

## Mémoire de stage

présenté par  
Simon PEYRIN

pour obtenir le diplôme de  
Mastère spécialisé Forêt, Nature et Société

Sujet :  
**Cartographie de biomasse en forêt tropicale guyanaise par lidar  
aéroporté**

soutenu publiquement le 02 février 2018

à AgroParisTech,  
centre de Montpellier

devant le jury suivant :

Dr Sylvie Durrieu  
Dr Grégoire Vincent  
Dr Raphaël Manlay

Examinatrice  
Tuteur de stage  
Enseignant-référent AgroParisTech



# REMERCIEMENTS

Je remercie chaleureusement mon tuteur de stage, Grégoire Vincent pour son infinie patience et son aide indispensable dans le traitement des données, l'amélioration de mes codes R et le temps passé à relire les différentes versions préparatoires du présent mémoire.

Merci à Raphaël Manlay, mon encadrant AgroParisTech, pour ses conseils et toute la partie administrative, ainsi qu'à Sylvie Durrieu pour avoir accepté le rôle d'examinatrice lors de la soutenance.

Un grand merci aux chercheurs de l'IRD qui m'ont apporté leur aide : Daniel Sabatier, Philippe Verley et Maxime Réjou-Méchain.

Merci à Gaëlle Jaouen d'Ecofog pour les données envoyées depuis la Guyane et le temps passé à répondre à mes demandes de précisions.

Merci enfin à Anthony Laybros pour sa relecture du mémoire.



## RESUME

La technologie lidar aéroporté représente un outil de cartographie de la biomasse en forêt tropicale pertinent à une échelle intermédiaire entre la parcelle d'inventaire et l'imagerie satellite. Au cours de cette étude, différentes méthodes ont été testées afin de réduire l'incertitude des prédictions de biomasse issues des modèles lidar en forêt guyanaise : correction des données d'apprentissage par la régionalisation des relations hauteur-diamètre, décomposition en modèles intermédiaires et normalisation des prédicteurs à l'aide du modèle numérique de canopée.

La prise en compte du facteur « hauteur environnante de canopée » a permis d'améliorer significativement l'estimation de la hauteur des arbres et par là, de leur biovolume, à partir de la connaissance de leur seul diamètre. Le modèle composite de prédiction de biovolume ne s'est pas montré plus performant que le modèle direct. Malgré une plus grande sensibilité à la densité de tirs lidar, les prédicteurs issus du modèle numérique de canopée se sont avérés meilleurs que ceux directement issus du nuage de points.

La prise en compte de variables topographiques et biogéographiques dans les modèles multivariés représente un axe d'amélioration prometteur. Par ailleurs, l'estimation séparée des composants « densité » et « biovolume » de la biomasse est recommandée.

## ABSTRACT

Airborne lidar technology is a tool for mapping tropical forest biomass which is relevant at an intermediate scale between inventory plot and satellite imagery. In this study, several methods were tested in order to reduce the uncertainty of models: correction of learning data through regionalization of height-diameter relationships, disaggregation into intermediate models and standardization of explanatory variables based on the canopy height model.

Taking into account the "canopy surrounding height" factor has significantly improved the diameter-based estimation of the height of the trees and hence of their biovolume. The composite model for biovolume prediction did not perform better than the direct one. Despite greater sensitivity to lidar pulse density, it was found that explanatory variables based on the canopy height model lead to better results than those directly derived from the point cloud.

The consideration of topographic and biogeographic variables in multivariate models represents a promising area for improvement. In addition, the split estimation of the "density" and "biovolume" components of biomass is recommended.



# TABLE DES MATIERES

REMERCIEMENTS .....	1
RESUME .....	3
ABSTRACT .....	3
TABLE DES MATIERES .....	5
1. INTRODUCTION .....	7
2. MATERIEL ET METHODES .....	11
2.1. Sites d'étude .....	11
2.2. Acquisition des données .....	12
2.2.1. Relevés dendrométriques .....	12
2.2.2. Acquisitions lidar .....	13
2.2.3. Fichiers de forme des placettes d'inventaire .....	14
2.3. Traitement des données .....	15
2.3.1. Création des rasters .....	15
2.3.2. Calcul des variables dépendantes .....	15
2.3.3. Prédicteurs issus du modèle numérique de canopée .....	16
2.3.4. Prédicteurs issus du nuage de points .....	16
2.3.5. Régression sur les prédicteurs .....	16
2.4. Analyse des données .....	17
3. RESULTATS .....	19
3.1. Amélioration des modèles .....	19
3.1.1. Régionalisation des relations hauteur-diamètre .....	19
3.1.1.1 Influence de la hauteur dominante de canopée .....	19
3.1.1.2 Calcul d'une allométrie locale .....	22
3.1.1.3 Spécificité des palmiers .....	23
3.1.2. Décomposition de la prédiction .....	24
3.1.3. Normalisation des prédicteurs .....	27
3.2. Production de cartes de biomasse .....	29
3.2.1. Résolution spatiale de prédiction .....	29
3.2.2. Choix des paramètres du modèle .....	31
3.2.3. Cartographie de biomasse à Paracou .....	33

4. DISCUSSION .....	35
4.1. Amélioration des données d'apprentissage	35
4.2. Forme des modèles	36
4.3. Normalisation des données	37
5. CONCLUSION .....	39
6. REFERENCES.....	41
7. LISTE DES ABREVIATIONS.....	45
8. TABLE DES TABLEAUX .....	47
9. TABLE DES FIGURES.....	49
ANNEXES .....	50
Annexe 1 : Distinction entre MNT, MNS et MNC	50
Annexe 2 : Emplacement des sites d'étude en Guyane Française	51
Annexe 3 : Site de Paracou	52
Annexe 4 : Site des Nouragues	53
Annexe 5 : Site de Montagne Tortue	54
Annexe 6 : Site de Montagne Plomb	55
Annexe 7 : Site de Saut Lavillette	56
Annexe 8 : Site d'Arbocel	57
Annexe 9 : Composition floristique	58
Annexe 10 : Structure des peuplements	60
Annexe 11 : Diamètres des arbres par site	61
Annexe 12 : Hauteur moyenne de canopée par site	62
Annexe 13 : Densité de bois par site	63
Annexe 14 : Dénivelé par site	64
Annexe 15 : Correction des mesures de diamètre	65
Annexe 16 : Jeu de données de Paracou et Montagne Tortue	66
Annexe 17 : Effet de la densité de tir sur les prédicteurs	67
Annexe 18 : Echelle spatiale de prédiction	69
Annexe 19 : Paramétrage des modèles	74



# 1. INTRODUCTION

La cartographie de la biomasse épigée (ou AGB pour *Above-Ground Biomass*) des forêts tropicales est nécessaire à l'identification des priorités de conservation, à l'exploitation durable des ressources, ainsi qu'au suivi des émissions de dioxyde de carbone dans l'atmosphère générées par leurs destruction et dégradation. Le carbone stocké dans la végétation hors sol représente en effet la réserve la plus importante et la plus directement impactée par ces dégradations (Gibbs et al., 2007). Le développement de méthodes d'estimation fiables est donc un enjeu majeur au regard des politiques internationales d'atténuation du réchauffement climatique (Asner, 2009).

La plus sûre de ces méthodes consiste à abattre un ensemble d'arbres constituant un échantillon caractéristique du type de forêt étudié, à les sécher, puis à les peser. L'approche destructive, bien que précise, est extrêmement coûteuse en temps, en argent, et impraticable sur de grandes étendues. Elle permet de constituer des bases de données ouvrant la voie à des estimations à large échelle mais la très faible surface inventoriée ne constitue pas à ce jour un ensemble réellement représentatif des forêts tropicales dans leur diversité. (Pan et al., 2011)

Une approche intermédiaire est basée sur des relevés dendrométriques effectués lors d'inventaires forestiers à l'échelle de quelques hectares, au cours desquels l'ensemble des arbres d'une zone donnée sont mesurés. La biomasse des arbres n'étant pas une information directement accessible depuis le sol, on fait appel à des modèles de régression permettant de la calculer à partir de grandeurs dendrométriques (diamètre et hauteur entre autres) (Chave et al., 2005; Chave et al., 2014; Feldpausch et al., 2012). Dans les forêts denses à canopée fermée, la hauteur est toutefois une donnée difficilement mesurable et rarement relevée (Hunter et al., 2013). Alors que l'inclusion de l'information sur la hauteur dans le calcul permet de réduire l'écart-type de l'erreur relative de prédiction biomasse totale par site (20 sites, 1808 arbres) de 19,5% à 12,5% (Chave et al., 2005), les estimations de biomasse sont fréquemment basées uniquement sur le diamètre à hauteur de poitrine DBH (*Diameter at Breast Height*) mesuré à 1,30 m du sol. Cela a pour conséquence d'introduire un risque de biais, c'est-à-dire une erreur systématique en cas de variations spatialement structurées des relations H-D (Hauteur-Diamètre) (Ngomanda et al., 2014; Simard et al., 2011).

Les calculs de biomasse et de biovolume (la biomasse étant le produit du biovolume par la densité du bois sec à l'air) effectués dans le cadre de ce stage sont tous basés sur les modèles de Chave et al. (2005). Assise sur un jeu de données de 2410 arbres abattus de plus de 5 cm de diamètre et issus de 27 sites sous les tropiques, cette étude représente un support solide pour l'estimation de biomasse à partir de données d'inventaire.

Les deux équations suivantes, applicables aux forêts tropicales humides (échantillon de 1402 arbres répartis sur 15 sites), permettent le calcul du biovolume (*Above-Ground Volume* AGV) en fonction du diamètre uniquement (Équation 1), ou en fonction du diamètre et de la hauteur (Équation 2) :

$$AGV(D) = \exp[-1,499 + 2,148 \cdot \ln(D) + 0,207 \cdot \ln(D)^2 - 0,0281 \cdot \ln(D)^3]$$

Équation 1 : Biovolume d'après Chave en fonction de D

$$AGV(D, H) = 0,0509 \cdot D^2 \cdot H$$

Équation 2 : Biovolume d'après Chave en fonction de H-D

Les techniques de télédétection par capteur embarqué à bord d'un avion ou d'un satellite permettent d'obtenir rapidement une grande quantité d'informations sur les surfaces continentales à l'échelle de paysages ou de régions entières. Les satellites à Très Haute Résolution spatiale (THR) comme Pléiades, Ikonos ou RapidEye rendent possible la détermination de la typologie de végétation à partir d'analyses texturales d'images, c'est à dire de patrons d'organisation spatiale des variations d'intensité spectrale (Couteron et al., 2006).

Le lidar (*light detection and ranging*) est un dispositif permettant la localisation d'objets dans l'espace dont le principe repose sur l'émission d'impulsions laser (tirs) à haute fréquence. La durée nécessaire à la lumière pour revenir au capteur après avoir été réfléchi par un obstacle donne une mesure de la distance les séparant. Il existe des capteurs mono-écho pour lesquels chaque tir ne donne lieu à l'enregistrement que d'un seul retour, et des capteurs multi-échocs dont chaque tir peut donner lieu à l'enregistrement de plusieurs retours. L'ensemble de ces relevés permet de générer un nuage de points (NDP) spatialisé du couvert forestier (Figure 1). La puissance du lidar réside dans sa capacité à pénétrer la végétation et ainsi obtenir à la fois des données caractérisant la surface du sol et la volumétrie de la canopée (Baghdadi et Mehrez, 2016).

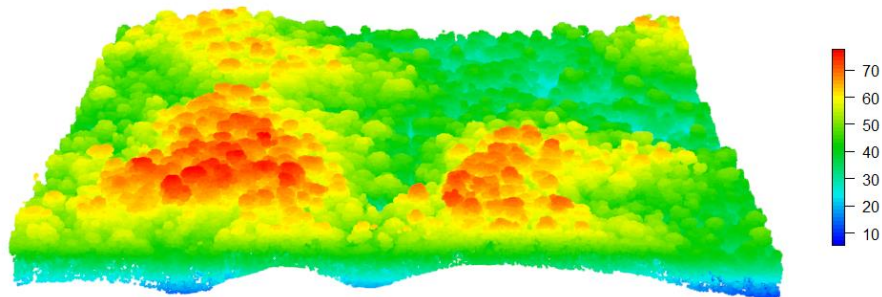


Figure 1 : Nuage de points lidar de 25 hectares à Paracou visualisé avec R  
(colorisation en fonction de l'altitude en m)  
(Source : IRD, UMR AMAP)

Le traitement des nuages de points lidar permet de générer trois rasters différents, c'est à dire trois mailles carrées régulières où une valeur d'altitude unique est attribuée à chaque pixel : un modèle numérique de terrain (MNT), de surface (MNS) et de canopée (MNC). Le MNT est la représentation spatialisée des variations d'altitude du sol par rapport au niveau de la mer. La surface forestière visible depuis le ciel constitue quant à elle le MNS. Enfin, le MNC qui est la différence (MNS-MNT), indique la hauteur du couvert forestier par rapport au sol (voir Annexe 1) (Baghdadi et Mehrez, 2016).

La technologie lidar aéroportée consiste en un émetteur lidar embarqué dans un avion équipé d'un système de Géo-Positionnement par Satellite (GPS), d'une centrale inertielle et d'un ordinateur collectant les données. Le GPS relève la position de l'avion en temps réel pendant que la centrale inertielle suit les mouvements axiaux de l'appareil. Avec les lidars récents, le signal reçu prend la forme d'une mesure continue de la quantité d'énergie réfléchiée en fonction de la distance au capteur. Chaque pic d'intensité indique un obstacle rencontré (canopée, sous-étage forestier, sol). Cette méthode permet donc une modélisation dense de la structure des zones boisées, et le dépassement des limites à la mesure in situ des métriques forestières à travers le traitement numérique de nuages de points couvrant de grandes surfaces de forêt (Pereira et Ferraz, 2014). Il pourrait donc à terme se substituer partiellement aux campagnes d'inventaires dendrométriques, ce qui nécessite le développement préalable de modèles prédictifs robustes, c'est-à-dire associés à une incertitude minimale sur la biomasse prédite.

La méthode de prédiction de biomasse à partir d'acquisitions lidar utilisée ici repose sur la recherche de corrélations entre les grandeurs statistiques caractérisant le NDP du couvert forestier, et la quantité de biomasse aérienne présente sur cette même zone.

La procédure est la suivante :

- Un ensemble de placettes d'inventaire est choisi conformément au type de végétation que l'on souhaite modéliser ;
- La biomasse (variable dépendante, expliquée par le modèle) présente sur chacune de ces placettes est calculée à partir des relevés de terrain et des équations allométriques appropriées ;
- Les grandeurs caractéristiques du NDP lidar correspondant à chaque placette sont calculées, par exemple : hauteur moyenne, déciles de la distribution des hauteurs, intensité de l'onde retour ou proportion de trouée (fraction des points de hauteur inférieure à 5 m). Ces grandeurs sont les variables explicatives du modèle, également appelées « prédicteurs » (Asner, 2009) ;
- Les prédicteurs expliquant le mieux les variations de la variable dépendante servent à la construction du modèle auquel est associée une incertitude. En 2002, Lefsky et al. proposent un modèle ajusté aux forêts tempérées et boréales basé sur le carré de la hauteur moyenne de canopée (*Mean Canopy Height* MCH) qui explique à lui seul 84% de la variance de biomasse épicée :  $AGB = 0,378 \cdot MCH^2$ .

Une fois qu'un modèle satisfaisant a été obtenu, il est possible de l'appliquer à l'ensemble de la surface forestière couverte par les acquisitions lidar à condition que le type de végétation soit suffisamment homogène.

L'UMR AMAP (Unité Mixte de Recherche en botAnique et Modélisation de l'Architecture des Plantes et des végétations) qui accueille ce stage, mène des recherches dans le développement de méthodes d'estimation de biomasse en contexte tropical.

