus sont

différents sur une forêt tropicale qui présente que des feuillus. Dans le cas d'une forêt de conifères ou mixte, il faudrait donc au moins une parcelle de forêt à haute densité de même type que l'ensemble de la forêt auquel on voudrait corriger le biais de l'estimation de la hauteur. Dans le cas de forêt tropicale, il semblerait qu'il n'aurait pas besoin de corriger le biais puisque ce modèle indique que le biais est faible sur ce site d'étude. La question est de savoir si le modèle est-il applicable sur tout type de territoire et de forêts. Par communication personnelle de Roussel et al. (2019) le modèle théorique devrait être généraliste puisqu'il est fondé sur la théorie des probabilités. En tout cas, nous allons observer l'effet de l'estimation de la hauteur sur la biomasse en fonction de plusieurs densités de points, et vérifier ainsi son modèle théorique.

## **II Effet de l'estimation de la hauteur sur la biomasse forestière du Brésil**

### **II.1 Cadre théorique :**

#### ***II.1.1 Définition de la biomasse***

Le terme biomasse forestière désigne en théorie tout le matériel biologique provenant des arbres, plantes. Il inclut les troncs, les branches, l'écorce, les aiguilles, les feuilles, la cime, la souche et les racines. Il est possible de convertir cette biomasse en biocombustibles solides, liquides, ou gazeux, que l'on peut brûler pour produire de l'énergie, ou utiliser comme combustibles de remplacement pour le transport ou les procédés industriels.

Les arbres sont utiles sur le plan énergétique, ils convertissent l'énergie du soleil en biomasse grâce à la photosynthèse : un processus qui absorbe du dioxyde de carbone de l'atmosphère. Il est considéré comme un poids anhydre des diverses composantes biologiques de l'écosystème, en tonnes/ha. Les arbres constituent donc un enjeu important au niveau planétaire puisqu'il absorbe le dioxyde de carbone, qui aujourd'hui est beaucoup rejeté dans l'atmosphère par les activités humaines.

La mesure et la surveillance de la biomasse aérienne de la forêt sont devenues un sujet de recherche important ces dernières années, en raison de son importance capitale pour les négociations climatiques internationales qui vise à réduire les émissions de gaz à effet de serre associé à la déforestation et à la dégradation des forêts ainsi qu'une meilleure

quantification des stocks de carbone (Leitold et al. 2015). Le projet international REDD + est mis en place pour cet objectif visant à réduire les émissions de carbone.

Pour bien effectuer cette quantification, plusieurs technologies de télédétection sont utilisées notamment le LiDAR, qui permet d'obtenir une mesure directe de variables de structure du couvert forestier par exemple la distribution de la hauteur de la canopée. C'est une technologie qui convient donc bien à l'estimation de la biomasse.

### ***II.1.2 Systèmes de télédétection pour quantifier la biomasse***

Divers systèmes de télédétection sont utilisés pour mesurer les caractéristiques biophysiques des forêts comme la biomasse. Ils sont classés généralement par type de capteur :

- Capteurs passifs : spectromètres, dispositifs qui mesurent le rayonnement réfléchi ou émis par les surfaces de la Terre
- Capteurs actifs : génèrent et émettent de l'énergie en interne puis mesurent l'énergie retournée.

Les capteurs passifs mesurent différentes longueurs d'onde du rayonnement solaire réfléchi donnant des informations géométriques pouvant être liées aux propriétés biophysiques de la végétation (Rosette et al. 2012). Le capteur du satellite Landsat (mesure du spectre visible) est un exemple de système passif, il est plus sensible aux propriétés optiques. Comme les capteurs optiques utilisent l'énergie réfléchie ou absorbée par les strates de végétation les plus élevées, ils sont moins sensibles à la structure de la végétation (Steininger et al., 2000). La résolution spatiale des systèmes optiques passifs change selon le capteur de 2 m à 250 m environ.

Par contre, les capteurs actifs émettent une énergie et mesurent cette énergie retournée. Les systèmes actifs comprennent le LiDAR et le radar. Ce dernier mesure la structure de la forêt avec la réflexion des micro-ondes selon une résolution spatiale élevée (25 -100 m), et son avantage est qu'il peut faire l'objet de mesure indépendamment de l'heure du jour ou de la couverture nuageuse.

Des capteurs optiques et radars combinés avec le LiDAR ont été utilisés pour développer des modèles améliorant les estimations de biomasse. La combinaison de ces plusieurs capteurs permet une meilleure précision que le LiDAR seul, même lui seul peut obtenir des précisions plus que convenables.

L'utilisation du LiDAR s'est rapidement accélérée ces dernières années. Les systèmes LiDAR émettent en général des longueurs d'onde dans le proche infrarouge entre 900 et 1064 nm, enregistrent le temps de parcours de l'impulsion laser émise et retournée au capteur. Cet intervalle temps-retour permet de calculer la distance parcourue de l'impulsion laser. Ces caractéristiques se rapportent à la largeur du faisceau laser et l'intensité de l'énergie du laser réfléchi. Le faisceau laser à faible empreinte au sol se dessine sur une surface d'un diamètre inférieur ou égal à 50 cm, fournissant des données de plus en plus utilisées dans le domaine forestier, mais appliquées à des zones relativement petites. Les capteurs à empreinte moyenne ou grande ont pour la plupart un diamètre supérieur à 50 cm au sol.

Deux types de capteurs LiDAR existent, le LiDAR à retour discret et le LiDAR à retour complet. Le LiDAR à retour discret enregistre chaque retour et combine du premier au dernier retour. Tandis que le LiDAR FWF<sup>9</sup> enregistre la forme de l'onde complète. Ces deux types de LiDAR permettent d'obtenir les métriques de la structure de la forêt telles que la hauteur maximale du couvert forestier, les hauteurs des strates de la végétation au-dessus du sol.

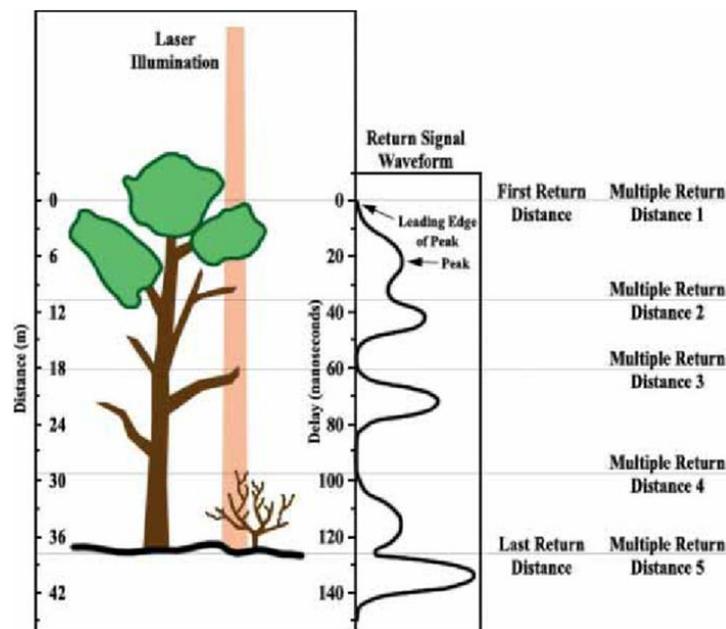


Figure 17: Retour de la forme des impulsions (Fallah et Onur (2011))

Ces différentes métriques dérivées du LiDAR peuvent être liées à l'estimation de la biomasse à l'aide de modèles corrélatifs, avec de mesures sur le terrain correspondantes (Rosette et al.

<sup>9</sup> FWF (Full WaveForm) : onde continue

(2012)). Le LiDAR est une technologie qui est rendue courante dans le domaine forestier en raison de son approche directe pour mesurer la structure de la forêt et pour les précisions relativement plus élevées rapportées dans la littérature.

### **II.1.3 Différentes approches pour mesurer la biomasse :**

#### **II.1.3.1 L'approche par zone**

L'objectif de l'approche par zone est de créer un modèle prédictif qui relie les variables LiDAR X à un attribut d'inventaire forestier Y, qui peut être mesuré à partir d'inventaire terrain. Par la suite, le modèle dérivé est utilisé pour estimer Y partout où X est connu. Les variables du LiDAR aérien sont utilisées comme prédicteurs dans le développement du modèle qui existent sous forme raster où chaque pixel représente une zone forestière d'environ 200 à 325 m<sup>2</sup> (taille pour une parcelle standard). En fonction de sa taille, un peuplement forestier peut être composé de centaines de pixels. Une fois que les modèles sont développés et appliqués, chacun des pixels aura une estimation de Y. Le pixel permet donc de fournir une description détaillée des attributs de la forêt dans le peuplement.

#### **II.1.3.2 L'approche par arbre individuel**

L'idée principale de cette approche est de segmenter les arbres individuels avec précision à partir du nuage de points LiDAR, puis extraire les métriques liées à l'arbre. Les métriques peuvent être n'importe quelles caractéristiques associées à l'arbre telles que le diamètre de la cime ou la hauteur. À partir de ces métriques ou les connaissances liées à la croissance de l'arbre ou d'équations allométriques, il est possible de lier les mesures de l'arbre à d'autres valeurs telles que le volume de bois, la biomasse et le type d'essence (Dalponte et al. 2018). Diverses méthodes ont été explorées pour détecter des arbres individuels dont beaucoup reposent sur le MHC. Il n'existe pas de méthodes standard, mais nous pouvons distinguer deux catégories d'algorithmes :

- Algorithmes basés sur le MHC
- Algorithmes basés sur le nuage de points bruts

Les algorithmes basés sur le MHC segmentent des arbres individuels en fonction d'une image raster du MHC. Ils sont basés sur des algorithmes de traitement d'images classiques qui ne sont pas liés au nuage de points. Le choix d'algorithme influence donc grandement le résultat, et ils existent beaucoup d'entre eux dans la littérature. Les deux familles génériques

d'algorithmes de traitement d'image sont le « watershed » et « region growing » qui ne sont pas spécifiques à un type d'image raster en particulier.

Les algorithmes basés sur le nuage de points brut ont été créés pour éviter le problème de dépendance à l'observation d'échelle. Travailler directement sur le nuage de points permet d'éviter les imprécisions de l'image raster du MHC. Plusieurs algorithmes à base de traitement d'image peuvent également être convertis en une version basée sur un nuage de points. Par exemple, Dalponte et Coomes (2016) utilisent la méthode par « region growing ». L'inconvénient de cette approche est qu'elle nécessite des données suffisamment denses pour détecter avec précision les cimes d'arbres individuelles (Kaartinen et al. 2012). Puis, elle requiert une puissance de calcul assez importante. Cependant, l'avantage est qu'elle peut potentiellement fournir plus d'informations au niveau de l'arbre.

### **II.1.4 Equations allométriques**

Avec ces différentes approches, nous pouvons mesurer la biomasse aérienne à partir d'équations allométriques.