Cette étude s'inscrit dans le cadre du projet « Intégration de données spatialisées multi-échelles pour la cartographie des types de forêt et de la biomasse en Amapá et Guyane » (Biomap) du programme franco-brésilien Guyamazon. L'objectif du stage est la production de cartes de biomasse sur plusieurs sites en Guyane française pour lesquels on dispose de relevés dendrométriques ainsi que d'acquisitions lidar. Ces cartes seront ensuite exploitées par Pierre COUTERON (IRD, UMR AMAP) comme base pour le développement de modèles de prédiction spatialisée de biomasse à partir de données satellites THR. Les données satellites permettent en effet d'extrapoler la prédiction de biomasse à plus large échelle sur la base d'analyses texturales de canopée.

L'intégralité du travail effectué l'a été à l'aide du logiciel R spécialisé dans le traitement statistique de données.

La problématique est la suivante : « Dans quelle mesure l'incertitude liée aux cartes spatialisées de biomasse peut-elle être réduite par l'amélioration des modèles prédictifs ? »

Afin d'y répondre, plusieurs pistes de réflexion ont été proposées :

- L'effet du facteur « hauteur moyenne de canopée » sur les relations H-D locales pour l'amélioration de la qualité des données d'apprentissage (Grégoire Vincent et al., 2012) ;
- La réduction d'incertitude associée à la décomposition des composantes de la biomasse épigée (Vincent et al., 2014) ;
- La normalisation des données en entrée des modèles de prédiction : modélisation préalable du nuage de point en profil de densité de végétation (Vincent et al., 2017), ou extraction d'un simple modèle de surface de la canopée ;
- L'intérêt d'une stratification préalable par site voire éventuellement par faciès forestier (Kennel et al., 2013).

L'étude de la stratification par type forestier (dernier point) a dû être laissée de côté par manque de temps.

## 2. MATERIEL ET METHODES

### 2.1. SITES D'ETUDE

Cette étude est basée sur un jeu de données issues de six dispositifs forestiers permanents localisés dans la moitié Nord de la Guyane française et appartenant au réseau Guyafor géré par l'ONF, le Cirad et le CNRS (<http://www.ecofog.gf/spip.php?article364>) :

- Paracou (5°18' N, 52°53' W) ;
- Nouragues (4°5' N, 52°41' W) ;
- Montagne Tortue (4°22' N, 52°21' W) ;
- Montagne Plomb (5°01' N, 52°95' W) ;
- Saut Lavillette (4°15' N, 52°20' W) ;
- Arbocel (5°30' N, 53°05' W).

Voir l'emplacement des sites en Annexe 2 et leur surface inventoriée au Tableau 1.

Le site de Paracou situé dans le nord de la Guyane française, est recouvert de forêt de terre ferme de type tropicale humide. Il est constitué d'un ensemble de 16 placettes de forme carrée dont 15 de 6,25 hectares et une de 25 hectares. Neuf de ces placettes furent exploitées en 1986 et présentent donc un stade de développement différent de celles non exploitées. (<https://paracou.cirad.fr/>)

Le site des Nouragues géré par le CNRS a été créé en 1986. Le substrat est un bouclier granitique parfois dégradé aux alentours de l'inselberg des Nouragues, et un substrat métamorphique ailleurs (gneiss Caraïbe). Les sols sont argileux à sableux-argileux (<http://www.nouragues.cnrs.fr/>). Il comporte trois grandes parcelles référencées dans Guyafor : Grand Plateau (10 ha), Petit Plateau (12 ha), et Saut-Pararé (6 ha), ainsi que huit petites placettes de 1 ha : Balanfois1, Balanfois2, Bas-fond1, Bas-fond2, Lhor, Parare Ridge, Ringler et Wemomax. Balanfois1 a été écarté de l'étude à cause d'un chablis survenu entre l'acquisition des données dendrométriques et celle des données lidar (Tableau 1). Bas-fond1, Bas-fond2 et Wemomax n'ont pas non plus été pris en compte à cause d'une incertitude sur la géolocalisation des arbres mais également d'une taille limitée (0,5 ha).

Le site de Montagne Tortue est constitué de trois parcelles installées en 2002 et gérées par l'ONF couvrant une surface de 14 hectares. Deux de ces parcelles de quatre hectares chacune ont subi une exploitation en 2006 alors que la troisième a été laissée intacte pour servir de témoin.

Le site d'Arbocel a subi une coupe rase en 1976 dans le cadre d'une étude visant à évaluer la formation, la composition et l'évolution du recru forestier dans les conditions réelles d'une exploitation par les papetiers (Sarraiil et al., 1990). Il s'agit donc d'un site en cours de régénération caractérisé par un couvert forestier peu élevé et des bois de faible densité.

Les Annexes 3 à 8 présentent le MNC des différents sites d'études ainsi que l'emplacement des parcelles d'inventaire.

Les Annexes 9 à 14 exposent les différences inter-sites en termes de composition floristique, structure des peuplements, diamètre des arbres, hauteur de canopée, densité de bois et dénivelé.

## 2.2. ACQUISITION DES DONNEES

Le Tableau 1 indique site par site la surface totale inventoriée, celle couverte par les acquisitions lidar disponibles, ainsi que l'année de chaque type de relevé. Pour la cohérence des modèles, il est important que les deux types de mesures soient réalisées à des dates aussi proches que possible l'une de l'autre (G. Vincent et al., 2012).

Tableau 1 : Dates des relevés dendrométriques et lidar par site

Site	Surface d'inventaire (ha)	Surface lidar (km <sup>2</sup> )	Année d'inventaire	Année de relevé lidar
Paracou	119	10	2015	2015
Nouragues	32	24,4	2010 & 2012	2012
Montagne Tortue	14	1,9	2014	2013
Montagne Plomb	6	22,7	2013	2014
Saut Lavillette	6	395	2012	2013
Arbocel	4	3,3	2010	2009
Total	181	494		

### 2.2.1. RELEVES DENDROMETRIQUES

Les données dendrométriques de tous les sites ont été transmises par Gaëlle Jaouen (Ecofog) sous la forme d'un tableau indiquant pour chaque arbre les informations suivantes :

- Identifiant ;
- État (vivant ou mort) ;
- Coordonnées locales (dans le repère de sa parcelle) et coordonnées UTM ;
- Famille, genre et espèce ;
- Circonférence des troncs corrigée en cm ;
- Année d'inventaire.

La correction de la mesure de circonférence des troncs concerne les arbres des parcelles suivies depuis au moins trois campagnes d'inventaire. Les mesures relatives aux individus montrant une croissance en diamètre supérieure à 5 cm ou une décroissance inférieure à 2 cm subissent une correction par régression linéaire établie à l'aide des mesures antérieures et postérieures. D'autres corrections s'appliquent spécifiquement aux Bugu Bugu (bois cathédrale) aux troncs très irréguliers, ainsi qu'aux très gros arbres (voir Annexe 15).

Il est à noter que seuls les individus de diamètre supérieur à 10 cm ont été mesurés et que parmi eux, seuls les arbres en vie ont été pris en compte dans cette étude. Les mesures de circonférences des troncs ont été généralement effectuées à hauteur de poitrine, c'est-à-dire à 1,30 m du sol ou au-dessus des contreforts. Le diamètre des arbres a été déduit à partir de leur circonférence en divisant cette dernière par  $\pi$ .

Gaëlle Jaouen a également transmis un tableau indiquant la densité de bois sec à l'air (12% d'humidité) de plus de 1500 essences tropicales ce qui a permis d'associer à la plupart des individus de notre jeu de donnée la valeur correspondante. En cas de donnée manquante,

il a été décidé d'attribuer à l'individu concerné la densité de bois moyenne dans son environnement proche (parcelle d'un hectare).

## 2.2.2. ACQUISITIONS LIDAR

Le Tableau 2 résume les conditions dans lesquelles les acquisitions lidar ont été effectuées. On constate une grande dispersion dans la densité moyenne de tir qui est un indicateur de la précision des relevés. La plupart des scanners lidar utilisés sont multi-échos. Seul celui, plus ancien (2009), ayant servi aux acquisitions d'Arbocel est mono-écho, donc moins pénétrant et ne relevant que les derniers retours.

Tableau 2 : Conditions d'acquisitions des données lidar

Site	Scanner lidar	Densité de tir (m <sup>-2</sup> )	Opérateur	Commanditaire
Paracou	LMS-Q780	32	Altoa	Labex CEBA
Nouragues	LMS-Q560	13	Altoa	CNRS
Montagne Tortue	LMS-Q560	6	Altoa	ONF
Montagne Plomb	LMS-Q560	10	Altoa	IRD
Saut Lavillette	LMS-Q560	5	Altoa	ONF
Arbocel	LMS-280i (mono écho)	5	Altoa	Cirad

Les nuages de points se présentent numériquement sous la forme de fichiers possédant l'extension .las ou .laz.

Parmi les métadonnées de ces fichiers, on trouve pour chaque écho :

- Coordonnées spatiales XYZ du point correspondant ;
- Identifiant du tir ;
- Temps GPS correspondant à l'instant d'émission du tir ;
- Numéro du retour ;
- Nombre total de retours correspondant au même tir ;
- Classification : points aberrants, points sol, points sursol (végétation, bâtiment) ;
- Angle de tir.

La sélection dans R des points classés « sol » et « sursol » a permis de supprimer les points aberrants car situés sous le niveau du sol ou au-dessus de la canopée (dus à des passages d'oiseaux par exemple).

### 2.2.3. FICHIERS DE FORME DES PLACETTES D'INVENTAIRE

Un fichier de forme (ou *shapefile*) constitue un format non topologique simple permettant de stocker les informations relatives à l'emplacement géométrique et aux attributs des entités géographiques. Les entités géographiques d'un fichier de formes peuvent être représentées par des points, des lignes et des polygones (ou surfaces). Il peut donc représenter aussi bien les limites d'une parcelle d'inventaire que l'emplacement de l'ensemble des arbres qui s'y trouvent (<http://desktop.arcgis.com/fr/>).

Un fichier de forme est en réalité composé de plusieurs fichiers. Les trois suivants sont requis :

- .shp fichier contenant la géométrie des entités ;
- .dbf fichier contenant les attributs au format dBase ;
- .shx fichier d'index.

Il inclut également un fichier ayant l'extension .prj qui contient les informations sur le système de coordonnées. Bien que ces informations soient très utiles elles ne sont pas obligatoires car le logiciel se base sur la valeur des coordonnées moyennes pour retrouver le système de projection ainsi que le datum adapté.

Afin de disposer pour notre modélisation d'un ensemble de zones de mêmes dimensions et d'en extraire des grandeurs statistiques comparables entre elles, toutes les parcelles ont été subdivisées en placettes carrées de 1 ha qui sont au nombre de 147 sur l'ensemble des sites (voir Annexes 3 à 8). La création des fichiers de forme des placettes des Nouragues, Montagne Tortue et Saut Lavillette a été effectuée sur R au cours du stage. Celles de Paracou, Montagne Plomb et Arbocel m'ont été fournies par l'UMR AMAP.

Le logiciel QGIS a permis de vérifier la cohérence entre les coordonnées GPS des arbres et celles des parcelles. Ainsi, le fichier de forme des placettes de Saut Lavillette a été décalé de 4,40 m vers le sud et 1,50 m vers l'est pour faire correspondre la répartition géographique des arbres aux limites de la zone.

Le système de projection utilisé est le WGS84-22N.



## 2.3. TRAITEMENT DES DONNEES

Les relevés dendrométriques, acquisitions lidar et fichiers de forme ont servi à produire les deux types de données d'apprentissage nécessaires à la construction des modèles prédictifs de biomasse, c'est-à-dire variables dépendantes et prédicteurs.

### 2.3.1. CREATION DES RASTERS

Sur chaque site, trois rasters ont été créés pour les besoins de l'étude : le MNC, le MNT et le H50.

Le MNC (résolution de 1 m) a été généré à partir du NDP grâce à la bibliothèque `{lidR}` de R. La fonction `lasnormalize` a permis de normaliser les données en plaçant tous les points identifiés « sol » à l'altitude 0. La fonction `grid_canopy` a ensuite sélectionné les points les plus hauts et les a rangés sous la forme d'un tableau à trois colonnes XYZ qui a été converti en raster à l'aide de la fonction `rasterFromXYZ` de la bibliothèque `{raster}`. Lorsque la faible densité du NDP lidar a induit des pixels non-renseignés dans le MNC, ces pixels ont reçu en post-traitement une valeur égale à la valeur médiane des huit points voisins sur la grille.

Le MNT (résolution de 1 m) est issu du même nuage de point que celui ayant servi à la création du MNC. La fonction `grid_terrain` `{lidR}` appliquée au fichier `.las` non normalisé a permis de générer le tableau des coordonnées spatiales qui a été converti en raster avec `rasterFromXYZ` `{raster}`.

La valeur de chaque pixel du H50 (résolution de 1 m) correspond à la hauteur moyenne de la canopée dans un carré de 51 m de côté dont il est le centre. Ce raster a été généré à partir du MNC grâce à la fonction `focal` `{raster}` de R.

### 2.3.2. CALCUL DES VARIABLES DEPENDANTES

Pour les besoins de l'étude, un ensemble de variables dépendantes ont été calculées à l'échelle de chacune des 147 placettes de 1 ha. Pour cela, il a fallu attribuer les arbres des parcelles d'inventaires à la placette correspondante ce qui a été fait à l'aide des fichiers de forme grâce à la fonction `over` `{sp}` de R.

Les variables dépendantes sont :

- N : nombre de tiges ;
- $D_g$  : diamètre quadratique moyen ( $\text{cm}^2$ ) ;
- G : surface terrière ( $\text{m}^2$ ) ;
- $\rho$  : densité moyenne de bois moyenne pondérée par le biovolume ( $\text{kg}/\text{dm}^3$ ) (voir calcul en Annexe 13) ;
- AGV(D) : biovolume calculé d'après l'Équation 1 p.8 en fonction du diamètre mesuré des arbres ( $\text{m}^3$ ) ;
- AGV(D,  $H_{(D,H_p)}$ ) : biovolume calculé selon l'Équation 2 p.8 en fonction du diamètre mesuré des arbres et de leur hauteur estimée d'après l'Équation 4 p.22 ( $\text{m}^3$ ) ;
- AGB(D) : biomasse, produit de la densité de bois et de AGV(D) (t) ;
- AGB(D,  $H_{(D,H_p)}$ ) : biomasse, produit de la densité de bois et de AGV(D,  $H_{(D,H_p)}$ ) (t).

### 2.3.3. PREDICTEURS ISSUS DU MODELE NUMERIQUE DE CANOPEE

Les 17 prédicteurs issus du MNC sont :

- `z_mean` : hauteur moyenne (m) ;
- `z_med` : hauteur médiane (m) ;
- `z_sd` : écart-type de la distribution des hauteurs (m) ;
- `z_dx`,  $x = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0,99\}$  : déciles et 99<sup>e</sup> percentile de la distribution des hauteurs (m) ;
- `z_inf5m` : proportion de points inférieurs à 5 m ;
- `z_skew` : coefficient d'asymétrie (*skewness*) de la distribution des hauteurs ;
- `z_kurt` : coefficient d'aplatissement (*kurtosis*) de la distributions des hauteurs ;
- `z_entr` : fonction `entropy {lidR}` appliquée à la distributions des hauteurs (mesure de la dispersion des valeurs).

### 2.3.4. PREDICTEURS ISSUS DU NUAGE DE POINTS

Les prédicteurs sont extraits du NDP normalisé avec `lasnormalize {lidR}` (points sol à l'altitude 0). A l'exception des deux derniers dans la liste, ils ont tous été calculés à l'aide de la fonction `lasmetrics` associée à `stdmetrics {lidR}`.

Ils sont au nombre de 26 :

- `zmax` : hauteur maximale (m) ;
- `z_mean` : hauteur moyenne (m) ;
- `z_sd` : écart-type de la distribution des hauteurs (m) ;
- `z_skew` : coefficient d'asymétrie (*skewness*) de la distribution des hauteurs ;
- `z_kurt` : coefficient d'aplatissement (*kurtosis*) de la distribution des hauteurs ;
- `z_entr` : fonction `entropy {lidR}` appliquée à la distribution des hauteurs ;
- `z_qx`,  $x = \{10, 20, 30, 40, 50, 60, 70, 80, 90\}$  : percentiles de la distribution des hauteurs ;
- `z_pcumx`,  $x = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$  : pourcentage cumulé de retours dans la x<sup>e</sup> couche de hauteur ;
- `cov` : rapport du nombre de premiers retour au-dessus de 1,37 m (valeur conventionnelle de hauteur de poitrine correspondant à 4,5 pieds) sur le nombre total de premiers retours ;
- `dens` : nombre de retours au-dessus de 1,37 m sur le nombre total de retours.

Pour plus de précisions sur les méthodes de calcul, voir :

<https://github.com/Jean-Romain/lidR/wiki/stdmetrics>.

### 2.3.5. REGRESSION SUR LES PREDICTEURS

Afin d'obtenir une estimation numérique des variables dépendantes (Y) en fonction des prédicteurs (X), trois fonctions de régression ont été testées : `plsr {pls}`, `autopls {autopls}`, et `randomForest {randomForest}`.

Les deux premières sont basées sur le principe de la régression des moindres carrés partiels (*Partial Least Squares PLS*) et sont particulièrement pertinentes quand on a besoin de prédire un ensemble de variables dépendantes à partir d'un ensemble très grand de prédicteurs qui peuvent être fortement corrélés entre eux. La *PLS* maximise la corrélation entre X et Y par la recherche de variables latentes liées à X et à Y servant à exprimer la régression de Y sur ces variables et finalement de Y sur X. A la différence de `plsr`,

autopls procède à une sélection préalable de variables.  
([https://doc.rero.ch/record/7954/files/mem\\_VancolenS.pdf](https://doc.rero.ch/record/7954/files/mem_VancolenS.pdf))

La fonction `randomForest` se base sur la méthode des forêts d'arbres décisionnels proposée en 2001 par Cutler et al.

## 2.4. ANALYSE DES DONNEES

La qualité des modèles de prédictions a été évaluée à l'aide de quatre indices :

- $R^2$  : coefficient de détermination de la régression linéaire des observations (valeurs prises par la variable dépendante prédite) en fonction des prédictions du modèle (droite en bleu sur la Figure 2) ;
- CCL : coefficient de corrélation linéaire (coefficient directeur de la régression linéaire des observations en fonction des prédictions) ;
- RMSE : erreur quadratique moyenne entre observations ( $x_i$ ) et prédictions ( $y_i$ ) ;
- $RMSE/\bar{x}$  : rapport entre erreur quadratique moyenne et moyenne des observations. Indicateur de l'incertitude relative du modèle de prédiction.

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

n : nombre d'observations

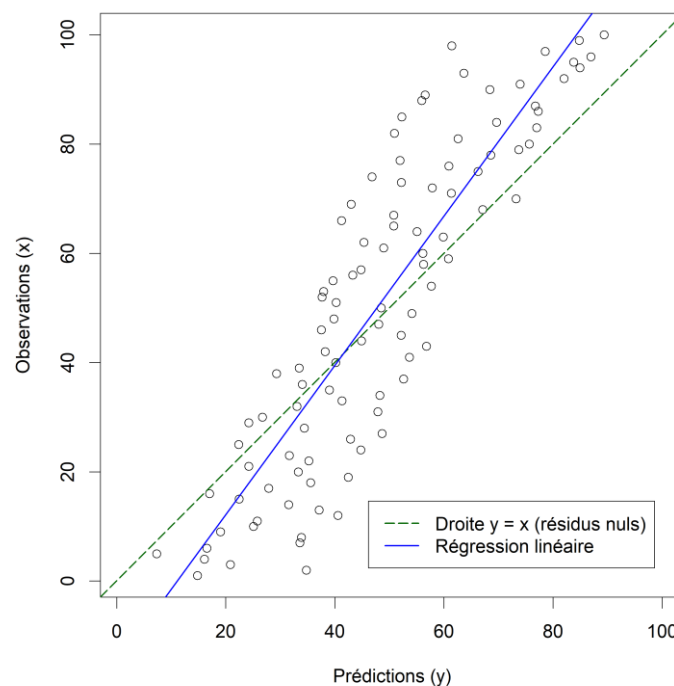


Figure 2 : Nuage des observations en fonction des prédictions



## 3. RESULTATS

### 3.1. AMELIORATION DES MODELES

#### 3.1.1. REGIONALISATION DES RELATIONS HAUTEUR-DIAMETRE

##### 3.1.1.1 Influence de la hauteur dominante de canopée

Comme précisé dans l'introduction, l'emploi de relations allométriques pantropicales pour l'estimation de biomasse présente localement un risque de biais. Un moyen de réduire ce biais est de tenir compte des effets de la hauteur de végétation sur l'allométrie locale. Des études menées à Paracou ont montré une corrélation positive et très significative entre l'élancement des tiges et la hauteur moyenne du couvert à l'échelle des placettes (Vincent et al., 2012). La première partie de cette étude consiste à confirmer cette tendance en intégrant d'autres sites dans le jeu de données et à estimer la réduction du biais permise par la prise en compte du paramètre « hauteur dominante de canopée » dans les calculs d'allométrie.

Nous nous sommes donc basés sur le jeu de données indépendant du projet BRIDGE (*Bridging Information on Tree Diversity in French Guiana, and a Test of Ecological Theories*). Ce projet impliquant neuf groupes de recherche Européens en écologie tropicale fut mené de 2007 à 2008 sur neuf placettes de 1 ha issues de plusieurs sites en Guyane française. Le but était d'améliorer les connaissances sur les différentes dimensions de la diversité biologique des arbres (spécifique, fonctionnelle, évolutive) (<http://www.ecofog.gf/bridge/indexfr.html>).

Afin de déterminer la relation allométrique hauteur-diamètre des arbres sur l'ensemble des sites, plusieurs fonctions de régression ont été testées grâce à la fonction `nls` (*nonlinear least squares*) de la bibliothèque `{stats}` de R :

- Weibull :  $H = A_W \cdot [1 - \exp(-B_W \cdot D^{C_W})]$  |  $A_W, B_W, C_W$  constantes ;
- Weibull simplifié :  $H = A_{WS} \cdot [1 - \exp(-B_{WS} \cdot D)]$  |  $A_{WS}, B_{WS}$  constantes ;
- Loi puissance  $H = A_{PS} \cdot D^{B_{PS}}$  |  $A_{PS}, B_{PS}$  constantes.

La Figure 3 montre la relation existant entre le diamètre et la hauteur des arbres du jeu de données BRIDGE ainsi que les représentations graphiques des trois fonctions de régression testées.

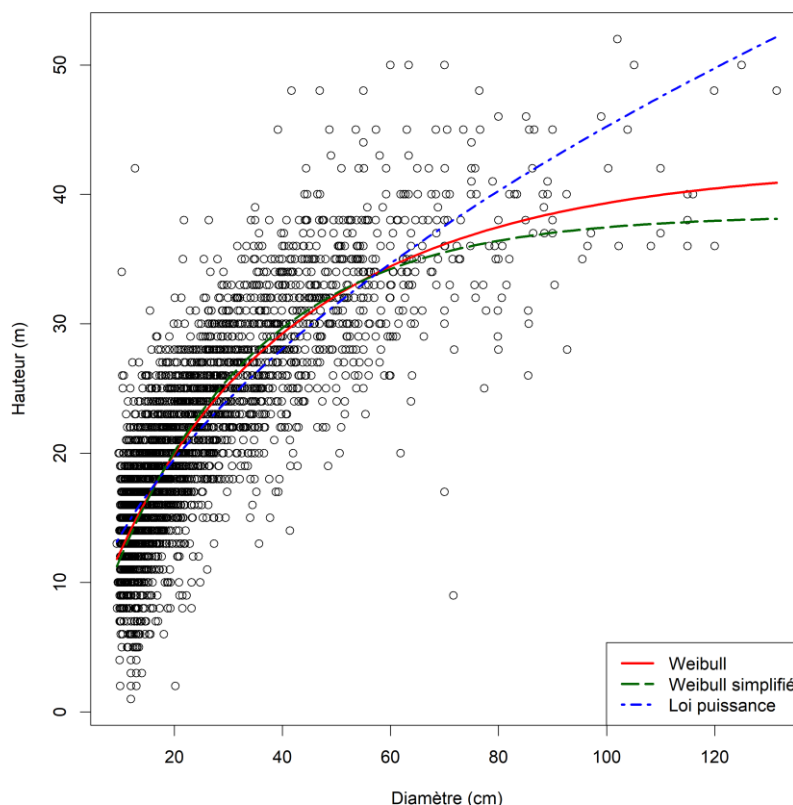


Figure 3 : Rapport hauteur-diamètre des arbres du jeu de données BRIDGE et fonctions de régression testées

Le critère d'information d'Akaike (AIC) nous permet d'effectuer une discrimination entre ces différents modèles de régression et de retenir celui offrant le meilleur compromis entre complexité (nombre de paramètres) et qualité de l'ajustement.

Tableau 3 : Qualité de régression des trois fonctions testées pour le jeu de données BRIDGE

	Weibull	Weibull simplifié	Loi puissance
$R^2$	0,71	0,71	0,70
RMSE	4,26	4,27	4,38
AIC	22 015	22 043	22 230

L'application du critère d'information d'Akaike amène à choisir la loi Weibull dont le caractère asymptotique rend compte du fait que la hauteur des arbres tend vers celle de la canopée.

On définit la hauteur dominante dans chaque placette comme la hauteur moyenne des 100 arbres de plus gros diamètre (Pardé, 1956). L'analyse de l'écart moyen par placette entre hauteurs observées et prédites par le modèle en fonction de la hauteur dominante de la canopée montre une dépendance linéaire vis-à-vis de cette dernière variable. Comme le montre la Figure 4, la hauteur moyenne des arbres est correctement estimée pour une hauteur de canopée égale à 31,1 m, surestimée en deçà de cette valeur, et sous-estimée au-delà.

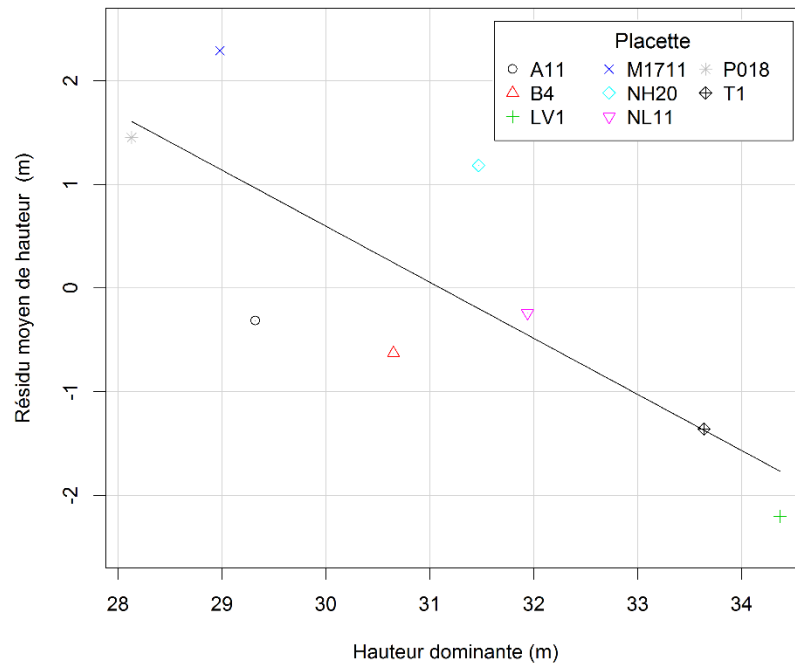


Figure 4 : Résidus moyens de hauteurs (prédictions–observations) en fonction de  $H_p$  par placette du jeu de données BRIDGE et droite de régression associée ( $R^2 = 0,63$ )

Il est donc possible d'améliorer la prédiction placette par placette par l'ajout dans l'équation allométrique du facteur correctif  $-a \cdot (H_p - 31,1)$  où  $a$  est le coefficient directeur de la droite de régression, et  $H_p$  la hauteur dominante dans la placette considérée.

On aboutit donc à l'équation suivante :

$$H = A_W \cdot [1 - \exp(-B_W \cdot D^{C_W})] - a \cdot (H_p - 31,1)$$

avec  $A_W = 42,4$  ;  $B_W = 4,55 \cdot 10^{-2}$  ;  $C_W = 0,880$  ;  $a = -0,541$ .

Par la correction de l'équation de régression initiale, la somme des valeurs absolues des résidus entre hauteur observée et hauteur prédite sur l'ensemble des arbres est réduite de 12461 m à 11928 m, soit 14 cm par arbre ou encore 7‰ de la hauteur mesurée moyenne. Nous pouvons ainsi confirmer l'hypothèse selon laquelle la relation hauteur-diamètre des arbres est dépendante de la hauteur dominante de la canopée.

### 3.1.1.2 Calcul d'une allométrie locale

De Baltzer (2015), stage portant sur l'étude des variations allométriques des arbres tropicaux, nous disposons d'un jeu de 2141 arbres parmi les 75 000 du site de Paracou dont la hauteur du houppier a été mesurée de façon fiable par des relevés lidar, et dont la correspondance tronc-houppier a été validée sur le terrain. À ces 2141 individus sur Paracou s'ajoutent 66 arbres de Montagne Tortue dont les hauteurs sont également dérivées de relevés lidars aéroportés. Cet échantillon nous sert de base pour établir une relation allométrique locale valable uniquement pour les dicotylédones (les palmiers ont été retirés, voir §3.1.1.3 p.23). Suite à l'étude de BRIDGE, les trois mêmes fonctions de régression hauteur-diamètre ont été testées (voir Annexe 16).

Le modèle représentant le mieux le NDP d'après le critère d'Akaike est de type « loi puissance ». Une régression linéaire en log-log est effectuée qui aboutit à :

$$\log(H_D) = m \cdot \log(D) + p$$

D'où :

$$H_D = \exp\left(m \cdot \log(D) + p + \frac{\sigma_D^2}{2}\right)$$

Équation 3 : Allométrie locale en fonction du diamètre

La fonction de régression linéaire  $\ln \{stats\}$  nous donne :  $m = 0,358$  et  $p = 2,05$  ( $R^2 = 0,52$  ;  $RMSE = 0,129$ ). Le terme  $\frac{\sigma^2}{2}$ , égal à la moitié du carré de la variance résiduelle du modèle, correspond au coefficient de Baskerville (Baskerville, 1972).

On définit de même la hauteur estimée corrigée en fonction de la hauteur dominante de canopée  $H_p$  :

$$\log(H_{(D,H_p)}) = q \cdot \log(D) + r \cdot \log(H_p) + s$$

D'où :

$$H_{(D,H_p)} = \exp\left(q \cdot \log(D) + r \cdot \log(H_p) + s + \frac{\sigma_{(D,H_p)}^2}{2}\right)$$

Équation 4 : Allométrie locale en fonction du diamètre et de la hauteur dominante

Avec  $q = 0,290$  ;  $r = 0,718$  et  $s = -0,0839$  ( $R^2 = 0,64$  ;  $RMSE = 0,112$ ).

Afin d'évaluer l'amélioration sur l'estimation de la hauteur apportée par la prise en compte de la hauteur dominante, l'erreur quadratique moyenne de la différence entre  $H_M$  (hauteur mesurée) et  $H_D$  a été comparée avec l'erreur entre  $H_M$  et  $H_{(D,H_p)}$  :

- $MSE(H_M - H_D) = 14,6 \text{ m}^2$  ;
- $MSE(H_M - H_{(D,H_p)}) = 10,9 \text{ m}^2$ .