L'allométrie est l'étude des relations entre la taille et la forme des organismes (King, 1996). Plus exactement, c'est l'étude des relations entre les caractéristiques (longueurs, volumes, formes) des différentes structures ou organes d'un organisme. Par exemple, pour une espèce donnée, on peut définir une relation entre la hauteur d'un arbre et le diamètre moyen de son tronc. Dans le cas de mesures LiDAR, la relation permet de relier une grandeur observable (hauteur d'arbre) à une grandeur non observable (la biomasse par exemple).

Cette relation est exprimée soit avec une fonction linéaire ou soit une fonction non linéaire : (Shi et Liu, 2017)

$$y = ax + b + e \text{ (fonction linéaire)}$$

$$y = ax^b + e \text{ ou } \log a + b \log x + e \text{ (fonction non linéaire)}$$

$y$  représente l'attribut de l'arbre (la biomasse par exemple) et  $x$  la taille et la forme de l'arbre.  $a$  et  $b$  sont des coefficients.  $e$  représente l'erreur. On peut définir les matrices suivantes :

$$Y = (n \times 1) \text{ vecteur d'observations}$$

$$X = (n \times p) \text{ matrice de forme connue}$$

Le modèle non linéaire peut-être linéarisé avec une transformation appropriée ou ne peut être linéarisé. Un exemple de modèle non linéaire qui peut être linéarisé avec un modèle de forme de variables multiplicatives :

$$y = ax_1^b + cx_2^d + ex_3^f + e_r \text{ (Forme non linéaire)}$$

Pour effectuer la linéarisation, on utilise une forme logarithmique :

$$\ln y = \ln a + b \ln x_1 + \ln c + d \ln x_2 + \ln e + f \ln x_3 + \ln e_r \text{ (Forme linéaire)}$$

Ainsi, il est possible d'estimer les coefficients à l'aide de la méthode par régression linéaire et moindres carrés.

Un modèle qui ne peut pas être converti en modèle linéaire est de type :

$$y = a x_1^b \times x_2^c \times x_3^d + e_r$$

La méthode par régression linéaire ne s'applique pas sur cette équation. L'application des moindres carrés nécessite donc une approche itérative pour pouvoir estimer les coefficients.

En écologie, la relation allométrique peut être utilisée pour calculer des facteurs écologiques et notamment la biomasse, en mesurant le diamètre à la hauteur de poitrine. Ces mesures peuvent se faire directement sur le terrain. Théoriquement, ce qui affecte la biomasse c'est la hauteur et le diamètre de l'arbre. Par conséquent, la biomasse peut être estimée à l'aide d'équations allométriques :

$$W = aD^b \times H^c + e \text{ ou } \log w = \log a + b \log D + c \log H + e$$

$D$  représente le diamètre de l'arbre à la hauteur de la poitrine (cm) et  $H$  représente la hauteur en (m),  $e$  l'erreur, et  $a$ ,  $b$ , et  $c$  sont des coefficients de régression.

Par contre, il n'est pas facile de relever la hauteur  $H$  sur le terrain contrairement à  $D$ . Des chercheurs ont donc mis en place des modèles prédicteur pour estimer  $H$  au lieu de le mesurer. De plus, certaines équations allométriques ne comprennent que  $D$  pour estimer la biomasse.

Une équation allométrique de la biomasse fonde sa base théorique sur des lois d'échelle allométrique en biologie (West 1997). Ces équations ne prennent pas en compte la biomasse souterraine ainsi que les composants organiques en décomposition (litière, bois mort ...).

Cependant, la biomasse dépend également de caractéristiques spécifiques à la nature et composition de la végétation. On retrouve alors plusieurs types d'équations allométriques :

- Mono-espèces ou spécifiques à un genre, par exemple celles établies par Basuki et al. (2009)

- Multi-espèces locales établies dans un site donné qui incluent le diamètre et/ou la hauteur, et la densité du bois dans le jeu des prédicteurs
- Multi-espèces régionales ou pantropicales avec le diamètre et/ou la hauteur, la densité du bois et le type de forêt (« wet » – « moist » – « dry ») dans le jeu des prédicteurs (Chave et al. 2005)

Les mesures obtenues par télédétection notamment le LiDAR sont injectées dans l'équation allométrique, mais doivent être vérifiées avec les données terrain. L'utilisation d'une équation allométrique est donc indispensable pour estimer la biomasse des arbres et forêts.

Les deux approches basées sur un seul arbre et les approches basées sur les zones impliquent l'utilisation de modèles empiriques ou semi-empiriques en utilisant une analyse de régression linéaire ou non linéaire. Les approches basées sur des arbres individuels nécessitent des données LiDAR avec des densités ponctuelles élevées et sont principalement basées sur des modèles de régression se concentrant sur une relation entre les paramètres individuels de l'arbre, par exemple, la hauteur, la dimension de la cime et des estimations de la biomasse sur le terrain. Les méthodes basées sur l'approche par zone peuvent utiliser des densités de points plus basses, mais nécessitent un grand nombre de données de référence. Ces méthodes analysent la distribution verticale des retours laser au niveau de la parcelle en dérivant des quantités statistiques importantes et estimant les paramètres d'inventaire forestier par zone.

En utilisant l'approche par zone il est donc possible de dériver des métriques des arbres avec le LiDAR, la question est de savoir quelle équation utiliser pour estimer la biomasse.

### ***II.1.5 Équations allométriques établies sur le territoire du Brésil :***

Pour faire un bon choix d'équation allométrique, il faut prendre en compte l'endroit du site d'étude qui a permis l'estimation des coefficients de l'équation et la précision de l'estimation de la biomasse en fonction des résultats obtenus avec les données LiDAR et terrain.

Pour reprendre un modèle utilisé dans la littérature, il faut examiner le type de végétation et de forêt avec lequel les paramètres du modèle ont été estimés. Effectivement, pour pouvoir faire correspondre le modèle à notre zone, il faudrait que le modèle du site d'étude utilisé soit à peu près similaire à celui qu'on veut estimer. Cette méthode d'approximation n'est peut-être pas rigoureuse, mais nous supposons qu'elle est valide quand nous rencontrons le même type de forêt.

Dans cette étude, nous avons choisi une équation qui était formée pour la forêt atlantique du Brésil, différente de la forêt de l'Amazonie.

Mis à part le site d'étude choisi, la précision de l'estimation de la biomasse est importante également. Même si l'objectif de notre étude n'est pas d'obtenir une précision optimale, nous avons décidé de choisir une précision correcte. Pour évaluer la précision de l'estimation de la biomasse, il existe principalement l'erreur quadratique moyenne (RMSE : Root Mean Square Error) qui est l'écart type des résidus, considéré comme l'erreur de prédiction. Les résidus sont en fait une mesure de la distance qui sépare les points de données de la droite de régression. Il indique ainsi la concentration des données autour de la droite de meilleur ajustement, il permet donc de vérifier les résultats expérimentaux. Il est noté :

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

Où  $n$  est le nombre de parcelles,  $y_i$  la valeur observée pour  $i$  parcelle et  $\hat{y}_i$  est la valeur prédite pour la parcelle  $i$ .

Le coefficient de détermination  $R$  est aussi un indicateur de qualité de régression,  $R^2$  est conçu pour un modèle de régression linéaire. C'est le quotient des variances des valeurs prédites des valeurs observées d'une variable dépendante. Il est exprimé en pourcentage donc toujours compris entre 0 % et 100 %. Il mesure donc la correspondance entre le modèle et les données observées.

- 0 % indique que le modèle utilisé n'explique absolument pas la distribution des points.
- 100 % indiquent le modèle utilisé, sachant que les paramètres ou coefficients calculés sont ceux qui déterminent la distribution des points.

Il est exprimé de la façon suivante :

$$r^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

Où  $y_i$  est la valeur observée pour chaque parcelle  $i$  et  $\hat{y}_i$  est la valeur prédite pour la parcelle  $i$ , et  $\bar{y}$  est la moyenne des valeurs observées.

Il est donc aussi un bon indicateur de corrélation entre la mesure de la biomasse et des métriques obtenues à partir de données LiDAR.

En réunissant tous ces éléments, il a fallu faire un choix pour identifier l'équation à utiliser. De plus, les équations doivent estimer pour un ensemble d'espèces d'arbres de type tropical et donc ne présente que les statistiques autour de la hauteur comme les percentiles de hauteur par exemple, car l'approche par zone ne donne pas l'estimation du Dbh (diamètre d'arbre à hauteur de poitrine).

## **II.1.6 Exemple de deux équations appliquées en Amazonie**

### **II.1.6.1 : Article Remote Sensing de (Sato et al. 2016)**

Sato et al. appliquent une équation non linéaire de la forme :  $y = ax_1^b \times cx_2^d \times ex_3^f \times gx_4^h$ , dans la partie Est de l'état d'Acre situé dans la forêt amazonienne.

$$AGB^{10} = 0.36u_z^{1.16} \times z_{75}^{0.78} \times z_{10}^{-0.18} \times k_z^{0.41} \quad (13)$$

$u_z$  est la hauteur moyenne,  $z_{75}$  est le 75<sup>ème</sup> percentile,  $z_{10}$  est le 10<sup>ème</sup> percentile,  $k_z$  est le kurtosis des distributions de tous les retours de chaque hauteur.

### **II.1.6.2 Article de Rappaport et al. 2018**

Dans le contexte de la déforestation et des incendies forestiers, Rappaport et al, ont développé une équation allométrique fonction de la hauteur maximale, à l'aide de données inventaire terrain et de LiDAR provenant des forêts amazoniennes intactes et dégradées. La zone d'étude se localise à l'étendue sud des forêts amazoniennes dans l'État brésilien du Mato Grosso.

$$AGB = 0.054 \times h_{max}^{1.76} \quad (14)$$

Où  $h_{max}$  désigne la hauteur maximale.

Ces deux équations construites à partir de données LiDAR et d'inventaire terrain se situent au Brésil, mais ne sont sans doute pas applicables dans la forêt atlantique du Brésil. Nous allons donc présenter deux équations applicables à la forêt atlantique du Brésil.

---

<sup>10</sup> AGB : Above-Ground Biomass

## **II.1.7 Exemple de deux équations appliquées en forêt atlantique du Brésil**

### **II.1.7.1 Article (Leitold et al. 2015) de Carbon Balance and Management**

Leitold et al. (2015) ont développé un modèle de régression linéaire simple à partir des neuf parcelles d'inventaire terrain et des données LiDAR au même site d'étude que notre projet.

$$AGB = 24.13 \times h_{moy} - 204.76 \quad (15)$$

Où  $h_{moy}$  est la hauteur moyenne de la canopée.  $AGB$  est prédit en Mg / ha.  $R^2 = 0.43$  et  $RMSE = 30.0$  Mg / ha.