L'introduction du facteur « hauteur de canopée » rendu disponible par le lidar représente donc une amélioration significative dans l'estimation de la hauteur des arbres à partir de leur diamètre.

Le Tableau 4 synthétise la biomasse sommée des 2207 arbres de Paracou et Montagne Tortue calculée de quatre façons différentes. **On fait l'hypothèse que l'estimation la plus proche de la valeur vraie est celle basée sur le diamètre et la hauteur mesurés de chaque individu selon l'Équation 2** de Chave et al. (2005). On constate que l'estimation basée sur l'Équation 1 (2<sup>e</sup> ligne du Tableau 4) induit une surestimation



importante de plus de 8%. La meilleure estimation de biomasse est celle tenant compte de la hauteur dominante de canopée car son RMSE est minimal (4<sup>e</sup> ligne du Tableau 4).

Tableau 4 : Estimation de trois façons différentes du biovolume total des 2207 arbres de Paracou et Montagne Tortue

AGV	Valeur (m <sup>3</sup> )	Diff. (%)	RMSE (dm <sup>3</sup> )
$\sum 0,0509 \cdot D^2 \cdot H_M$	7997	-	-
$\sum \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	8686	+ 8,62	855
$\sum 0,0509 \cdot D^2 \cdot H_D$	7991	- 0,08	653
$\sum 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	8031	+ 0,43	610

Note : en toute rigueur, le calcul de l'AGV devrait être noté  $\sum_{i=1}^N 0,0509 \cdot D_i^2 \cdot H_i$  avec N le nombre d'arbres. Pour le confort de lecture, une notation simplifiée a été adoptée.

### 3.1.1.3 Spécificité des palmiers

À Paracou, les palmiers (famille des *Arecaceae*) représentent 1871 individus sur les 72 671 arbres vivants que comporte le site, soit 2,6% du peuplement. Dans les calculs de biomasse, ils sont généralement traités de la même manière que les dicotylédones qui forment le reste des individus alors que leur modèle de croissance est radicalement différent. En effet, lorsque leur tronc cesse de croître en diamètre, il poursuit son développement vertical contrairement aux dicotylédones (Goodman et al., 2013). Les équations de Chave et al. (2005) étant explicitement invalides pour décrire l'allométrie des palmiers, nous aurons recours au modèle de Goodman et al. (2013) exprimant la biomasse en fonction du diamètre uniquement (la hauteur étant indisponible) :

$$\log(AGB) = -3.3488 + 2,7483 \cdot \log(D)$$

Équation 5 : Biomasse d'après Goodman en fonction de D

Dans le Tableau 5 a été recensé le nombre de palmiers sur Paracou, Nouragues et Montagne Tortue par rapport au nombre total d'arbres.

Tableau 5: Nombre de palmiers par site

Site	Nombre total d'arbres en vie	Nombre de palmiers	Proportion de palmiers
Paracou	72 671	1871	2,57%
Nouragues	11 693	290	2,50%
Mt. Tortue	7174	149	2,08%

La colonne « Chave » du Tableau 6 indique la biomasse totale (en tonnes) par site calculée d'après l'Équation 2 :

$$AGB_{Chave} = \sum_{\text{dicot+palm}} 0,0509 \cdot \rho \cdot D^2 \cdot H_{(D,Hp)}$$

Dans la colonne « Chave/Goodman » la biomasse totale a été calculée en différenciant celle des dicotylédones de celle des palmiers :

$$AGB_{\text{Chave/Goodman}} = \sum_{\text{dicot}} 0,0509 \cdot \rho \cdot D^2 \cdot H_{(D,Hp)} + \sum_{\text{palm}} \exp(-3.3488 + 2,7483 \cdot \log(D))$$

Tableau 6 : Biomasse totale par site en tonnes avec prise en compte (Chave/Goodman) et sans prise en compte (Chave) de l'allométrie spécifique des palmiers

Site	Chave	Chave/Goodman	Différence
Paracou	53 843	53 765	1,45‰
Nouragues	12 046	12 031	1,25‰
Mt Tortue	5485	5479	1,09‰

L'impact de la prise en compte de la spécificité des palmiers sur l'estimation de la biomasse est faible mais l'erreur systématique doit être corrigée. La biomasse des palmiers sera donc estimée à l'aide du modèle de Goodman et al. (2013).

### 3.1.2. DECOMPOSITION DE LA PREDICTION

Une même quantité de biomasse peut être présente sur une même surface sous la forme de peuplements ayant des structures ou des densités moyenne de bois très différentes. Par exemple, des individus de grande taille peu nombreux et espacés, versus de petits individus organisés très densément. Certains modèles de prédiction pourraient être pénalisés pour aboutir à une prédiction identique dans les deux cas. D'autre part il a été observé (Vincent et al., 2010) que le diamètre quadratique moyen est prédit de façon relativement précise avec les statistiques issues du nuage de point alors que la densité de tige l'est moins.

L'hypothèse testée ici est que l'incertitude sur la prédiction du biovolume grâce au lidar peut être réduite à l'aide d'un modèle composite. Il s'agit de le déterminer non pas directement, mais indirectement à l'issue d'un ensemble de calculs intermédiaires dont les variables feraient chacune l'objet d'un modèle de prédiction lidar.

Le biovolume total sur une surface donnée est le produit du nombre de tiges  $N$  et du biovolume moyen des arbres présents  $\overline{AGV}$  :

$$AGV = N \cdot \overline{AGV}$$

D'après (Vincent et al., 2014),  $\overline{AGV}$  dépend linéairement de  $Dg_{10}^2$ , carré du diamètre quadratique moyen au-dessus de 10 cm. D'où :

$$\overline{AGV} = a_{CH} \cdot Dg_{10}^2 + b_{CH}$$

Avec :

$$Dg_{10}^2 = \frac{1}{N} \cdot \sum_{i=1}^N (D_i - 10)^2$$

On a donc :

$$AGV = N \cdot \overline{AGV} = N \cdot (a_{CH} \cdot Dg_{10}^2 + b_{CH})$$

Équation 6 : Décomposition du calcul de la biomasse

Les coefficients  $a_{CH}$  et  $b_{CH}$  de la régression linéaire de  $\overline{AGV}$  en fonction de  $Dg_{10}^2$  dépendent uniquement de la hauteur dominante de la canopée  $H_p$  :

$$\overline{AGV}(H_p) = a_{CH}(H_p) \cdot Dg_{10}^2 + b_{CH}(H_p)$$

Sur la Figure 5 ont été tracés les biovolumes moyens de chacune des 85 placettes de 1 ha de Paracou en fonction du carré du diamètre quadratique moyen au-dessus de 10 cm. Ces biovolumes –  $AGV(D, H_{(D,H_p)})$  – ont été calculés à partir des coefficients de l'Équation 4 et de sept valeurs de hauteur dominante comprises entre 20 et 50 par pas de 5.

$$\overline{AGV}(H_p) = \frac{1}{N} \sum_{i=1}^N 0,0509 \cdot D_i \cdot H(H_p)_i$$

Avec :

$$H(H_p)_i = \exp \left( q \cdot \log(D_i) + r \cdot \log(H_p) + s + \frac{\sigma_{(D,H_p)}^2}{2} \right)$$

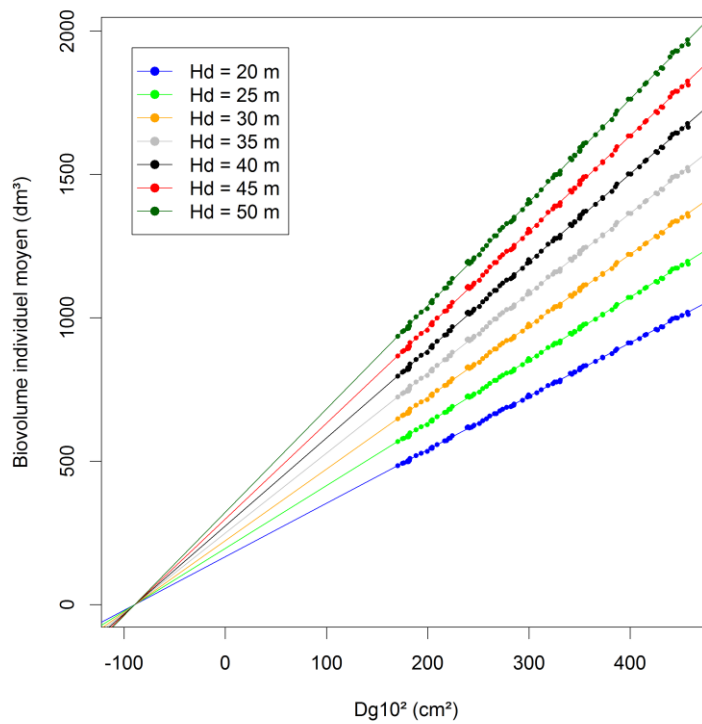


Figure 5 : Estimation du biovolume moyen par arbre de chacune des 85 placettes de 1 ha de Paracou en fonction du diamètre quadratique moyen au-dessus de 10 cm et de  $H_p$ .

Le Tableau 7 donne le coefficient directeur  $a_{CH}$  et l'ordonnée à l'origine  $b_{CH}$  de la régression linéaire de chacune des sept droites de régression obtenues grâce à la fonction  $\text{lm}\{\text{stats}\}$ .

Tableau 7: Coefficients de régression de l'AGV moyen en fonction de Hp

Hp (m)	a <sub>CH</sub> (dm <sup>3</sup> .cm <sup>-2</sup> )	b <sub>CH</sub> (dm <sup>3</sup> )
20	1,87	167
25	2,19	196
30	2,50	223
35	2,79	249
40	3,07	274
45	3,34	299
50	3,60	322

La régression linéaire de a<sub>CH</sub> en fonction de Hp (R<sup>2</sup> > 0,99) donne :

$$a_{CH}(Hp) = 0,0578 \cdot Hp + 0,743$$

Les droites de régression des sept ensembles  $\overline{AGV}(Hp)$  en fonction de  $Dg_{10}^2$  se rejoignent en un point unique de coordonnées (-89,4 ; 0) en lequel le biovolume s'annule. En effectuant un changement de repère, on peut donc écrire :

$$\overline{AGV}(Hp) = a_{CH}(Hp) \cdot (Dg_{10}^2 + 89,4)$$

D'où :

$$AGV(Hp) = N \cdot a_{CH}(Hp) \cdot (Dg_{10}^2 + 89,4)$$

Équation 7 : Décomposition du biovolume finalisé

Le Tableau 8 présente les valeurs moyennes du biovolume et son écart-type sur les placettes de Paracou calculées de trois façons différentes. La première est  $AGV(D, H_{(D,Hp)})$  (estimation la plus proche de la valeur vraie). La suivante –  $AGV(\widehat{D}, \widehat{H}_{(D,Hp)})$  – est la prédiction directe par la fonction `autopls` {`autopls`} de  $AGV(D, H_{(D,Hp)})$  à partir des métriques issues du MNC. La troisième est l'application de l'Équation 7 avec  $\widehat{N}$  et  $\widehat{Dg_{10}^2}$ , valeurs prédites par `autopls` de  $N$  et  $Dg_{10}^2$ . Le choix d'employer cette fonction est justifié en §3.2.1 p.29. Le RMSE de  $AGV(\widehat{D}, \widehat{H}_{(D,Hp)})$  est le moins élevé des deux. On retiendra donc que la simple décomposition de l'AGB n'apporte pas de gain de précision avec les modèles PLS utilisés.

Tableau 8 : Estimation du biovolume à l'aide de la méthode composite et comparaison avec la prédiction directe

Biovolume moyen par placette	Valeur (m <sup>3</sup> /ha)	RMSE (m <sup>3</sup> /ha)
$\sum 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	561 ±93	-
$AGV(\widehat{D}, \widehat{H}_{(D,Hp)})$	561 ±80	46,8
$\widehat{N} \cdot a_{CH}(Hp) \cdot (\widehat{Dg_{10}^2} + 89,4)$	560 ±79	47,7

### 3.1.3. NORMALISATION DES PREDICTEURS

Les modèles statistiques par analyse de nuages de points dépendent potentiellement fortement des conditions d'acquisition, entre autres : pénétration (mono-écho ou multi-échos), densité de tir, taille de l'empreinte (dépendante de la hauteur de vol) ou angle de scan (Silva et al., 2017). Afin de s'affranchir au moins partiellement de ces différences qualitatives (Tableau 2 p.13), et de générer des prédictions de biomasse indépendamment des conditions d'acquisition, il peut être nécessaire de normaliser l'information en entrée des modèles.

Une stratégie basée sur le calcul du profil de densité de végétation (Vincent et al., 2017) bien qu'initialement envisagée n'a pas été abordée car la mise en œuvre d'une telle approche sur de grandes surfaces n'était pas encore opérationnelle.

Afin d'évaluer les effets de la densité de tir sur l'information issue des nuages de points, une dalle couvrant une surface de 25 ha a été sélectionnée parmi celles de Paracou. A partir de cette dalle, quatre nuages de points comportant 50%, 20%, 10% et 5% du nombre de tirs du NDP d'origine ont été créés par décimation, en ne gardant les échos que de 1 tir sur 2, 1 tir sur 5, 1 tir sur 10, et 1 tir sur 20 respectivement et par ordre chronologique d'émission. Les cinq nuages de points ont été normalisés grâce à la fonction `lasnormalize` de la bibliothèque `{lidR}` et les cinq MNC correspondants ont été générés avec `grid_canopy` et `rasterfromXYZ {raster}`. Neuf fichiers de forme carrée de 1 ha ont été créés à l'intérieur de la zone afin d'en extraire les métriques des neuf placettes correspondantes (Figure 6). La moyenne et la médiane des neuf distributions de hauteur ont été moyennées pour chacun des cinq taux de densité du NDP (100%, 50%, 20%, 10% et 5%). Les résultats détaillés sont présentés en Annexe 17.

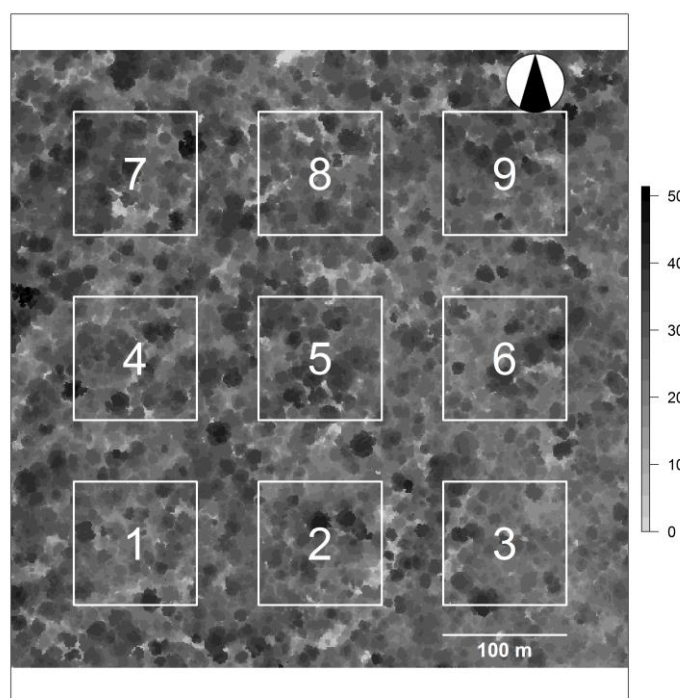


Figure 6 : Répartitions des neuf placettes de 1 ha sur la dalle de 25 ha de Paracou

La Figure 7 montre l'effet de la décimation sur l'aspect du NDP de la placette n°5.