### **II.1.7.2 Article (Becknell et al. 2018) de Biotropica**

Becknell et al. 2018 ont choisi un modèle sigmoïdal avec une seule variable : 95<sup>ème</sup> percentile de hauteur. Ils ont comparé les modèles linéaires et sigmoïdaux, mais les modèles linéaires prédisaient une biomasse négative, ils ont donc sélectionné le modèle avec le plus petit RMSE à partir des 30 parcelles d'inventaire terrain et des mesures LiDAR.

$$AGB = \frac{590.2}{1 + e^{-0.202(P95-23.24)}} \quad (16)$$

Où  $P95$  est le 95<sup>ème</sup> percentile de hauteur.  $AGB$  est prédit en Mg / ha.  $R^2 = 0.62$  et  $RMSE = 44,85$  Mg / ha.

## **II.2 Matériels et méthodes :**

### **II.2.1 Site d'étude et données**

Le site d'étude et les données sont les mêmes que cité dans la première partie.

### **II.2.2 Pré-traitement**

#### **II.2.2.1 Échantillonnage des points**

Les données ont été sous-échantillonnées de façon aléatoire (8 points / m<sup>2</sup>, 3 points / m<sup>2</sup>, 1 points / m<sup>2</sup>). La proportion de points est conservée en fonction de la densité de points que l'on souhaite obtenir. En revanche, la fonction utilisée « lasfilterdecimate » ne peut pas augmenter le nombre de points, car si ce nombre que l'on souhaite est plus élevé que le nombre de points existant alors il renvoie un ensemble de points inchangé. Cette fonction

peut aussi supprimer des impulsions aléatoires au lieu de points aléatoires. Trois algorithmes peuvent être appliqués :

- L'algorithme « random » qui élimine aléatoirement les points ou les impulsions pour atteindre la densité souhaitée
- L'algorithme « homogenize » crée une grille avec une résolution définie et filtre le nuage de points en sélectionnant au hasard les points dans chaque cellule. Il est conçu pour produire des nuages de points de densités homogènes dans toute la zone calculée. La taille de cellule doit être adaptée pour la densité existante. Par exemple dans un nuage de 2 points / m<sup>2</sup>, la résolution de 5 x 5 m est préférable pour calculer une densité de points cohérente.
- L'algorithme « highest » crée une grille avec une résolution définie et filtre le nuage de points en sélectionnant le point le plus élevé de chaque cellule

Nous avons opté pour l'algorithme « homogenize » pour produire des nuages de points de densités homogènes sur toute la zone avec une résolution de 1 m<sup>2</sup>.<sup>11</sup>

### **II.2.3 Traitement**

#### **II.2.3.1 Application de l'équation allométrique choisie**

Après réflexion, nous avons choisi l'équation ( 16 ) de Becknell et al. (2018). Tout d'abord, parce qu'il concernait le même type de forêt que notre zone d'étude, puis parce qu'il présentait un coefficient de détermination R<sup>2</sup> satisfaisant (R<sup>2</sup>=0.62) avec un RMSE plutôt petit. De plus, l'équation ( 15 ) de l'article de Leitold et al. 2015 est de forme linéaire avec un terme négatif. Ainsi, quand la valeur de la hauteur moyenne de certains pixels était environ inférieure à 8.50 m, nous avons une valeur de biomasse négative. Il était donc préférable d'utiliser l'équation ( 16 ) de Becknell et al. 2018, qui utilise une équation sigmoïdale avec des valeurs prédites de biomasse positive. Il faut préciser que Leitold et al. 2015., a calculé la biomasse sur les neuf parcelles d'inventaire forestier où certainement il n'y a que des arbres. Cependant, notre objectif d'étude est d'estimer la biomasse sur tout le territoire de notre zone d'étude.

Le traitement de l'application de cette équation a été fait sur le logiciel R.

---

<sup>11</sup> Voir annexe 3

Nous avons tout d'abord sous-échantillonné notre jeu de données avec la méthode décrite dans la partie précédente (Échantillonnage des points) à 1, 3, 7 points / m<sup>2</sup>.

En utilisant l'approche par zone, nous avons calculé ensuite une série statistique descriptive basée sur les 95<sup>ème</sup> percentiles au sein de chaque pixel avec une résolution de 1x1 m<sup>2</sup>. Enfin, nous avons injecté ces valeurs de percentiles de hauteur dans l'équation ( 16 ) pour en déduire les valeurs de biomasse.

## II.3 Résultats

Après avoir appliqué l'équation allométrique ( 16 ) sur toute la zone d'étude sur différents jeux de données échantillonnés à 1, 3, 7, 18 points / m<sup>2</sup>, nous obtenons une cartographie de la biomasse sur toute la zone.

### II.3.1 Estimation de la biomasse avec des données à faible densité : 1, 3 points / m<sup>2</sup> <sup>12</sup>

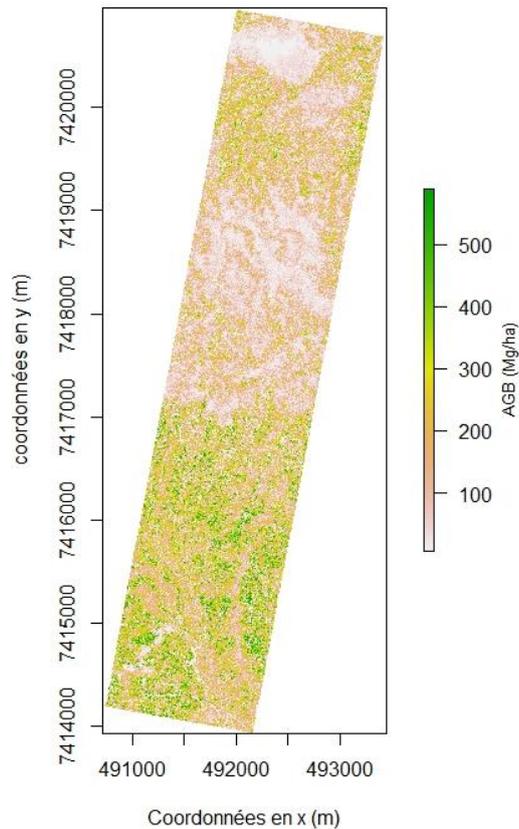


Figure 18: cartographie de la biomasse aérienne à partir des données LiDAR de densité de 1 point / m<sup>2</sup>

---

<sup>12</sup> Voir annexe 4

**II.3.2 Estimation de la biomasse avec des données à moyenne densité :  
7 points / m<sup>2</sup><sup>13</sup>**

**II.3.3 Estimation de la biomasse avec des données à haute densité : 18  
points / m<sup>2</sup>**

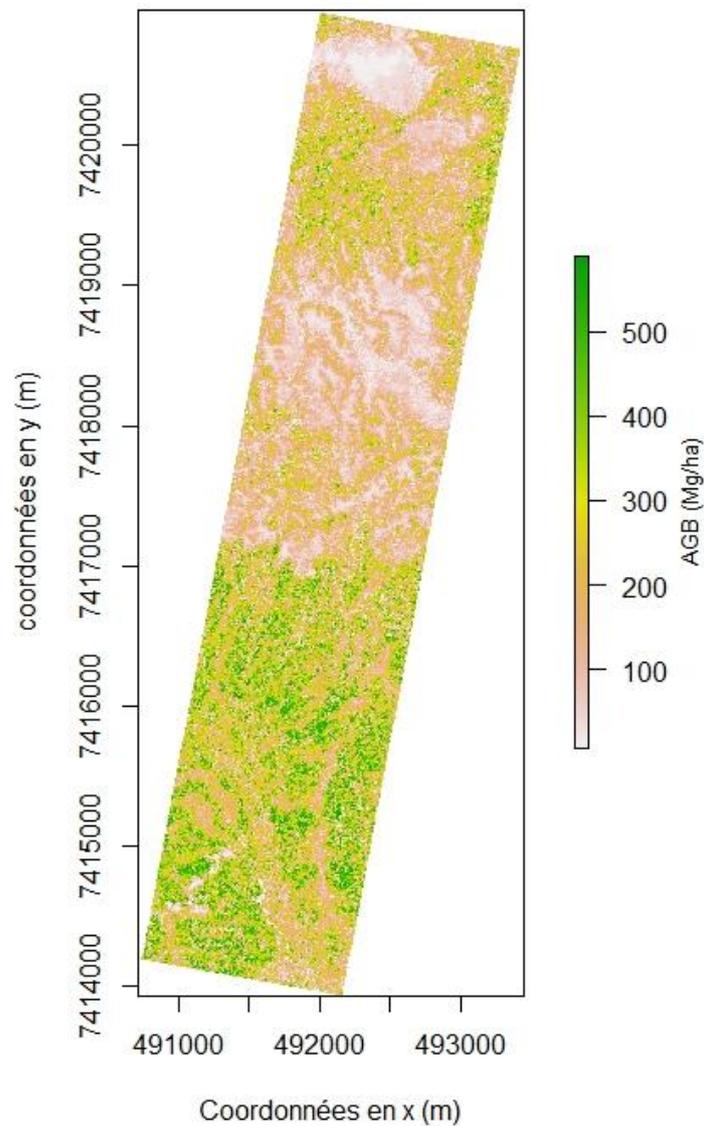


Figure 19: cartographie de la biomasse forestière à partir de densité de 18 points / m<sup>2</sup>