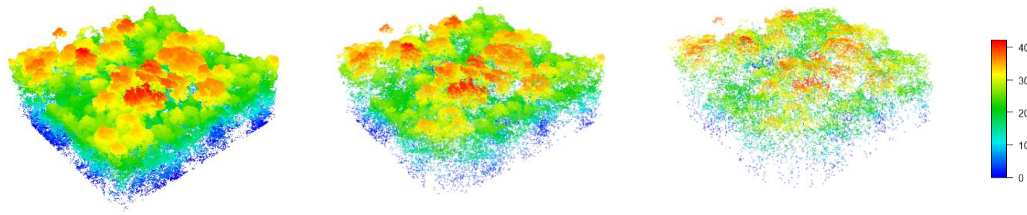


Figure 7 : Décimation du NDP de la placette 5. De gauche à droite, le taux de densité est respectivement de 100%, 20%, et 5% (colorisation en fonction de l'altitude en m).

La Figure 8 montre la manière dont les métriques (moyenne et médiane) du MNC et du NDP sont affectés par la décimation. On remarque que la moyenne et la médiane du NDP restent constantes alors que celles du MNC décroissent exponentiellement avec la diminution de densité du NDP.

Les statistiques issues du MNC apparaissent donc comme davantage sensibles aux paramètres d'acquisition lidar que ne le sont les statistiques tirées du NDP.

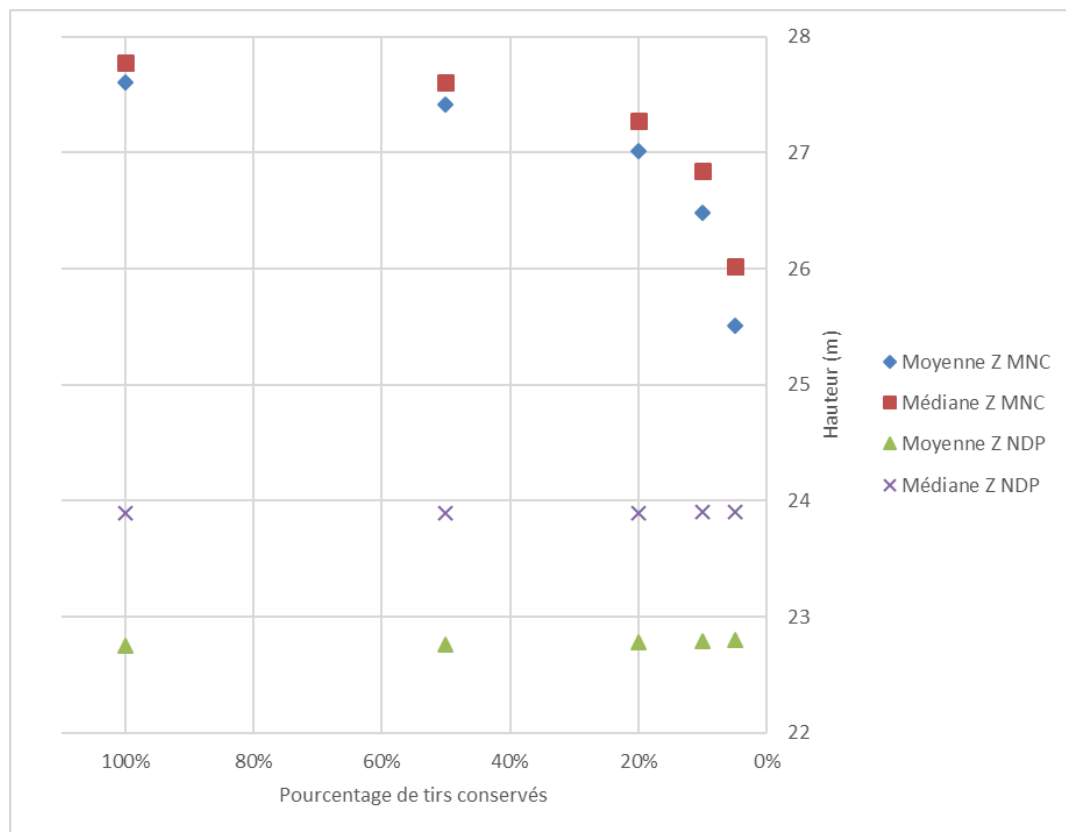


Figure 8 : Effet de la densité de tirs sur le MNC et le NDP (données en Annexe 17)

## 3.2. PRODUCTION DE CARTES DE BIOMASSE

### 3.2.1. RESOLUTION SPATIALE DE PREDICTION

Afin d'étudier les effets de la résolution sur la qualité de la prédiction, cinq échelles spatiales ont été testées. Pour cela, les arbres à l'intérieur des parcelles d'inventaire de Paracou ont été réunis au sein de placettes carrées. Cette géométrie compacte a permis d'émettre une hypothèse d'uniformité végétale à l'intérieur de la surface. Le Tableau 9 expose, pour chacune des cinq échelles spatiales testées, les dimensions des placettes, leur nombre, la surface totale échantillonnée (produit de la surface unitaire par le nombre de placettes) et l'écart-type de la distribution des mesures d'AGB( $D, H_{(D,H_p)}$ ). La surface totale échantillonnée n'est pas systématiquement égale à 119 ha car 250 et 500 (longueurs en m des côtés des parcelles de Paracou) n'étant pas multiples à la fois de 100 et 200, il n'était pas possible dans ces deux cas de couvrir l'ensemble de la surface inventoriée avec des placettes carrées de 100 et 200 m de côté.

Les variations de l'écart-type des distributions de biomasse s'expliquent d'une part par le nombre de placettes d'une échelle à l'autre, d'autre part par les variations locales de biomasse qui sont mieux captées sur de petites unités de surfaces.

Tableau 9 : Dimensions des placettes et écart-type de la biomasse mesurée à Paracou

Surface (ha)	Dimensions (m)	Nombre de placettes	Surface totale échantillonnée (ha)	Ecart-type de AGB( $D, H_{(D,H_p)}$ ) (t/ha)
0,25	50×50	475	119	106
1,00	100×100	85	85	74
1,56	125×125	76	119	74
4,00	200×200	19	76	63
6,25	250×250	19	119	63

Pour chacune de ces différentes échelles, les trois fonctions de régression décrites en §2.3.5 p.16 ont servi à prédire le biovolume et la biomasse d'après les statistiques du MNC. L'erreur quadratique moyenne du biovolume d'après la fonction et l'échelle de prédiction utilisée a été synthétisée dans la Figure 9.

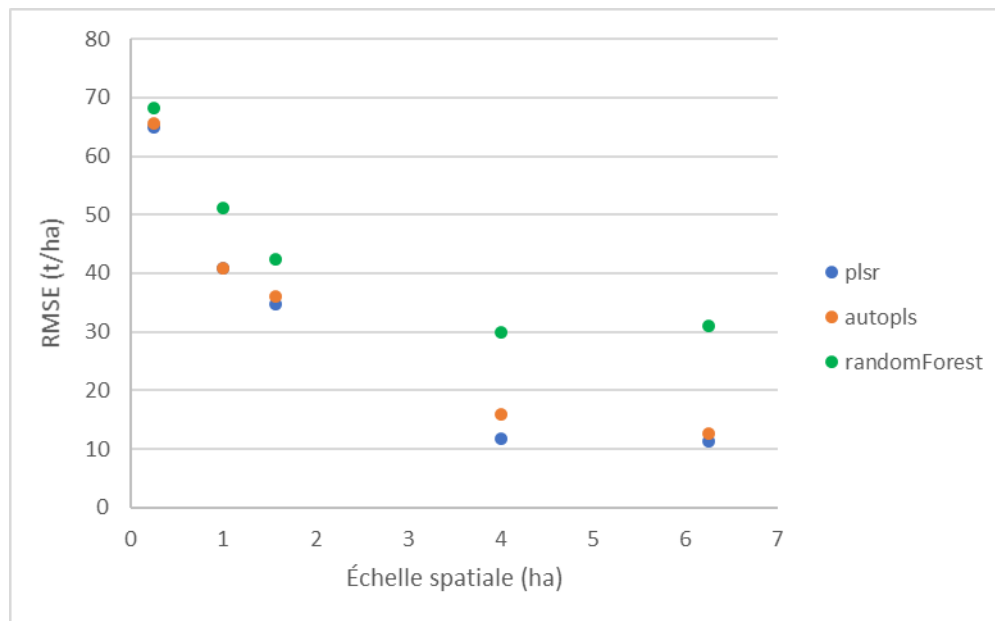


Figure 9 : RMSE de la prédiction de  $AGB(D, H_{(D,H_p)})$  en fonction de l'échelle spatiale et de la fonction de régression utilisée (données en Annexe 18)

La fonction `plsr` est la plus efficace globalement car elle minimise l'erreur quadratique moyenne à toutes les échelles sauf à celle de 1 ha où elle est supplantée par `autopl`. Bien que l'incertitude diminue avec l'augmentation de l'échelle de prédiction, il a été décidé d'effectuer la cartographie à l'échelle de 1 ha en utilisant la fonction de régression `autopl`. Cette unité de surface représente l'aire minimale de la majorité des placettes d'inventaire et de plus, une résolution plus fine est davantage représentative de l'hétérogénéité spatiale du couvert : les zones non-boisées comme les routes et les fleuves apparaissent nettement sur une carte de haute résolution, alors qu'elles constituent du bruit sur une carte grossièrement pixelisée.

Les résultats des simulations sont présentés en Annexe 18.



### 3.2.2. CHOIX DES PARAMETRES DU MODELE

Afin d'étudier la possibilité d'un modèle de prédiction unique et valable sur l'ensemble des sites, des essais ont été effectués en faisant varier cinq paramètres :

- Variable dépendante : AGV(D, H<sub>(D,Hp)</sub>) ou AGB(D, H<sub>(D,Hp)</sub>) ;
- Origine des prédictors : MNC ou NDP ;
- Modèle : de type loi puissance basé sur la hauteur moyenne du MNC/NDP ( $Y = A \cdot \bar{Z}^B$ ) ou autopl<sub>s</sub> ;
- Domaine d'apprentissage : ensemble des placettes d'où sont extraites les données d'apprentissage (ensemble des sites ou un site particulier) ;
- Domaine d'application : ensemble des placettes sur lesquelles le modèle est testé (ensemble des sites ou un site particulier) ;

Une tendance nette se dégage de l'analyse des résultats (voir Annexe 19) : la prédiction est systématiquement meilleure ( $R^2$  plus élevé, RMSE plus faible) lorsqu'elle est basée sur les statistiques du MNC plutôt que sur celles du NDP. De plus, la fonction autopl<sub>s</sub> de R se montre plus efficace que la loi puissance dans les deux cas.

Le  $R^2$  et le RMSE de la prédiction AGV recalée sur les 147 placettes à partir des statistiques du MNC valent respectivement 0,75 et 70 t/ha avec le modèle puissance, et 0,76 et 69 t/ha avec le modèle autopl<sub>s</sub>. Les mêmes grandeurs issues de la prédiction autopl<sub>s</sub> de l'AGV recalée sur les 147 placettes à partir des statistiques du NDP valent 0,74 et 73 t/ha.

La Figure 10 montre sous forme de boîtes à moustache l'erreur relative de la prédiction autopl<sub>s</sub> de AGV(D, H<sub>(D,Hp)</sub>) et AGB(D, H<sub>(D,Hp)</sub>) recalée sur les 147 placettes à partir des statistiques du MNC.

L'erreur relative a été calculée comme suit :

$$\epsilon_{AGV} = \frac{AGV_{PRED} - AGV_{OBS}}{AGV_{OBS}} \times 100$$

$$\epsilon_{AGB} = \frac{AGB_{PRED} - AGB_{OBS}}{AGB_{OBS}} \times 100$$

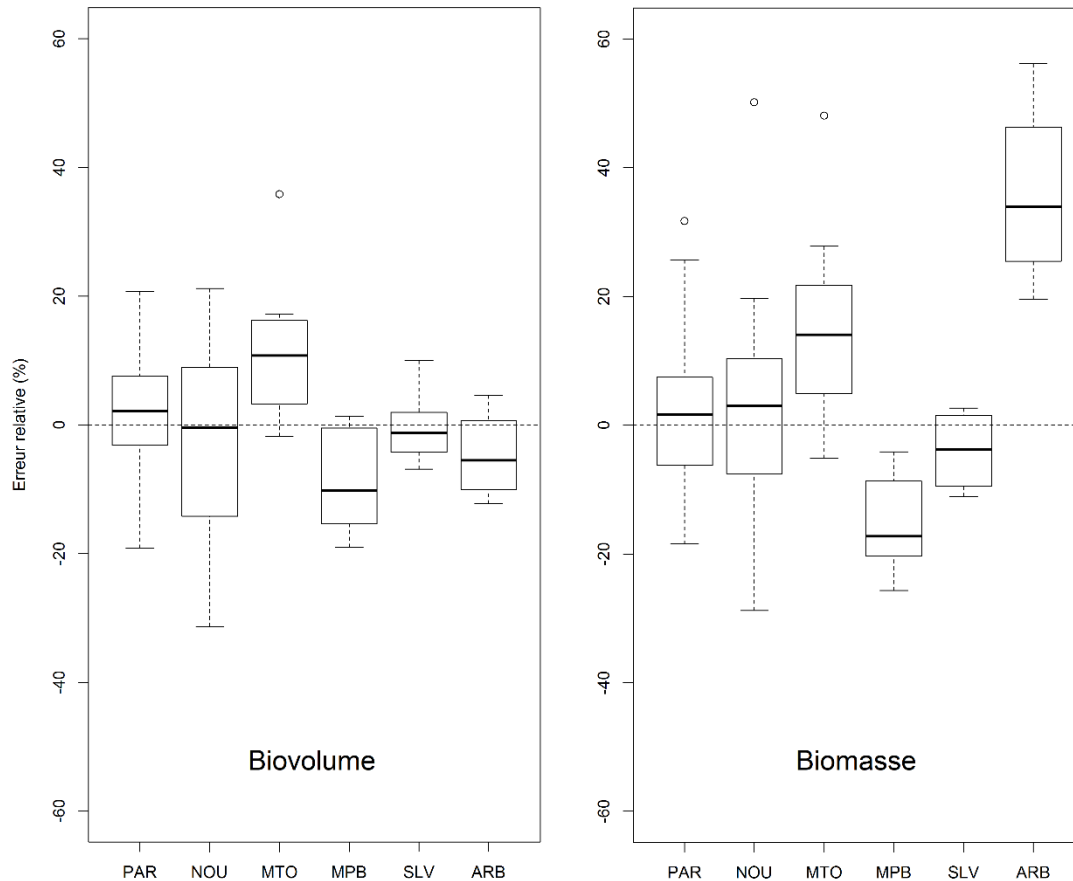


Figure 10 : Erreur relative de prédiction par site du biovolume et de la biomasse.  
En abscisse, les noms des sites ont été compactés : PAR (Paracou), NOU (Nouragues),  
MTO (Montagne Tortue), MPB (Montagne Plomb), SLV (Saut Lavillette) et ARB (Arbocel).

On observe une surestimation sensible du biovolume calculé sur Montagne Tortue et inversement, une sous-estimation sur Montagne Plomb. Ce phénomène lié au site est accentué dans le cas de la prédiction de la biomasse. La surestimation importante de la biomasse sur Arbocel est due à la densité du bois puisque la même tendance n'est pas observée avec le biovolume. Ce graphique ne rend toutefois pas compte des disparités importantes en termes de taille d'échantillonnage d'un site à l'autre : Arbocel par exemple comporte quatre placettes de 1 ha alors que Paracou en inclut 85.

L'intégration au sein des variables explicatives du modèle d'un facteur topographique (le dénivelé) permet d'augmenter le coefficient  $R^2$  du modèle `autopls` de 0,76 à 0,81 pour AGV et de 0,71 à 0,76 pour AGB. Après prise en compte de la topographie, il subsiste un « effet site » significatif : 10% de variance expliquée pour AGV et 15% pour AGB. Un tiers de l'effet site serait donc lié à la densité du bois.