---

<sup>13</sup> Voir annexe 5

### II.3.4 Comparaison de l'estimation de la biomasse selon les différentes données échantillonnées

Evolution de la biomasse en fonction de plusieurs densités de données lidar

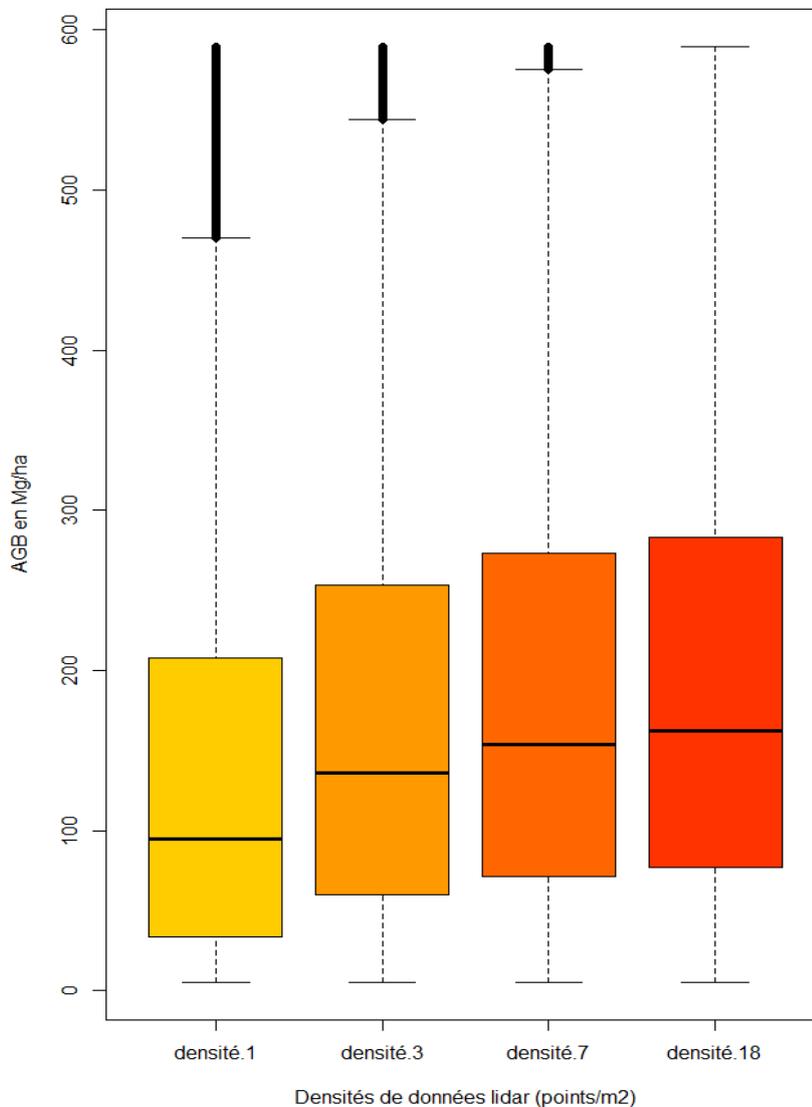


Figure 20: répartition de la biomasse aérienne sur toute la zone d'étude pour différentes densités de points issues du même jeu de données avec un modèle non linéaire basé sur les percentiles de hauteur. Densité 1, 3, 7, 18 correspondent aux données de biomasse issues de différentes densités de données LiDAR, par exemple densité 1 représente les données de biomasse issues de densité de 1 point / m<sup>2</sup>.

La Figure 20 nous montre les écarts de valeurs de biomasse pour différentes densités de points, effectivement nous remarquons un écart moyen d'environ 50 Mg / ha entre le jeu de données de densité 1 point / m<sup>2</sup> et le jeu de donnée de densité 18 points / m<sup>2</sup>. Nous observons

que la plus grande différence d'écart moyen est entre le jeu de données de densité 1 point / m<sup>2</sup> et 3 points / m<sup>2</sup>. L'écart moyen de biomasse aérienne diminue de plus en plus après 3 points / m<sup>2</sup>, il n'est plus très important après 7 points / m<sup>2</sup> et 18 points / m<sup>2</sup>, de l'ordre de 10 Mg / ha. Ce qui représente une faible sous-estimation de biomasse aérienne. Globalement, nous avons une évolution d'écart moyen qui diminue au fur à mesure que la densité de points / m<sup>2</sup> augmente.

## II.4 Discussion

Nous pouvons observer l'effet de l'estimation de la biomasse sur plusieurs jeux de données à densités de points différents, nous remarquons avec la Figure 20 une sous-estimation évidente des valeurs de biomasse en fonction de ces densités différentes. La densité 1 est celle qui sous-estime le plus les valeurs de biomasse, ainsi que cette sous-estimation diminue davantage au fur à mesure que la densité de points augmente. Pourquoi nous observons cet effet ?

Tout d'abord, il faut rappeler que l'équation utilisée est liée directement à une variable de hauteur. Par conséquent, les valeurs de hauteur dans chaque pixel vont influencer les valeurs de biomasse pour chaque pixel. L'estimation de la hauteur est donc liée à l'estimation de la biomasse. Or, le modèle de Roussel et al. (2019) indique une sous-estimation de la hauteur à partir de 1 point par m<sup>2</sup> et qui diminue de plus en plus. Par comparaison, nous retrouvons cet effet également sur les résultats de biomasse. Peut-on donc affirmer que son modèle est applicable sur notre site d'étude ? Nous avons choisi expressément ces densités différentes : 1 point / m<sup>2</sup>, 3 points / m<sup>2</sup>, 7 points / m<sup>2</sup>, 18 points / m<sup>2</sup>. 1 point / m<sup>2</sup> n'est pas couramment utilisé dans le domaine de la télédétection par LiDAR aéroporté, qui représente un niveau bas de qualité de donnée. La densité de 3 points / m<sup>2</sup> est la plus utilisée, et est un niveau normal dans ce domaine. La densité de 7 points / m<sup>2</sup> représente un niveau élevé et est plus rarement utilisée, qui aura en conséquence un coût important d'acquisition de données. La densité de 18 points / m<sup>2</sup> est considérée comme des données à haute densité présentant un niveau d'informations riches et est très rare dans le domaine de la télédétection. Grâce à l'obtention de ces résultats, il suffirait donc d'acquérir ces données à 3 points / m<sup>2</sup> dans le but d'estimer la biomasse sans avoir besoin nécessairement de corriger l'erreur moyenne sur l'estimation de la hauteur.

Néanmoins dans notre cas, nous avons repris une équation liée à une seule variable qui est la hauteur. Dans la littérature, pour obtenir des estimations de biomasse plus précises,

certaines études utilisent plusieurs variables pour réduire les erreurs d'estimation de biomasse au maximum. Par exemple, une étude en Indonésie montre que l'estimation de la biomasse est plus précise en utilisant d'autres variables comme la surface terrière et la densité du bois. (Coomes et al. 2017). Cependant, l'objectif de notre étude n'est pas d'obtenir une estimation de biomasse précise par rapport à l'estimation de biomasse par relevé terrain, mais de montrer l'effet de l'estimation de la hauteur sur la biomasse.

En revanche, il serait intéressant d'étudier à titre de perspective la dépendance d'autres métriques sur l'estimation de la biomasse. Le LiDAR ne donne pas directement les valeurs de biomasse, il permet d'obtenir la distribution géométrique d'un certain nombre d'objet et dont la plupart sont des feuilles. La hauteur n'est donc pas forcément la métrique la meilleure pour l'estimation de la biomasse. Asner et Mascaro (2014) proposent une méthodologie universelle de cartographie du stock aérien de carbone par LiDAR en forêt tropicale. Celle-ci repose sur la corrélation entre la hauteur de peuplement, la surface terrière et la densité du bois. La surface terrière représente la surface qu'occupe l'arbre à 1,30 m de hauteur, elle résume la densité et le diamètre d'un peuplement. En général, nous rencontrons ces variables-ci dans la littérature pour l'estimation de la biomasse aérienne par LiDAR. Mais il pourrait y avoir d'autres métriques que celles liées à la hauteur du LiDAR aérien pour calculer la biomasse aérienne. Peu de métriques liées aux structures forestières sont exploitées comme la densité de la canopée, le volume de la canopée, le pourcentage des retours de la végétation. Elles s'avèrent utiles pour prédire la biomasse. En effet, les métriques sur la couverture de la canopée et la densité peuvent compléter les métriques de hauteur pour une meilleure caractérisation de la structure forestière, car elles sont liées aux structures de végétation horizontale et verticale.

Il serait aussi également intéressant d'étudier l'effet de l'angle de balayage sur la pente car notre site d'étude révèle un terrain avec des pentes abruptes. Au fur à mesure que l'angle de balayage s'éloigne du nadir, l'angle d'incidence au sol augmente. La cause de ce phénomène est que le sol se comporte moins comme un diffuseur et ressemble plus à un réflecteur directionnel, facilitant la réflectance du chemin. La distance parcourue par le faisceau laser augmente également. Dans le cas d'acquisition en pente, la distance parcourue par l'impulsion et l'angle d'incidence au sol peuvent encore augmenter. Par conséquent, le taux de pénétration de l'impulsion laser dans la canopée est directement lié aux angles de balayage. Les pentes les plus abruptes devraient produire des données avec un minimum d'artefacts liés à l'angle de balayage (Gatziolis et Andersen (2008)). En revanche, trois

erreurs peuvent affecter la précision du laser sur la pente : le déplacement horizontal d'un point balayé au sol, les distorsions liées à l'empreinte laser au sol et à l'impulsion laser. (Cristi Gherase (2009)). L'angle d'incidence aurait également un impact sur la précision du laser sur la topographie du sol mais est peu étudié. Roussel et al. (2018) a montré l'effet de l'angle incidence du laser sur les métriques liées à la hauteur de la canopée mais non sur la topographie du sol. La grande majorité des études ont porté sur les effets de la végétation sur les mesures, car la présence de végétation est la principale cause d'erreur et la cartographie de la végétation est elle-même une application importante de la technologie du LiDAR. La structure de la forêt influencerait donc l'angle de visée et donc l'angle d'incidence sur la végétation. Par exemple, dans le cas de plantations alignées la réponse de l'impulsion laser ne serait pas la même selon le sens de rangées des arbres. Une forêt isotrope simplifierait l'étude de l'angle de visée comme c'est le cas de notre site d'étude, car elle n'a pas de direction privilégiée, peu importe la direction de vol. Il serait donc intéressant d'analyser l'influence de l'angle de balayage sur la végétation et la topographie du sol dans notre cas pour une autre étude. L'angle de balayage a été cartographié sur l'ensemble avec les données de 2017<sup>14</sup>, car les données de 2012 ne contenaient pas les valeurs de l'angle de balayage. Le modèle du système LiDAR des données 2017 est le même, en revanche les paramètres d'acquisition sont différents avec un champ de vision du système de balayage de 15° au lieu de 11° provenant des données 2012.

## **Conclusion :**

L'objectif de cette étude consistait à vérifier si le modèle de Roussel et al (2017) s'appliquait sur d'autres jeux de données telles que des données acquises en forêt tropicale. Nous avons vu que le biais moyen de la hauteur maximale est nettement moins important en forêt tropicale qu'en forêt d'Amérique du Nord. Pour vérifier son modèle, nous avons observé l'effet de l'estimation de la hauteur sur la biomasse selon différentes densités de points, et il est évident qu'on remarque une sous-estimation de la biomasse qui augmente au fur à mesure que la densité diminue. Mais la sous-estimation en moyenne de la biomasse n'est pas si élevée, elle est d'environ 50 Mg / ha. En revanche, lorsque la densité de points atteint 7 points / m<sup>2</sup>, l'erreur diminue fortement, la différence entre la densité de 18 points / m<sup>2</sup> et de

---

<sup>14</sup> Voir annexe 7

7 points / m<sup>2</sup> est quasiment négligeable, elle est nettement inférieure à 30 Mg / ha. Ce qui permet de répondre à la question posée lors de l'introduction, y a-t-il un biais important dans l'estimation de la hauteur sur le calcul de la biomasse à faible densité ? Cette étude montre qu'il n'y a pas de biais important dans l'estimation de la biomasse donc de l'estimation de la hauteur sur des données à faible densité.

Ces résultats confirment donc les résultats prédits du modèle de Roussel et al (2017). Son application est donc vérifiée à travers cette étude. Effectivement, le biais est peu important pour qu'il soit corrigé. Cette étude démontre qu'il n'est pas forcément nécessaire d'acquérir des données de haute densité sur la forêt atlantique du Brésil pour une étude de biomasse basée sur l'estimation de la hauteur. L'estimation de la hauteur pour ce type de forêt ne va pas beaucoup influencer la précision de l'estimation de la biomasse à faible densité, cependant la précision dépend-elle d'autres paramètres d'acquisition de données LiDAR ?

## Bibliographie

- Alexander, Cici, Amanda H. Korstjens, et Ross A. Hill. 2018. « Influence of Micro-Topography and Crown Characteristics on Tree Height Estimations in Tropical Forests Based on LiDAR Canopy Height Models ». *International Journal of Applied Earth Observation and Geoinformation* 65: 105-13. <https://linkinghub.elsevier.com/retrieve/pii/S0303243417302283> (28 juin 2019).
- Asner, Gregory P., et Joseph Mascaro. 2014. « Mapping Tropical Forest Carbon: Calibrating Plot Estimates to a Simple LiDAR Metric ». *Remote Sensing of Environment* 140: 614-24. <https://linkinghub.elsevier.com/retrieve/pii/S003442571300360X> (28 juin 2019).
- Basuki, T.M., P.E. van Laake, A.K. Skidmore, et Y.A. Hussin. 2009. « Allometric Equations for Estimating the Above-Ground Biomass in Tropical Lowland Dipterocarp Forests ». *Forest Ecology and Management* 257(8): 1684-94. <https://linkinghub.elsevier.com/retrieve/pii/S0378112709000516> (28 juin 2019).
- Becknell, Justin M. et al. 2018. « Landscape-Scale Lidar Analysis of Aboveground Biomass Distribution in Secondary Brazilian Atlantic Forest ». *Biotropica* 50(3): 520-30. <http://doi.wiley.com/10.1111/btp.12538> (28 juin 2019).
- Ben-Arie, Joshua R. et al. 2009. « Development of a Pit Filling Algorithm for LiDAR Canopy Height Models ». *Computers & Geosciences* 35(9): 1940-49. <https://linkinghub.elsevier.com/retrieve/pii/S0098300409000624> (28 juin 2019).
- Chave, J. et al. 2005. « Tree Allometry and Improved Estimation of Carbon Stocks and Balance in Tropical Forests ». *Oecologia* 145(1): 87-99. <https://www.jstor.org/stable/20062389> (28 juin 2019).
- Chen, Ziyue, Bingbo Gao, et Bernard Devereux. 2017. « State-of-the-Art: DTM Generation Using Airborne LIDAR Data ». *Sensors (Basel, Switzerland)* 17(1). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5298723/> (28 juin 2019).
- Coomes, David A. et al. 2017. « Area-Based vs Tree-Centric Approaches to Mapping Forest Carbon in Southeast Asian Forests from Airborne Laser Scanning Data ». *Remote Sensing of Environment* 194: 77-88. <https://linkinghub.elsevier.com/retrieve/pii/S0034425717301098> (28 juin 2019).
- Dalponte, Michele et al. 2018. « Predicting stem diameters and aboveground biomass of individual trees using remote sensing data ». *Ecological Indicators* 85: 367-76. <http://www.sciencedirect.com/science/article/pii/S1470160X17307033> (28 juin 2019).
- Dalponte, Michele, et David A. Coomes. 2016. « Tree-Centric Mapping of Forest Carbon Density from Airborne Laser Scanning and Hyperspectral Data ». *Methods in Ecology and Evolution* 7(10): 1236-45. <http://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12575> (28 juin 2019).

- Gatziolis, Demetrios, et Hans-Erik. Andersen. 2008. *A Guide to LIDAR Data Acquisition and Processing for the Forests of the Pacific Northwest*. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. <https://www.fs.usda.gov/treearch/pubs/30652> (8 août 2019).
- Gaveau, David L.A., et Ross A Hill. 2003. « Quantifying Canopy Height Underestimation by Laser Pulse Penetration in Small-Footprint Airborne Laser Scanning Data ». *Canadian Journal of Remote Sensing* 29(5): 650-57. <http://www.tandfonline.com/doi/abs/10.5589/m03-023> (28 juin 2019).
- Gherase, Cristi. 2009. « The effect of laser beam incidence angle and of terrain slope on the vertical accuracy ». Van Amsterdam. [https://www.google.com/search?rlz=1C1CHBF\\_frFR770FR770&ei=AiIMXbzDNI Xj\\_Aam0oTYBg&q=the+effect+of+laser+beam+incidence+angle+and+of+terrain+slope+on+the+vertical+accuracy&oq=the+effect+of+laser+beam+incidence+angle+and+of+terrain+slope+on+the+vertical+accuracy&gs\\_l=psy-ab.3...6126.12991..13304...0.0..0.209.2342.22j2j1.....0....1..gws-wiz.....33i10.Z1Jxfuf8Sn0&ved=&uact=5](https://www.google.com/search?rlz=1C1CHBF_frFR770FR770&ei=AiIMXbzDNI Xj_Aam0oTYBg&q=the+effect+of+laser+beam+incidence+angle+and+of+terrain+slope+on+the+vertical+accuracy&oq=the+effect+of+laser+beam+incidence+angle+and+of+terrain+slope+on+the+vertical+accuracy&gs_l=psy-ab.3...6126.12991..13304...0.0..0.209.2342.22j2j1.....0....1..gws-wiz.....33i10.Z1Jxfuf8Sn0&ved=&uact=5) (8 août 2019).
- Gobakken, Terje, et Erik Næsset. 2008. « Assessing Effects of Laser Point Density, Ground Sampling Intensity, and Field Sample Plot Size on Biophysical Stand Properties Derived from Airborne Laser Scanner Data ». *Canadian Journal of Forest Research* 38(5): 1095-1109. <http://www.nrcresearchpress.com/doi/10.1139/X07-219> (28 juin 2019).
- Hirata, Yasumasa. 2004. « The effects of footprint size and sampling density in Airborne Laser Scanning to extract individual trees in mountainous terrain ». *Laser-Scanners for Forest and Landscape Assessment Volume XXXVI-8/W2* 36. (8 août 2019)
- Höfle, Bernhard, et Norbert Pfeifer. 2007. « Correction of laser scanning intensity data: Data and model-driven approaches ». *ISPRS Journal of Photogrammetry and Remote Sensing* 62(6): 415-33. <http://www.sciencedirect.com/science/article/pii/S0924271607000603> (28 juin 2019).
- Isenburg, Martin, Yuanxin Liu, Jonathan Shewchuk, et Jack Snoeyink. 2006. « Streaming Computation of Delaunay Triangulations ». : 8. <https://www.cs.unc.edu/~isenburg/papers/ilss-scdt-06.pdf>.
- Jakubowski, Marek K., Qinghua Guo, et Maggi Kelly. 2013. « Tradeoffs between Lidar Pulse Density and Forest Measurement Accuracy ». *Remote Sensing of Environment* 130: 245-53. <https://linkinghub.elsevier.com/retrieve/pii/S0034425712004567> (28 juin 2019).
- Kaartinen, Harri et al. 2012. « An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning ». *Remote Sensing* 4(4): 950-74. <http://www.mdpi.com/2072-4292/4/4/950> (28 juin 2019).
- Khosravipour, Anahita et al. 2014. « Generating Pit-Free Canopy Height Models from Airborne Lidar ». *Photogrammetric Engineering & Remote Sensing* 80(9): 863-72.

- <http://openurl.ingenta.com/content/xref?genre=article&issn=0099-1112&volume=80&issue=9&spage=863> (28 juin 2019).
- Khosravipour et al. 2015. « Effect of Slope on Treetop Detection Using a LiDAR Canopy Height Model ». *ISPRS Journal of Photogrammetry and Remote Sensing* 104: 44-52. <https://linkinghub.elsevier.com/retrieve/pii/S0924271615000581> (28 juin 2019).
- King, David A. 1996. « Allometry and Life History of Tropical Trees ». *Journal of Tropical Ecology* 12(1): 25-44. <https://www.jstor.org/stable/2560163> (28 juin 2019).
- Leitold, Veronika et al. 2015. « Airborne Lidar-Based Estimates of Tropical Forest Structure in Complex Terrain: Opportunities and Trade-Offs for REDD+ ». *Carbon Balance and Management* 10(1). <https://cbmjournal.biomedcentral.com/articles/10.1186/s13021-015-0013-x> (14 mars 2019).
- Nasset, E. 2004. « Effects of Different Flying Altitudes on Biophysical Stand Properties Estimated from Canopy Height and Density Measured with a Small-Footprint Airborne Scanning Laser ». *Remote Sensing of Environment* 91(2): 243-55. <https://linkinghub.elsevier.com/retrieve/pii/S0034425704000963> (28 juin 2019).
- Rappaport, Danielle I et al. 2018. « Quantifying Long-Term Changes in Carbon Stocks and Forest Structure from Amazon Forest Degradation ». *Environmental Research Letters* 13(6): 065013. <http://stacks.iop.org/1748-9326/13/i=6/a=065013?key=crossref.f1492542a5956e3bb67b5b5f65707411> (28 juin 2019).
- Rosette, Jacqueline et al. 2012. « Lidar Remote Sensing for Biomass Assessment ». In *Remote Sensing of Biomass - Principles and Applications*, éd. Lola Fatoyinbo. InTech. <http://www.intechopen.com/books/remote-sensing-of-biomass-principles-and-applications/lidar-remote-sensing-for-biomass-assessment> (28 juin 2019).
- Roussel et al. 2019. *R package for airborne LiDAR data manipulation and visualisation for forestry application: Jean-Romain/lidR*. . R. <https://github.com/Jean-Romain/lidR> (28 juin 2019).
- Roussel, Jean-Romain et al. 2017. « Removing Bias from LiDAR-Based Estimates of Canopy Height: Accounting for the Effects of Pulse Density and Footprint Size ». *Remote Sensing of Environment* 198: 1-16. <https://linkinghub.elsevier.com/retrieve/pii/S0034425717302316> (14 mars 2019).
- Roussel, Jean-Romain, Martin Béland, John Caspersen, et Alexis Achim. 2018. « A Mathematical Framework to Describe the Effect of Beam Incidence Angle on Metrics Derived from Airborne LiDAR: The Case of Forest Canopies Approaching Turbid Medium Behaviour ». *Remote Sensing of Environment* 209: 824-34. <https://linkinghub.elsevier.com/retrieve/pii/S0034425717305849> (14 mars 2019).
- Ruiz, Luis, Txomin Hermosilla, Francisco Mauro, et Miguel Godino. 2014. « Analysis of the Influence of Plot Size and LiDAR Density on Forest Structure Attribute Estimates ». *Forests* 5(5): 936-51. <http://www.mdpi.com/1999-4907/5/5/936> (28 juin 2019).