### 3.2.3. CARTOGRAPHIE DE BIOMASSE A PARACOU

La Figure 11 correspond à l'estimation de AGB( $D, H_{(D,H_p)}$ ) sur Paracou à l'aide de la fonction `autopls` et à une résolution de 1 ha ( $R^2 = 0,65$  ;  $RMSE = 47$  t/ha). Les données d'apprentissage sont les métriques du MNC issues des 85 placettes du site. Les 16 parcelles d'inventaire dont elles sont issues ont été représentées en trait fin.

La zone en blanc à l'est correspond à la route et la piste menant au campement. Il s'agit de zones déboisées, donc exclues de l'aire de prédiction. Les trois pixels blancs à l'ouest correspondent à des zones où la hauteur moyenne de la canopée est inférieure à 23 m, donc en-dehors du champ des données d'apprentissage (voir Annexe 12). Elles ont également été masquées pour ne pas introduire de potentielles erreurs d'extrapolation dans la carte.

On observe une concentration de biomasse au niveau de la parcelle n°16 (la plus grande, au sud-ouest) qui n'a pas été exploitée. Cette carte constitue la meilleure estimation de biomasse sur Paracou que l'on a pu obtenir dans le cadre de cette étude.

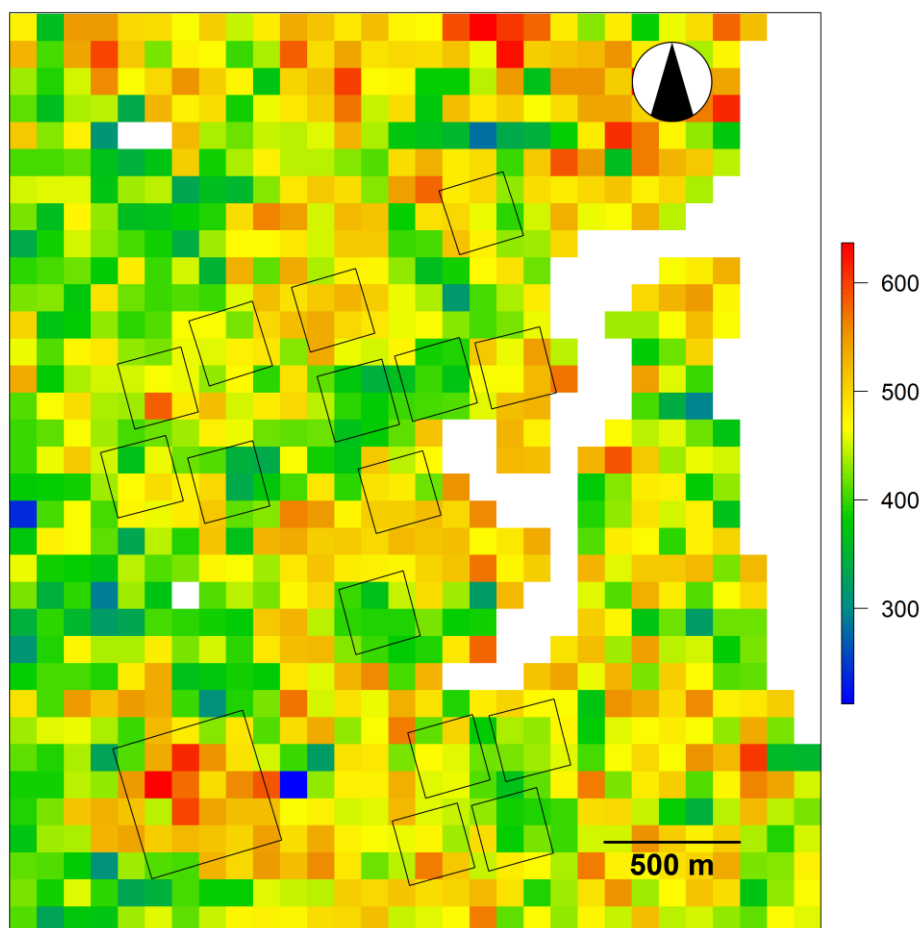


Figure 11 : Répartition de la biomasse sur Paracou (en t/ha)



## 4. DISCUSSION

L'objectif de cette étude était de tester différentes pistes d'amélioration des modèles de prédiction de biomasse forestière et d'en dresser des cartes sur nos différents sites d'étude en Guyane française. Pour cela, un travail a été fait sur les données d'apprentissage (influence de la hauteur dominante de canopée, prise en compte de l'allométrie spécifique des palmiers, normalisation des prédicteurs), sur la forme des modèles (fonction de régression employée, décomposition de la prédiction), ainsi que sur la résolution spatiale des cartes.

De multiples sources d'incertitudes et potentiels d'amélioration demeurent cependant.

### 4.1. AMELIORATION DES DONNEES D'APPRENTISSAGE

Les erreurs possibles liées à l'estimation de biomasse basée sur les inventaires de terrain sont nombreuses. Tout d'abord, seuls les arbres de plus de 10 cm de diamètre ont été pris en compte. Pourtant, des estimations menées au cours du stage sur deux jeux de données situés à Luquillo (Porto Rico) et sur l'île Barro Colorado (Panama) ont montré que les individus de faible diamètre (inférieur à 10 cm) peuvent représenter environ 3% du biovolume sur une parcelle de forêt. Les arbres morts sur pied n'ont pas été comptabilisés alors qu'ils contribuent également aux NDP lidar. Enfin les lianes, localement très présentes aux Nouragues (Tymen et al., 2016), ne rentrent pas non plus dans les inventaires forestiers. Or dans des zones fortement enliannées, si elles ne sont pas prises en compte dans les analyses, la surface terrière peut diminuer jusqu'à 8 % (Bongers et al., 2001). Les biovolumes forestiers peuvent ainsi être localement significativement sous-évalués.

La prise en compte de l'allométrie spécifique des palmiers (voir §3.1.1.3 p.23) n'a pas modifié significativement les estimations de biomasse calculées à partir des relevés de terrain. Cela est dû d'une part à la faible présence de ces derniers sur nos placettes d'inventaires, et d'autre part à l'indisponibilité de l'information sur la hauteur. Le diamètre est en effet un paramètre faiblement corrélé à la biomasse de ces arbres qui se développent verticalement sans élargir leur tronc.

Les équations allométriques de Chave pour le calcul du biovolume comportent une estimation implicite de la hauteur des arbres en fonction de leur diamètre. En combinant l'Équation 1 et l'Équation 2, on aboutit à :

$$H = \frac{\exp[-1,499 + 2,148 \cdot \ln(D) + 0,207 \cdot \ln(D)^2 - 0,0281 \cdot \ln(D)^3]}{0,0509 \cdot D^2}$$

Équation 8 : Relation H-D implicite de Chave

Sur la Figure 12 ont été tracées l'équation de la loi puissance issue de la régression de la hauteur en fonction du diamètre des arbres (Équation 3) et la relation implicite de Chave et al. (2005). On observe que cette dernière surestime sensiblement (de 5 m environ) la hauteur des arbres de gros diamètre par rapport à ce que prédit le modèle de régression simple basé sur le jeu de données de Paracou et Montagne Tortue. Cette différence est certainement due aux différences de distribution H-D entre la base de données de l'étude de Chave et la nôtre. On voit donc que, bien qu'adaptée aux forêts tropicales humides, l'Équation 1 induit localement un biais important sur l'estimation de la hauteur en fonction du diamètre. En poussant l'analyse plus loin, on peut se demander dans quelle mesure notre jeu de données local est représentatif de la végétation présente sur place. La régression a en

effet été menée sur un ensemble d'arbres de canopée dont le houppier a pu être identifié sur le MNC et les orthophotographies du site, puis associé à un individu de la base de données. Cet ensemble dont est issue la loi allométrique locale ne comporte donc aucun arbre de sous-étage et implique l'hypothèse de l'existence d'une continuité allométrique entre les arbres de canopée et ceux constituant le sous-étage forestier. La prise en compte du paramètre « hauteur dominante de canopée » rendu disponible par les acquisitions lidar a cependant permis d'obtenir des estimations de biovolume plus fines que par la seule connaissance du diamètre des individus.

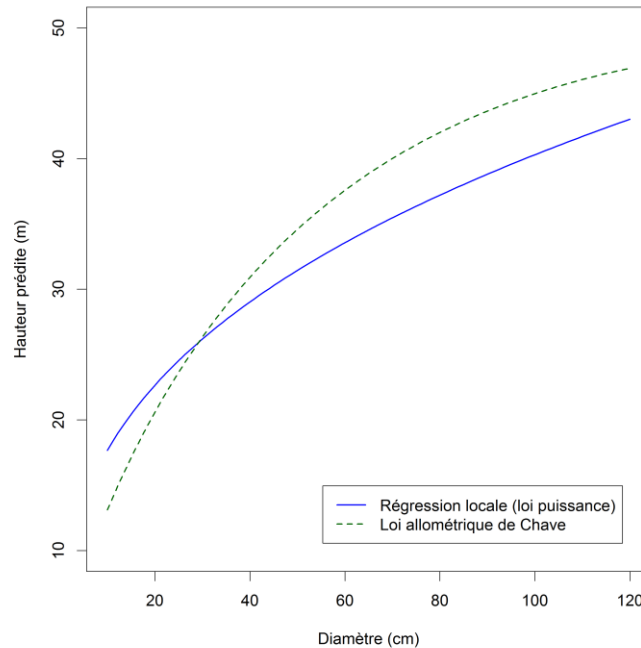


Figure 12 : Hauteur prédite des arbres selon la régression locale (Équation 3) et la loi allométrique de Chave et al. (2005).

## 4.2. FORME DES MODELES

La méthode de prédiction du biovolume par la prédiction séparée du nombre de tiges et du diamètre quadratique moyen au-dessus de 10 cm s'est montrée légèrement moins efficace que la méthode de prédiction directe (voir §3.1.2 p.24).

Les modèles multivariés issus de la prédiction `autopls` sont systématiquement meilleurs que ceux de type « loi puissance » mais de façon marginale seulement, ce qui démontre le rôle majeur que joue la hauteur moyenne du MNC/NDP dans la prédiction du biovolume.

L'étude des résidus de la prédiction du biovolume par le modèle multivarié `autopls` prenant les métriques du MNC comme prédicteurs a mis à jour un effet important lié au site (surestimation importante à Montagne Tortue, sous-estimation importante à Montagne Plomb). Le dénivelé des placettes a été analysé car les variations d'altitude du sol peuvent engendrer des erreurs dans le calcul du MNC et ainsi fausser les estimations de hauteurs des arbres (Leitold et al., 2015). L'introduction du facteur topographique n'a permis d'expliquer qu'en partie les différences observées. L'effet site résiduel peut être lié à la composition floristique. Stratifier au préalable par ensemble floristique homogène (Guitet et al., 2015) afin de pouvoir proposer un modèle adapté à chaque faciès forestier permettrait éventuellement de réduire ce biais.

Les incertitudes sur la biomasse sont systématiquement supérieures à celles sur le biovolume et donc imputables aux difficultés du modèle à prédire correctement la densité des bois. Des recherches ont été menées dans le but de trouver de nouveaux prédicteurs possibles de la densité et il apparaît que la dynamique forestière mesurable par des survols lidar répétés soit une voie prometteuse (projet DynForDiv de l'UMR AMAP).

### 4.3. NORMALISATION DES DONNEES

Nous avons constaté une grande hétérogénéité dans l'information lidar disponible (densité de tir, acquisitions mono-écho/multi-échos). Afin de rendre cette information en entrée des modèles indépendante des conditions d'acquisitions, une normalisation a été testée consistant à utiliser les statistiques du MNC plutôt que celles du NDP. L'hypothèse étant que la structure du premier serait moins sensible à la variabilité des paramètres d'acquisition. Les vérifications faites en §3.1.3 p.27 ont cependant montré un effet inverse pour le paramètre de densité de tir : les statistiques testées (moyenne et médiane) issues du MNC sont nettement plus sensibles à la densité du NDP ayant servi à générer le MNC que les métriques directement issues du NDP. Bien qu'attendu, cet effet est d'intensité marquée et cette analyse de sensibilité mériterait d'être menée pour l'ensemble des métriques utilisées dans les modèles de prédiction. L'effet de la densité de tirs sur la hauteur moyenne de la canopée a également été noté dans la littérature (Leitold et al., 2015).

Pourtant, les estimations basées sur les prédicteurs issus du MNC donnent de meilleurs résultats localement et globalement. La raison de la moindre performance des modèles basées sur le NDP reste peu claire. Une explication possible serait liée au choix des métriques retenues en entrée de la fonction `autopls`. Il a été constaté que les variables explicatives sélectionnées par la fonction changent selon le nombre de variables qui lui est proposé : l'addition d'information supplémentaire est susceptible de dégrader la prédiction. Les résultats `autopls` sont donc dépendants de la quantité de prédicteurs proposés à la fonction et il n'existe pas de garantie d'optimalité dans la sélection des variables retenues. Ce problème ne se pose pas dans le cadre d'un modèle monovarié basé sur la hauteur moyenne du MNC/NDP.

Il sera utile dans une prochaine étude de tenter une normalisation par profil de densité de végétation (distribution de la surface foliaire par unité de volume en fonction de la hauteur) ou bien par premiers échos afin de mieux capter la structure de la canopée. Une telle démarche restera cependant soumise aux conditions d'acquisitions lidar et à la difficulté d'intégrer des appareils lidar « anciens » pour lesquels un seul retour est enregistré par tir laser émis.





## 5. CONCLUSION

La principale difficulté de la prédiction de biomasse à partir de données lidar aéroporté réside dans la multiplicité des sources d'incertitudes. Relevés dendrométriques, sélection d'un jeu de données, choix d'une équation allométrique H-D, affectation de la densité des arbres, acquisitions des points lidar, traitement numérique des informations ou encore sélection d'une fonction de régression entre prédicteurs et variables dépendantes. Chacune de ces étapes implique des choix susceptibles d'affecter le résultat final. Toute investigation portant sur l'estimation de biomasse mériterait donc d'être accompagnée d'une étude de propagation des erreurs qui n'a pas pu être faite ici. Les futures recherches dans ce domaine pourront s'appuyer sur {BIOMASS} : une bibliothèque récente de R qui intègre ce type de calcul. (Réjou-Méchain et al., 2017).

L'amélioration de la prédiction de la biomasse épigée par la décomposition en fonction du diamètre quadratique, du nombre de tiges par hectare et de la densité de bois est nulle avec les métriques « standard » testées. Cependant l'utilisation de caractéristiques complémentaires (notamment de texture) permettrait peut-être un saut qualitatif. Des recherches en ce sens (associant par la signature fractale de la canopée ou la dynamique de végétation) ont donné des résultats préliminaires prometteurs pour la prédiction de la densité des bois.

En l'absence de connaissances précises sur la structure géologique, géomorphologique, climatique et biogéographiques des écosystèmes, il est extrêmement difficile de déterminer l'origine des biais observés sur nos prédictions de biomasse. L'intégration des caractéristiques topographiques permet d'améliorer sensiblement le modèle de prédiction de biomasse et devrait donc être considéré systématiquement.

Il reste un arbitrage à faire entre l'emploi d'un modèle global appliqué à l'ensemble des sites de l'étude et celui d'un ensemble de modèles locaux issus d'une régression effectuée site par site et plus performants localement. Les modèles locaux ne sont recommandables que si les données d'apprentissage sont suffisamment abondantes pour être considérées comme représentatives de la zone d'extrapolation. Par ailleurs, les modèles multivariés de type `autopls` sont sensiblement meilleurs mais induisent de fortes erreurs lorsqu'appliqués en dehors de leur zone d'échantillonnage (valeurs de biomasse négatives). Ce n'est pas le cas des modèles simples de type puissance qui prédisent une biomasse nulle pour une hauteur de végétation nulle (mais intègrent moins naturellement des co-variables comme les données issues du MNT). Il serait intéressant de dresser simultanément deux cartes, chacune issue d'une des deux méthodes, afin d'observer leurs différences de prédiction en fonction des variations spatiales de topographie et de végétation.