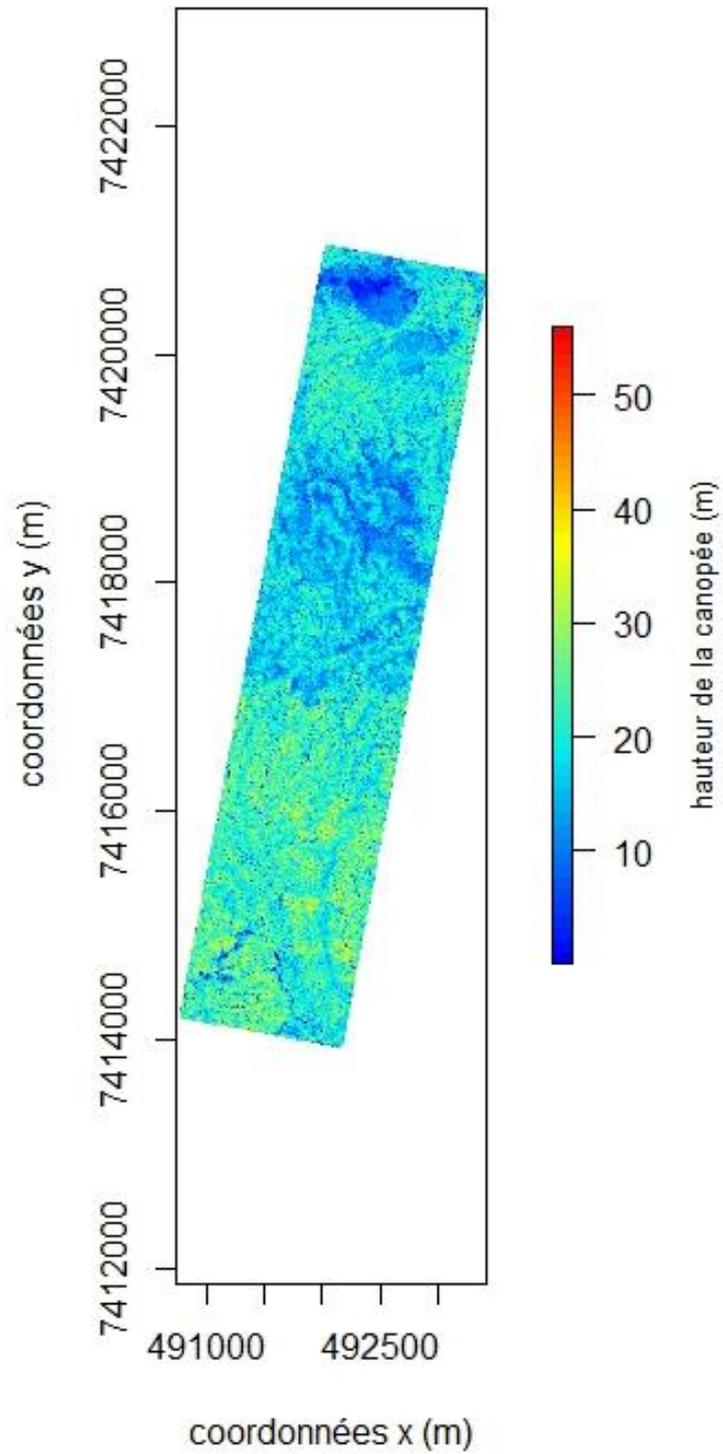
- Sato, Luciane et al. 2016. « Post-Fire Changes in Forest Biomass Retrieved by Airborne LiDAR in Amazonia ». *Remote Sensing* 8(10): 839. <http://www.mdpi.com/2072-4292/8/10/839> (28 juin 2019).
- Shi, Lei, et Shirong Liu. 2017. « Methods of Estimating Forest Biomass: A Review ». In *Biomass Volume Estimation and Valorization for Energy*, éd. Jaya Shankar Tumuluru. InTech. <http://www.intechopen.com/books/biomass-volume-estimation-and-valorization-for-energy/methods-of-estimating-forest-biomass-a-review> (28 juin 2019).
- Steininger et al., M.K. 2000. « Satellite estimation of tropical secondary forest above-ground biomass: Data from Brazil and Bolivia: International Journal of Remote Sensing: Vol 21, No 6-7 ». <https://www.tandfonline-com.acces.bibl.ulaval.ca/doi/abs/10.1080/014311600210119> (28 juin 2019).
- Vega, C. et al. 2014. « PTrees: A Point-Based Approach to Forest Tree Extraction from Lidar Data ». *International Journal of Applied Earth Observation and Geoinformation* 33: 98-108. <https://linkinghub.elsevier.com/retrieve/pii/S0303243414001135> (28 juin 2019).
- Véga, Cédric, Jean-Pierre Renaud, Sylvie Durrieu, et Marc Bouvier. 2016. « On the Interest of Penetration Depth, Canopy Area and Volume Metrics to Improve Lidar-Based Models of Forest Parameters ». *Remote Sensing of Environment* 175: 32-42. <https://linkinghub.elsevier.com/retrieve/pii/S0034425715302613> (28 juin 2019).
- West, G. B. 1997. « A General Model for the Origin of Allometric Scaling Laws in Biology ». *Science* 276(5309): 122-26. <http://www.sciencemag.org/cgi/doi/10.1126/science.276.5309.122> (28 juin 2019).
- White, Joanne C. et al. 2013. « A Best Practices Guide for Generating Forest Inventory Attributes from Airborne Laser Scanning Data Using an Area-Based Approach ». *The Forestry Chronicle* 89(06): 722-23. <http://pubs.cif-ifc.org/doi/10.5558/tfc2013-132> (28 juin 2019).

## Table des annexes

Annexe 1 MHC : Modèle de Hauteur de Canopée de la zone d'étude au Brésil à Serra do Mar.....	59
Annexe 2 Échantillon de données échantillonnées à 1 point / m <sup>2</sup> à gauche et à 18 points / m <sup>2</sup> à droite.....	60
Annexe 3 Cartographie de la biomasse aérienne à partir des données LiDAR de densité de 3 points / m <sup>2</sup> .....	61
Annexe 4 Cartographie de la biomasse aérienne à partir des données LiDAR de densité de 7 points / m <sup>2</sup> .....	62
Annexe 5 Comparaison de l'estimation des 95 èmes percentiles de hauteur selon différentes densités de données LiDAR .....	63
Annexe 6 Cartographie de l'angle de balayage du système LiDAR aérien à partir des données originales 2017 .....	64

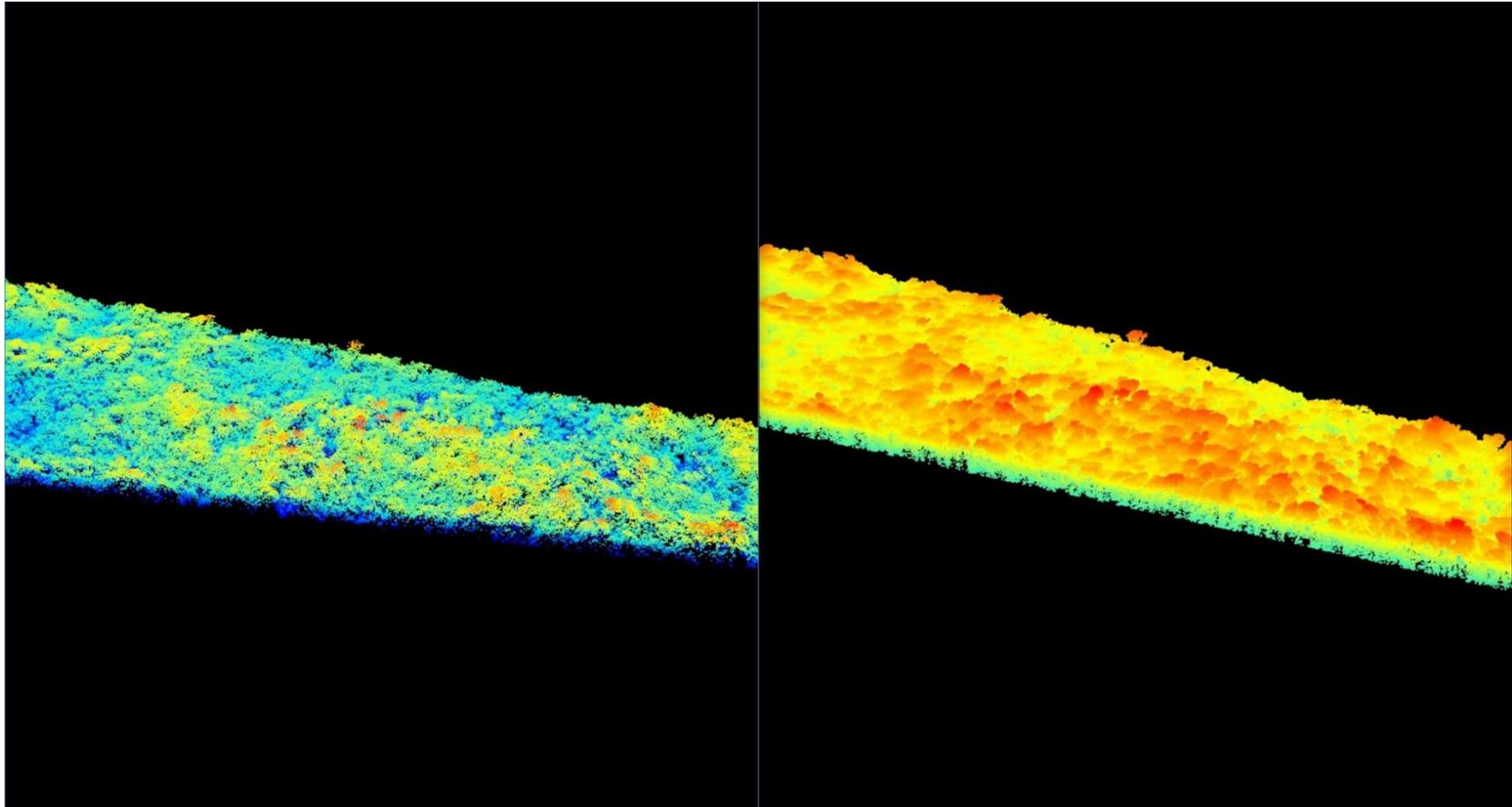
## Annexe 1

### MHC : Modèle de Hauteur de Canopée de la zone d'étude au Brésil à Serra do Mar



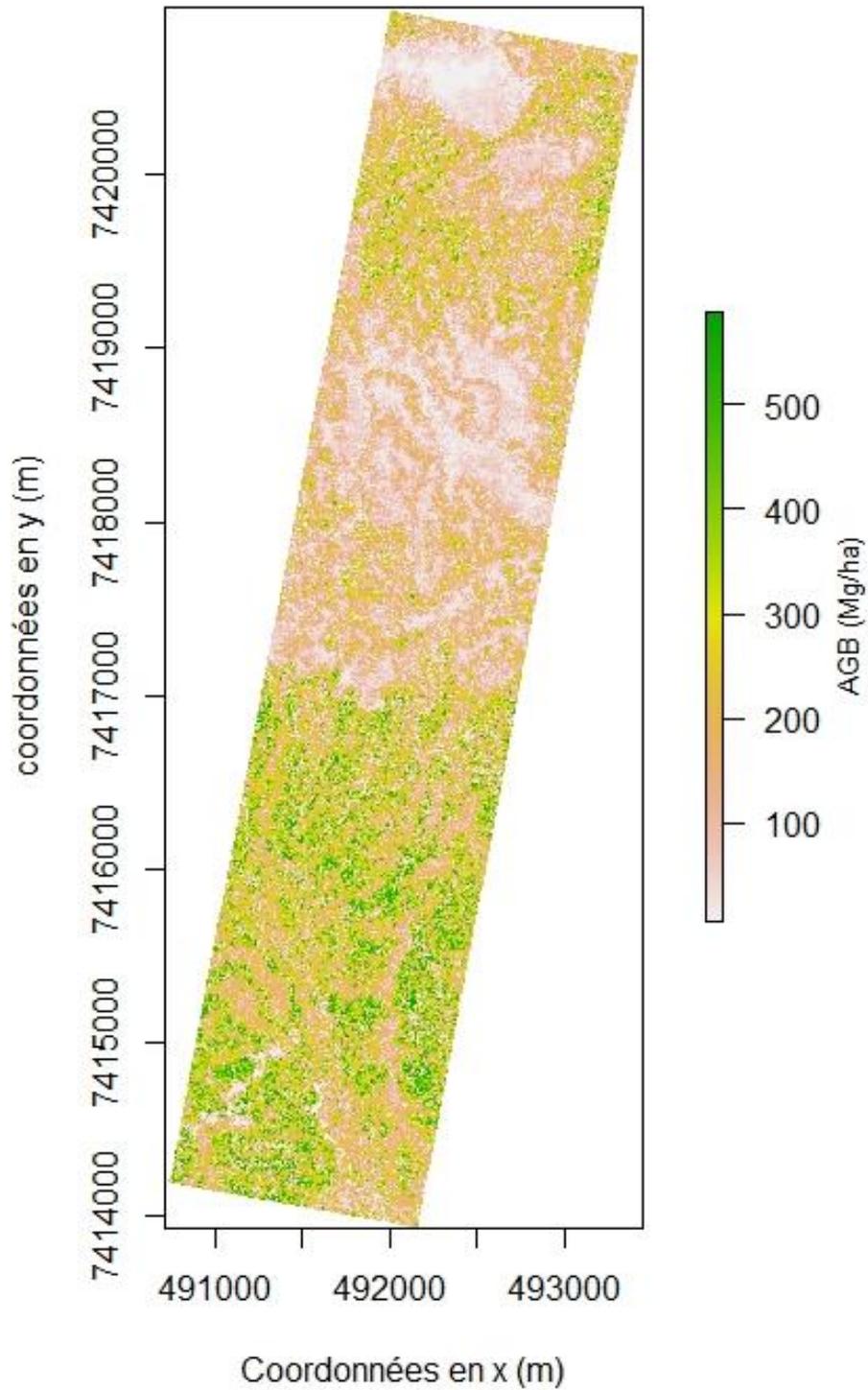
## Annexe 2

Échantillon de données échantillonnées à 1 point / m<sup>2</sup> à gauche et à 18 points / m<sup>2</sup> à droite