## 6. REFERENCES

- Asner G.P., 2009. Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters*, 4 (3), p. 034009.
- Baghdadi N., Mehrez Z., 2016. *Observation des surfaces continentales par télédétection optique*. ISTE.
- Baltzer C., 2015. *Fusion de données de télédétection et dendrométriques pour l'étude des variations allométriques des arbres tropicaux*. UMR AMAP.
- Baskerville G.L., 1972. Use of Logarithmic Regression in the Estimation of Plant Biomass. *Canadian Forestry Service*.
- Bongers F., Charles-Dominique P., Forget P.-M., Théry M., 2001. *Nouragues : Dynamics and Plant-Animal Interactions in a Neotropical Rainforest*.
- Chave J., Andalo C., Brown S., Cairns M.A., Chambers J.Q., Eamus D., Fölster H., Fromard F., Higuchi N., Kira T., Lescure J.-P., Nelson B.W., Ogawa H., Puig H., Riéra B., Yamakura T., 2005. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, 145 (1), pp. 87-99.
- Chave J., Réjou-Méchain M., Búrquez A., Chidumayo E., Colgan M.S., Delitti W.B.C., Duque A., Eid T., Fearnside P.M., Goodman R.C., Henry M., Martínez-Yrizar A., Mugasha W.A., Muller-Landau H.C., Mencuccini M., Nelson B.W., Ngomanda A., Nogueira E.M., Ortiz-Malavassi E., Pélissier R., Ploton P., Ryan C.M., Saldarriaga J.G., Vieilledent G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20 (10), pp. 3177-3190.
- Couteron P., Barbier N., Gautier D., 2006. Textural Ordination Based on Fourier Spectral Decomposition: A Method to Analyze and Compare Landscape Patterns. *Landscape Ecology*, 21 (4), pp. 555-567.
- Cutler A., Cutler D.R., Stevens J.R., 2012. Random Forests. In: *Ensemble Machine Learning*. Springer, Boston, MA, pp. 157-175. Disponible sur Internet: [https://link.springer.com/chapter/10.1007/978-1-4419-9326-7\\_5](https://link.springer.com/chapter/10.1007/978-1-4419-9326-7_5) [Consulté le 22/02/2018].
- Feldpausch T.R., Lloyd J., Lewis S.L., Brien R.J.W., Gloor M., Monteagudo Mendoza A., Lopez-Gonzalez G., Banin L., Abu Salim K., Affum-Baffoe K., Alexiades M., Almeida S., Amaral I., Andrade A., Aragão L.E.O.C., Araujo Murakami A., Arets E.J.M.M., Arroyo L., Aymard C. G.A., Baker T.R., Bánki O.S., Berry N.J., Cardozo N., Chave J., Comiskey J.A., Alvarez E., de Oliveira A., Di Fiore A., Djangbletey G., Domingues T.F., Erwin T.L., Phillips O.L., et al., 2012. Tree height integrated into pantropical forest biomass estimates. *Biogeosciences*, 9 (8), pp. 3381-3403.
- Gibbs H.K., Brown S., Niles J.O., Foley J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2 (4), p. 045023.

- Goodman R., Phillips O., Torres D., Freitas L., Tapia Cortese S., Monteagudo A., R. Baker T., 2013. Amazon palm biomass and allometry. *Forest Ecology and Management*, 310 , pp. 994-1004.
- Guitet S., Péliissier R., Brunaux O., Jaouen G., Sabatier D., 2015. Geomorphological landscape features explain floristic patterns in French Guiana rainforest. *Biodiversity and Conservation*, 24 (5), pp. 1215-1237.
- Hunter M.O., Keller M., Victoria D., Morton D.C., 2013. Tree height and tropical forest biomass estimation. *Biogeosciences*, 10 (12), pp. 8385-8399.
- Lefsky M., B. Cohen W., Harding D., Parker G., A. ACKERand S., Gower S., 2002. Lidar Remote Sensing of Aboveground Biomass in Three Biomes. *Global Ecology and Biogeography - GLOBAL ECOL BIOGEOGR*, 11 , pp. 393-399.
- Leitold V., Keller M., Morton D.C., Cook B.D., Shimabukuro Y.E., 2015. Airborne lidar-based estimates of tropical forest structure in complex terrain: opportunities and trade-offs for REDD+. *Carbon Balance and Management*, 10 , p. 3.
- Ngomanda A., Engone Obiang N.L., Lebamba J., Moundounga Mavouroulou Q., Gomat H., Mankou G.S., Loumeto J., Midoko Iponga D., Kossi Ditsouga F., Zinga Koumba R., Botsika Bobé K.H., Mikala Okouyi C., Nyangadouma R., Lépengué N., Mbatchi B., Picard N., 2014. Site-specific versus pantropical allometric equations: Which option to estimate the biomass of a moist central African forest? *Forest Ecology and Management*, 312 , pp. 1-9.
- Pan Y., Birdsey R.A., Fang J., Houghton R., Kauppi P.E., Kurz W.A., Phillips O.L., Shvidenko A., Lewis S.L., Canadell J.G., Ciais P., Jackson R.B., Pacala S.W., McGuire A.D., Piao S., Rautiainen A., Sitch S., Hayes D., 2011. A Large and Persistent Carbon Sink in the World's Forests. *Science*, 333 (6045), pp. 988-993.
- Pardé J., 1956. Une notion pleine d'intérêt: la hauteur dominante des peuplements forestiers. *Revue Forestière Française*, n° 12, pp. 851-856.
- Pereira L., Ferraz A., 2014. Estimation de la Biomasse Aérienne à partir de données lidar aéroporté. *Revue Francaise de Photogrammetrie et de Teledetection*.
- Réjou-Méchain M., Tanguy A., Pioniot C., Chave J., Hérault B., 2017. biomass: an r package for estimating above-ground biomass and its uncertainty in tropical forests. *Methods in Ecology and Evolution*, 8 (9), pp. 1163-1167.
- Sarrailh J.M., Foresta H. de, Maury-Lechon G., Prévost M.-F., 1990. La régénération après coupe papetière : parcelle Arbocel. In: Sarrailh J. M. (Éd.), *Mise en valeur de l'écosystème forestier guyanais (opération ECEREX)*. Paris, INRA, pp. 187-208. Ecologie et Aménagement Rural.
- Silva C.A., Hudak A.T., Vierling L.A., Klauberg C., Garcia M., Ferraz A., Keller M., Eitel J., Saatchi S., 2017. Impacts of Airborne Lidar Pulse Density on Estimating Biomass Stocks and Changes in a Selectively Logged Tropical Forest. *Remote Sensing*, 9 (10), p. 1068.

- Simard M., Pinto N., Fisher J.B., Baccini A., 2011. Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences*, 116 (G4), p. G04021.
- Tymen B., Réjou-Méchain M., Dalling J.W., Fauset S., Feldpausch T.R., Norden N., Phillips O.L., Turner B.L., Viers J., Chave J., 2016. Evidence for arrested succession in a liana-infested Amazonian forest. *Journal of Ecology*, 104 (1), pp. 149-159.
- Vincent G., Antin C., Laurans M., Heurtebize J., Durrieu S., Lavalley C., Dauzat J., 2017. Mapping plant area index of tropical evergreen forest by airborne laser scanning. A cross-validation study using LAI2200 optical sensor. *Remote Sensing of Environment*, 198 , pp. 254-266.
- Vincent Grégoire, Caron F., Sabatier D., Blanc L., 2012. LiDAR shows that higher forests have more slender trees. *Bois et Forêts des Tropiques*, (314), pp. 51-56.
- Vincent G., Sabatier D., Blanc L., Chave J., Weissenbacher E., Pélissier R., Fonty E., Molino J.-F., Coutron P., 2012. Accuracy of small footprint airborne LiDAR in its predictions of tropical moist forest stand structure. *Remote Sensing of Environment*, 125 , pp. 23-33.
- Vincent G., Sabatier D., Rutishauser E., 2014. Revisiting a universal airborne light detection and ranging approach for tropical forest carbon mapping: scaling-up from tree to stand to landscape. *Oecologia*, 175 (2), pp. 439-443.
- Vincent G., Weissenbacher E., Sabatier D., Blanc L., Proisy C., Coutron P., 2010. Détection des variations de structure de peuplements en forêt dense tropicale humide par Lidar aéroporté. *Revue Française de Photogrammetrie et de Teledetection*.



## 7. LISTE DES ABREVIATIONS

- AGB : *Above-Ground Biomass*
- AGV : *Above-Ground Volume*
- AIC : *Akaike Information Criterion*
- Cirad : Centre international pour la recherche en agronomie et pour le développement
- CNRS : Centre National de la Recherche Scientifique
- DBH : *Diameter at Breast Height*
- GPS : *Global Positioning System*
- Lidar : *Light detection and ranging*
- MCH : *Mean Canopy Height*
- MNC : Modèle Numérique de Canopée
- MNS : Modèle Numérique de Surface
- MNT : Modèle Numérique de Terrain
- MSE : *Mean Squared Error*
- NDP : Nuage De Points
- Radar : *Radio detection and ranging*
- RF : *Random Forest*
- RMSE : *Root Mean Squared Error*
- SIG : Système d'Information Géographique
- UMR AMAP : Unité Mixte de Recherche en botanique et Modélisation de l'Architecture des Plantes et des végétations





## 8. TABLE DES TABLEAUX

Tableau 1 : Dates des relevés dendrométriques et lidar par site .....	12
Tableau 2 : Conditions d'acquisitions des données lidar .....	13
Tableau 3 : Qualité de régression des trois fonctions testées pour le jeu de données BRIDGE .....	20
Tableau 4 : Estimation de trois façons différentes du biovolume total des 2207 arbres de Paracou et Montagne Tortue .....	23
Tableau 5: Nombre de palmiers par site .....	23
Tableau 6 : Biomasse totale par site en tonnes avec prise en compte (Chave/Goodman) et sans prise en compte (Chave) de l'allométrie spécifique des palmiers .....	24
Tableau 7: Coefficients de régression de l'AGV moyen en fonction de Hp...	26
Tableau 8 : Estimation du biovolume à l'aide de la méthode composite et comparaison avec la prédiction directe .....	26
Tableau 9 : Dimensions des placettes et écart-type de la biomasse mesurée à Paracou .....	29

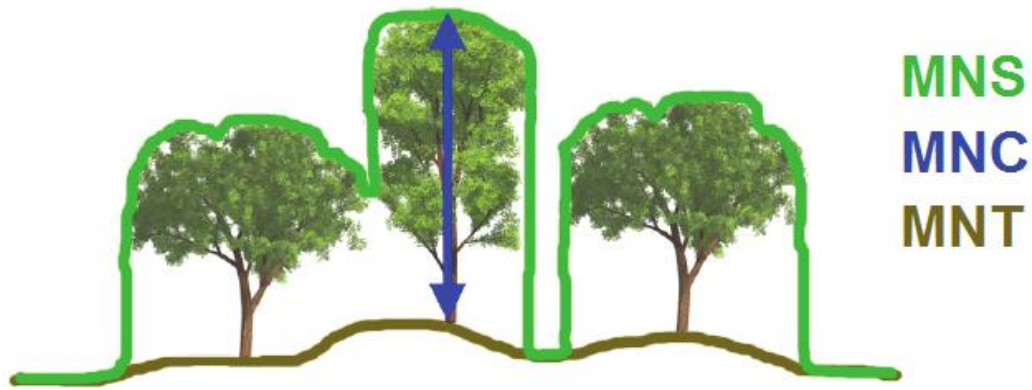


## 9. TABLE DES FIGURES

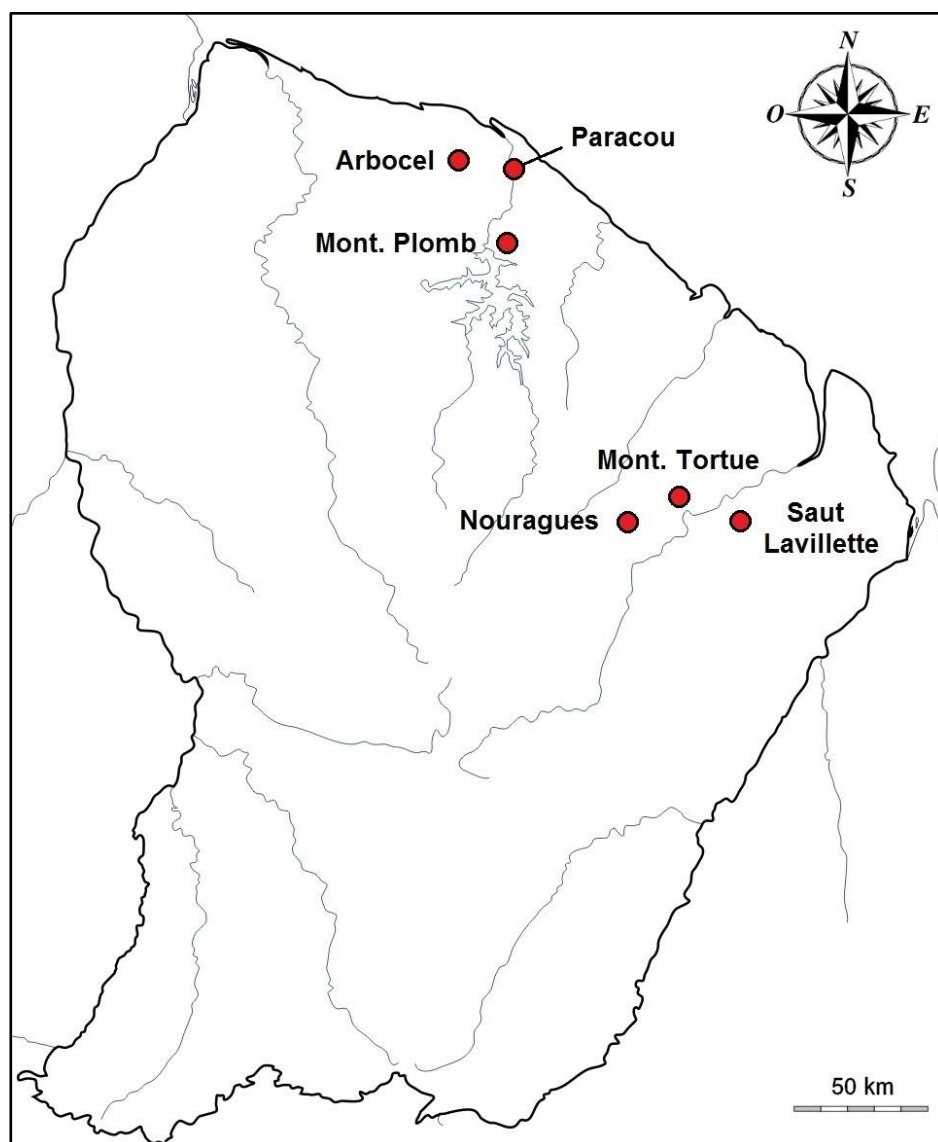
Figure 1 : Nuage de points lidar de 25 hectares à Paracou visualisé avec R (colorisation en fonction de l'altitude en m) (Source : IRD, UMR AMAP) .....	8
Figure 2 : Nuage des observations en fonction des prédictions .....	17
Figure 3 : Rapport hauteur-diamètre des arbres du jeu de données BRIDGE et fonctions de régression testées .....	20
Figure 4 : Résidus moyens de hauteurs (prédictions–observations) en fonction de $H_p$ par placette du jeu de données BRIDGE et droite de régression associée ( $R^2 = 0,63$ ) .....	21
Figure 5 : Estimation du biovolume moyen par arbre de chacune des 85 placettes de 1 ha de Paracou en fonction du diamètre quadratique moyen au-dessus de 10 cm et de $H_p$ .....	25
Figure 6 : Répartitions des neuf placettes de 1 ha sur la dalle de 25 ha de Paracou .....	27
Figure 7 : Décimation du NDP de la placette 5. De gauche à droite, le taux de densité est respectivement de 100%, 20%, et 5% (colorisation en fonction de l'altitude en m). .....	28
Figure 8 : Effet de la densité de tirs sur le MNC et le NDP (données en Annexe 17) .....	28
Figure 9 : RMSE de la prédiction de $AGB(D, H_{(D,H_p)})$ en fonction de l'échelle spatiale et de la fonction de régression utilisée (données en Annexe 18)....	30
Figure 10 : Erreur relative de prédiction par site du biovolume et de la biomasse. En abscisse, les noms des sites ont été compactés : PAR (Paracou), NOU (Nouragues), MTO (Montagne Tortue), MPB (Montagne Plomb), SLV (Saut Lavillette) et ARB (Arbocel). .....	32
Figure 11 : Répartition de la biomasse sur Paracou (en t/ha).....	33
Figure 12 : Hauteur prédite des arbres selon la régression locale (Équation 3) et la loi allométrique de Chave et al. (2005). .....	36