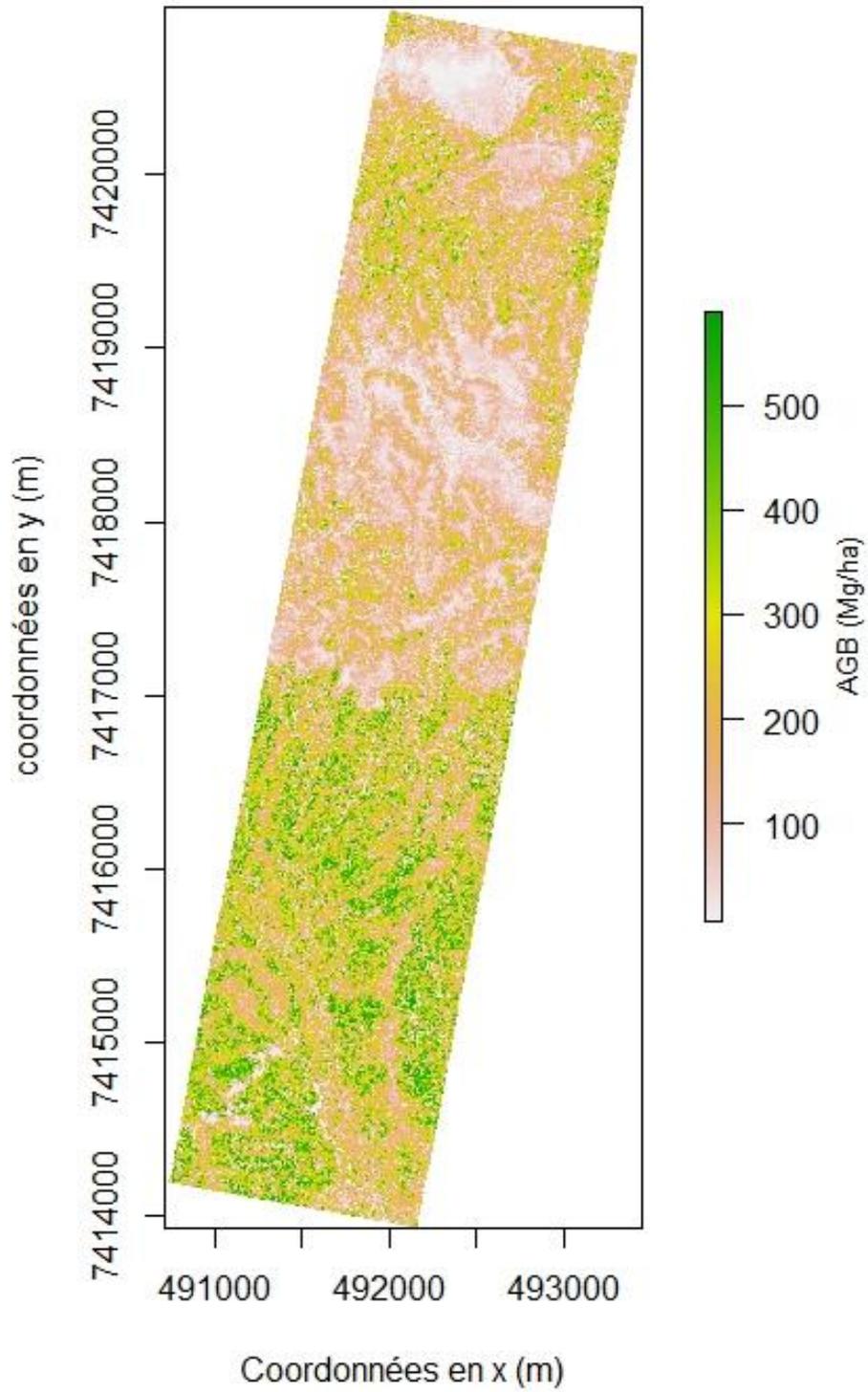
### Annexe 3

## Cartographie de la biomasse aérienne à partir des données LiDAR de densité de 3 points / m<sup>2</sup>



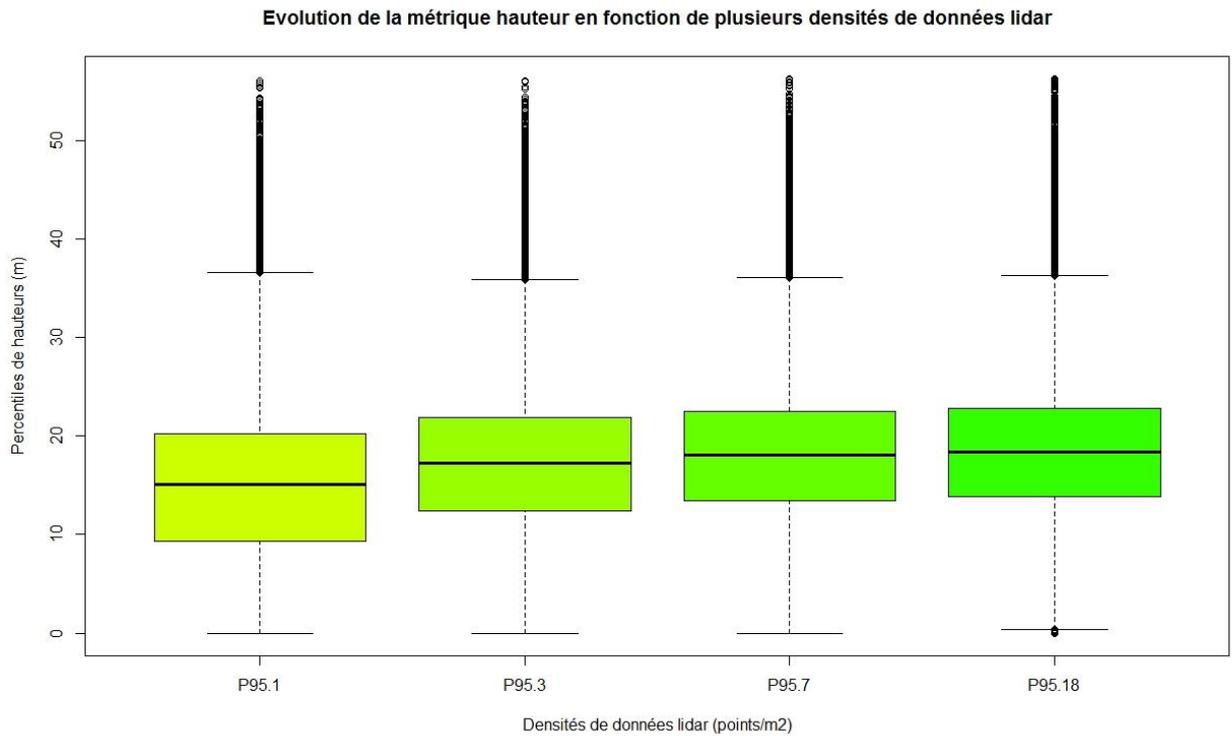
## Annexe 4

### Cartographie de la biomasse aérienne à partir des données LiDAR de densité de 7 points / m<sup>2</sup>



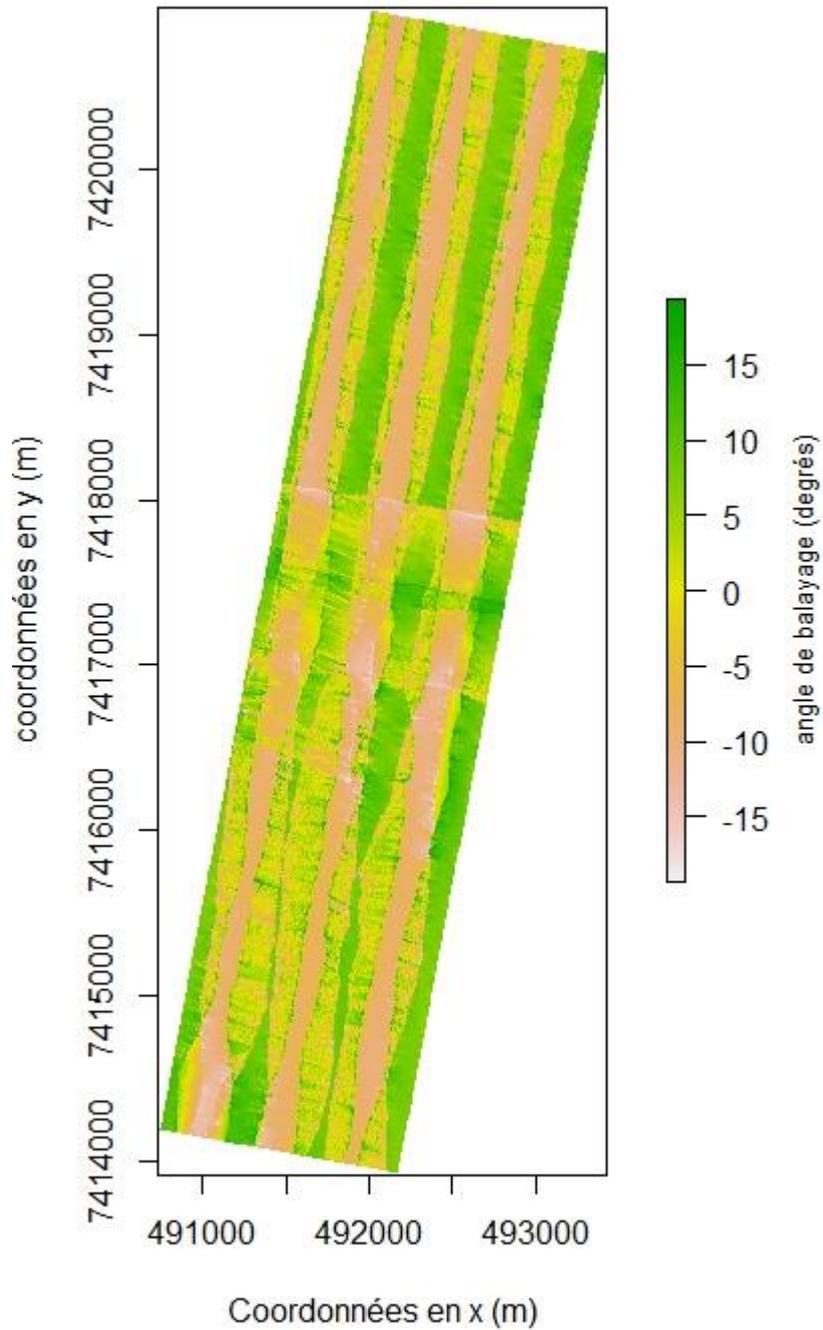
## Annexe 5

### Comparaison de l'estimation des 95 èmes percentiles de hauteur selon différentes densités de données LiDAR



## Annexe 6

### Cartographie de l'angle de balayage du système LiDAR aérien à partir des données originales 2017



## Liste des figures

Figure 1: canopée en forêt tropicale d'Amérique du Sud, source : <a href="https://www.aquaportail.com/synonyme-11461-canopee.html">https://www.aquaportail.com/synonyme-11461-canopee.html</a> .....	9
Figure 2: (a) probabilité qu'une impulsion retourne une hauteur selon une canopée toute plate. Une impulsion peut forcément toucher la canopée avec la hauteur $h_0$ suivant une probabilité $p_0$ , dont $p_0 = 1$ . (b) $h_0$ est la valeur attendue de $h_{max}$ , peu importe le nombre de points échantillonnés. ....	13
Figure 3: (a) Probabilité qu'une impulsion revienne à une hauteur $h$ selon une canopée plate avec une singularité. Une impulsion peut revenir à une hauteur $h_0$ ou $h_1$ suivant les probabilités $p_0$ et $p_1$ . (b) La $h_{max}$ a été simulée par Roussel et al (2017) en fonction du nombre de points utilisés $n$ et la valeur $h$ attendue a été calculée en utilisant l'équation ( 4 ).....	14
Figure 4: (a) Probabilité d'une impulsion retourne une hauteur selon une canopée à deux singularités. Une impulsion peut retourner trois valeurs de hauteur : $h_0$ , $h_1$ , $h_2$ avec une probabilité respective de $p_0$ , $p_1$ , $p_2$ . (b) Graphique de Roussel et al (2017) qui représente la hauteur maximale observée qui a été simulée, $h$ a été calculée avec l'équation ( 8 ).....	16
Figure 5: (a) Probabilité d'une impulsion retourne une hauteur $h$ selon une canopée continue (b) la hauteur attendue a été calculée par Roussel et al (2017) avec l'équation ( 10 ).....	17
Figure 6: effet de la taille de l'empreinte sur la hauteur maximale observée $h_{max}$ . La couleur orange en arrière-plan appartient à une empreinte dix fois plus grande que les petites empreintes. Les colonnes rouges appartiennent aux impulsions de faible empreinte. L'empreinte la plus grande a une forme d'onde plus large car l'intensité est intégrée sur une plus grande surface. Par conséquent, la grande empreinte sous-estime la hauteur maximale enregistrée par les plus petites empreintes, cela signifie que la hauteur du pic de la forme d'onde orange est inférieure au pic rouge le plus élevé.....	18
Figure 7: emplacement des données LiDAR.....	21
Figure 8: affichage sur le logiciel R de l'ensemble des données LiDAR. ....	22
Figure 9: Illustration de Véga et al (2014) de l'effet de la normalisation sur la géométrie du nuage de points dans le cas de fortes pentes. ....	26
Figure 10 : résultat de la normalisation d'une partie des données affichées sur le logiciel Quick Terrain Reader .....	26
Figure 11 illustration qui montre le principe du modèle de hauteur de canopée, source: <a href="https://www.researchgate.net/figure/Illustration-du-principe-de-creation-du-MNC-a-partir-des-MNS-et-MNT_fig1_301228831">https://www.researchgate.net/figure/Illustration-du-principe-de-creation-du-MNC-a-partir-des-MNS-et-MNT_fig1_301228831</a> .....	27

Figure 12: graphique des hauteurs de chaque impulsion en fonction des probabilités de chaque retour associé à une hauteur donnée suivant une canopée de forêt tropicale au Brésil, calculé avec une résolution de pixel de 400 m <sup>2</sup> .....	31
Figure 13: résultats de Roussel et al (2017) : graphique des hauteurs de chaque impulsion en fonction des probabilités de chaque retour associé à une hauteur donnée suivant une canopée de forêt typique d'Amérique du Nord, calculé avec une résolution de pixel de 400 m <sup>2</sup> . .....	31
Figure 14: évaluation du biais de l'estimation de la hauteur maximale avec une résolution de 400 m <sup>2</sup> en fonction de la densité d'impulsions .....	32
Figure 15: évaluation du biais de l'estimation de la hauteur maximale avec une résolution de 4 m <sup>2</sup> en fonction de la densité d'impulsions .....	33
Figure 16 : résultats de Roussel et al (2017): évaluation du biais de l'estimation de la hauteur maximale en fonction de la densité d'impulsions.....	33
Figure 17: Retour de la forme des impulsions (Fallah et Onur (2011)) .....	38
Figure 18: cartographie de la biomasse aérienne à partir des données LiDAR de densité de 1 point / m <sup>2</sup> .....	47
Figure 19: cartographie de la biomasse forestière à partir de densité de 18 points / m <sup>2</sup> .....	48
Figure 20: répartition de la biomasse aérienne sur toute la zone d'étude pour différentes densités de points issues du même jeu de données avec un modèle non linéaire basé sur les percentiles de hauteur. Densité 1, 3, 7, 18 correspondent aux données de biomasse issues de différentes densités de données LiDAR, par exemple densité 1 représente les données de biomasse issues de densité de 1 point / m <sup>2</sup> .....	49

## **Estimation de la biomasse forestière du Brésil à partir des données LiDAR.**

**Mémoire d'Ingénieur C.N.A.M., Le Mans 2019**

---

### **RESUME**

Les forêts tropicales sont considérées comme des stocks de carbone et biodiversité importants. Le LiDAR est une technologie qui permet de quantifier la biomasse forestière sur de grands territoires. L'estimation de la biomasse aérienne dépend de l'estimation de la hauteur qui est elle-même influencée par la densité de points acquis par LiDAR aérien. Le modèle mathématique de cette étude permet d'évaluer la précision de l'estimation de la hauteur maximale de la canopée en fonction de la densité de points. Ce mémoire montre que le biais de l'estimation de la hauteur est moins important en forêt tropicale qu'en forêt boréale du Nord du Canada. Ce phénomène est certainement dû à la forme de la canopée qui diffère puisque la forêt du nord du Canada contient des conifères de forme triangulaire, alors que la forêt tropicale du Brésil présente une canopée plutôt plate. Cette étude montre ensuite l'effet de l'estimation de la hauteur sur la biomasse forestière du Brésil. La sous-estimation est en moyenne 50 Mg / ha pour les données de densité de 1 point / m<sup>2</sup>, mais à partir de 3 points / m<sup>2</sup> la sous-estimation est quasiment négligeable. Le biais de l'estimation de la biomasse aérienne ne paraît donc peu important sur des données à faible densité. Cette étude démontre qu'il n'est donc pas nécessaire d'acquérir de données à haute densité sur la forêt Atlantique du Brésil pour une étude de biomasse basé sur l'estimation de la hauteur.

Mots clés : forêts, LiDAR, biomasse aérienne, canopée, densité de points, hauteur, estimation.

---

### **SUMMARY**

Tropical forests are considered important stocks of carbon and biodiversity. LiDAR is a technology that quantifies forest biomass in large areas. The estimation of aboveground biomass depends on the estimation of height which is itself influenced by the density of points acquired by aerial LiDAR. The mathematical model of this study evaluates the accuracy of estimating the maximum height of the canopy according to point density. This essay shows that the bias in height estimation is less in tropical forest than in boreal forest in northern Canada. This is certainly due to the different shape of the canopy, since the forest in northern Canada contains triangular conifers, while the Brazilian rainforest presents a rather flat canopy. This study then shows the effect of height estimation on forest biomass in Brazil. The underestimation is on average 50 Mg / ha for density data of 1 point / m<sup>2</sup>, but from 3 points / m<sup>2</sup> the underestimation is almost negligible. However, the bias of aboveground biomass estimation does not appear to be important on low density data. Therefore, this study shows that it is not necessary to acquire high-density data on the Atlantic forest of Brazil for a biomass study based on height estimation.

Key words : forests, LiDAR, aboveground biomass, canopy, point density, height, estimation.