# ANNEXES

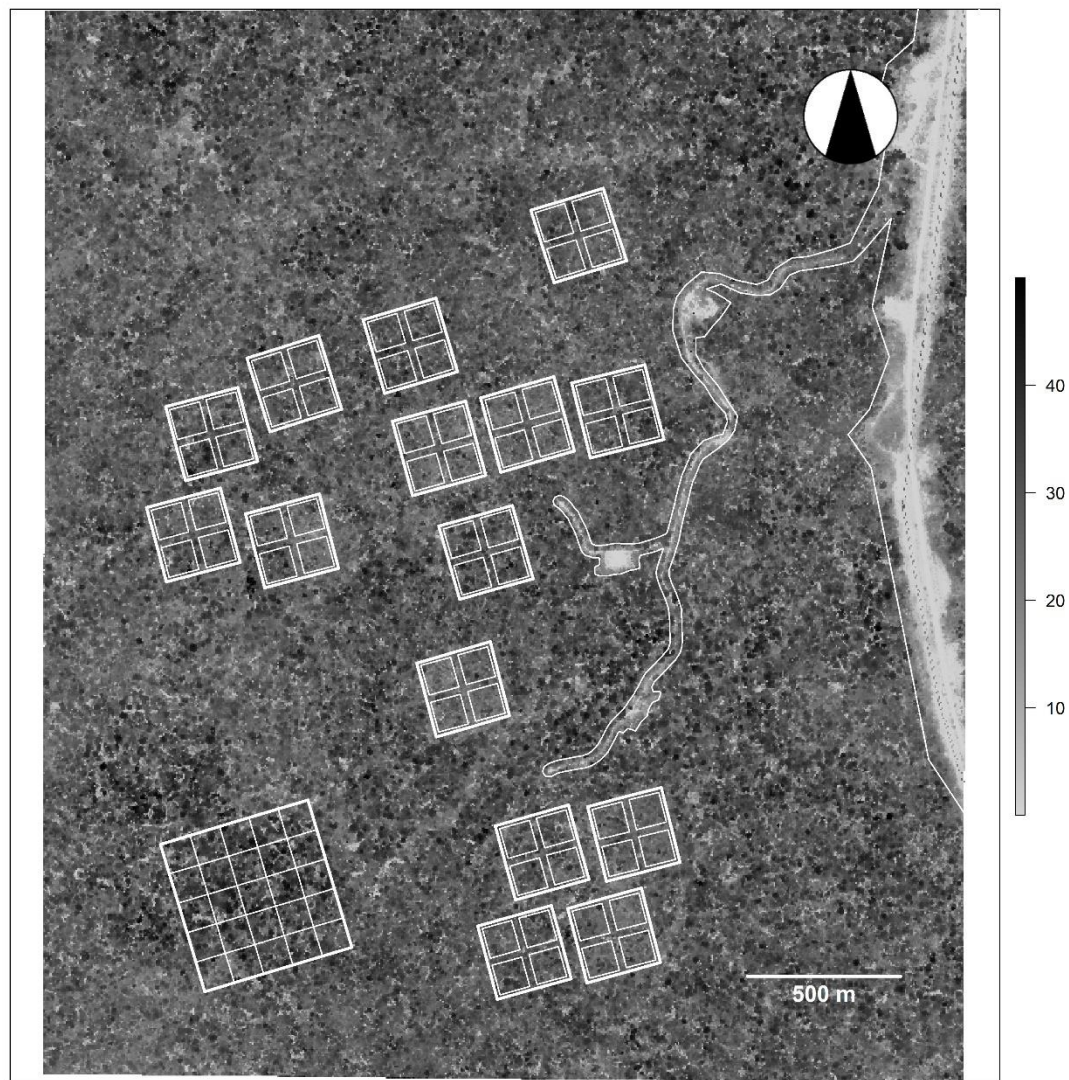
## ANNEXE 1 : DISTINCTION ENTRE MNT, MNS ET MNC



## ANNEXE 2 : EMBLACEMENT DES SITES D'ETUDE EN GUYANE FRANÇAISE



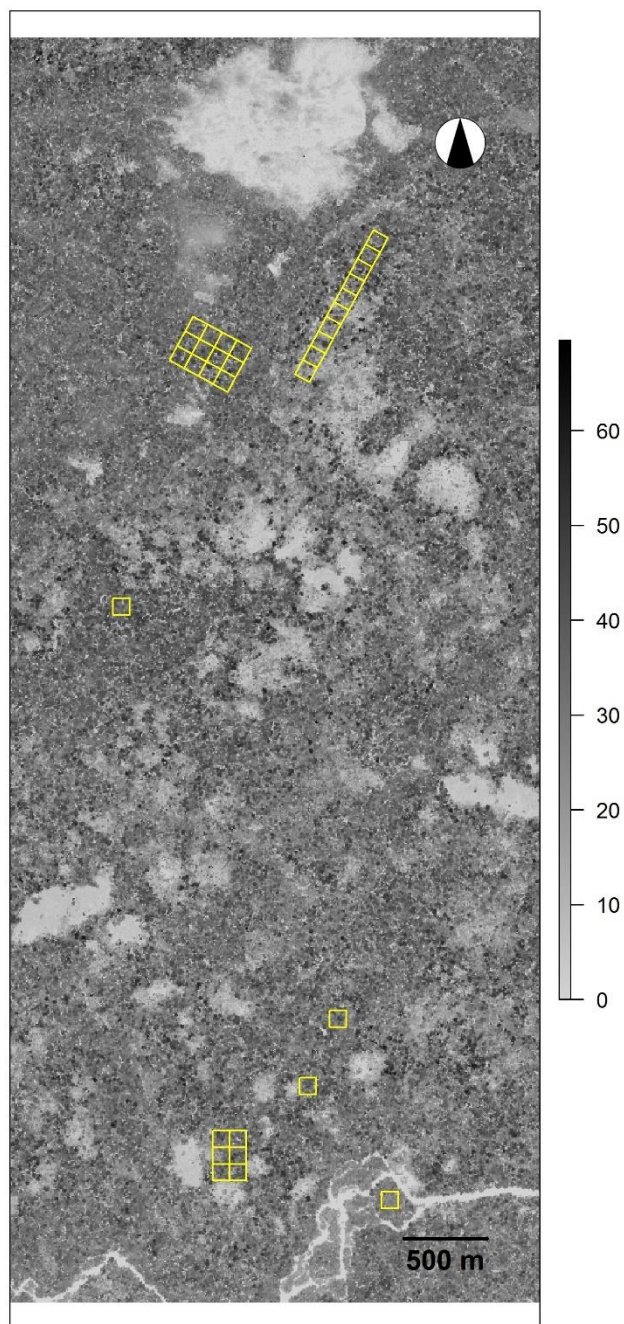
## ANNEXE 3 : SITE DE PARACOU



MNC de Paracou (hauteur en m)

En trait épais, les parcelles d'inventaire Guyafor et en trait fin, les 85 placettes de 1 ha ayant servi à l'étude. Autour de la route et de la piste (à l'est) a été définie une zone d'exclusion où l'estimation de biomasse à partir du lidar n'est pas possible.

## ANNEXE 4 : SITE DES NOURAGUES

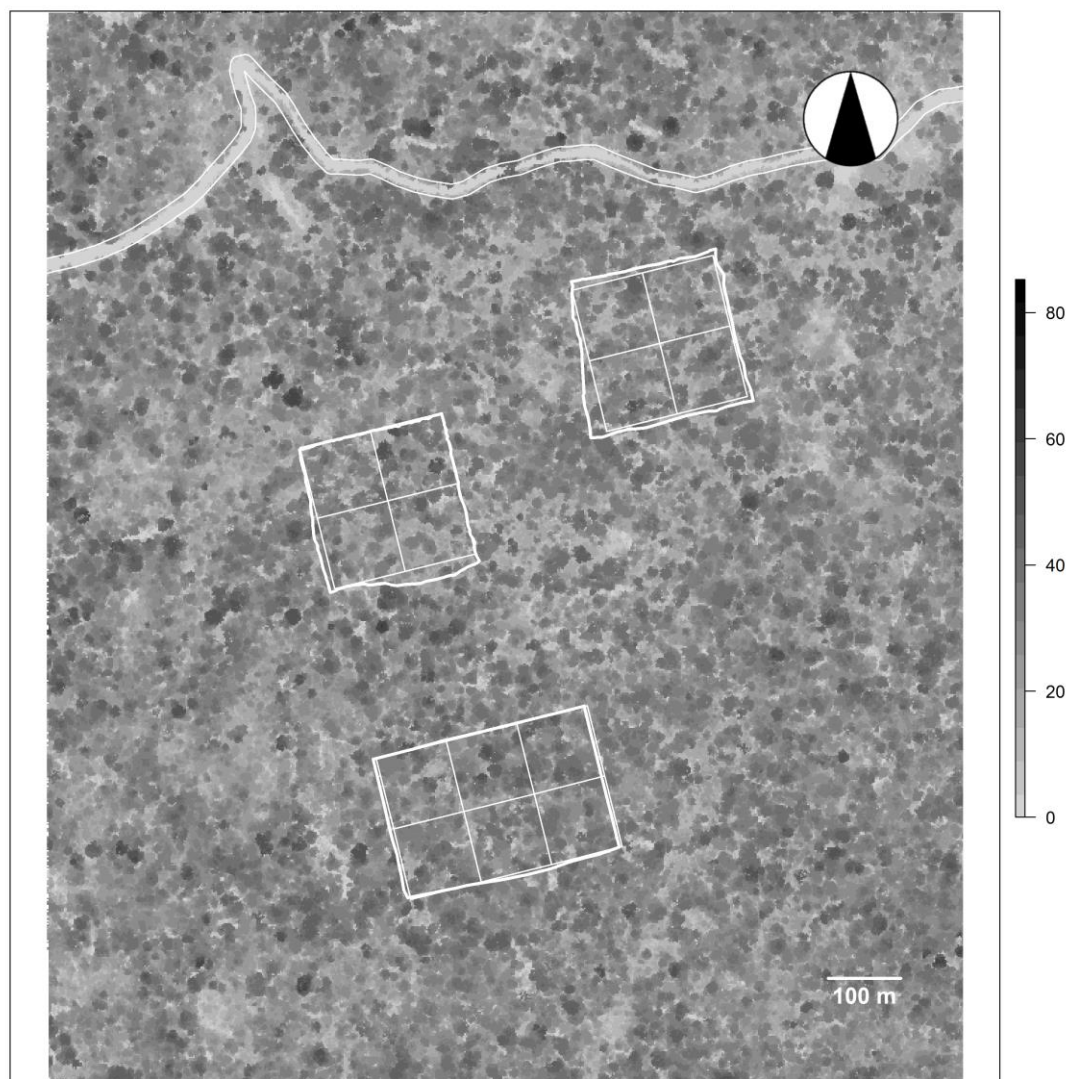


MNC des Nouragues (hauteur en m)

En jaune : les 32 placettes d'inventaire de 1 ha.



## ANNEXE 5 : SITE DE MONTAGNE TORTUE

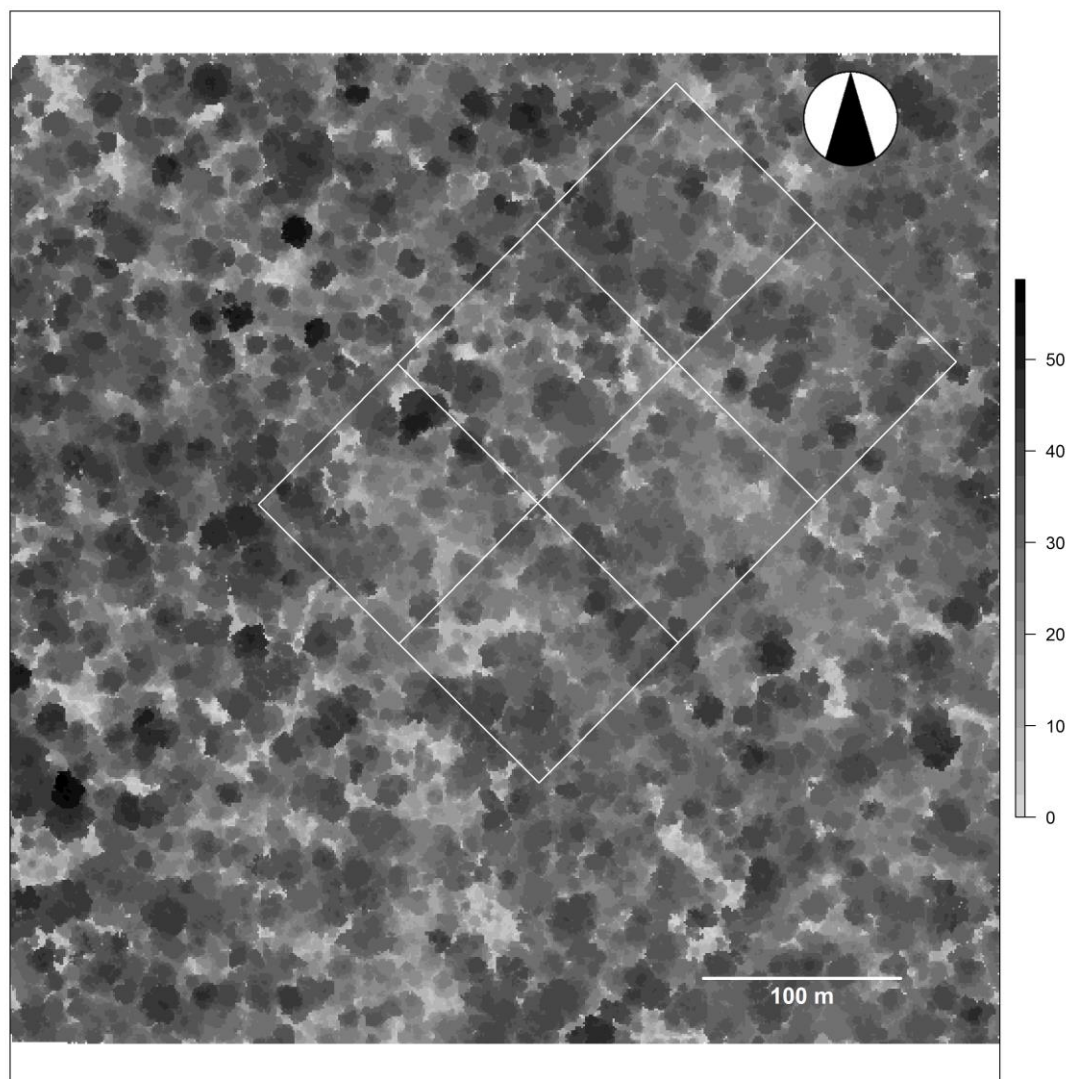


MNC de Montagne Tortue (hauteur en m)

En trait épais, les parcelles d'inventaire Guyafor et en trait fin, les 14 placettes de 1 ha ayant servi à l'étude.

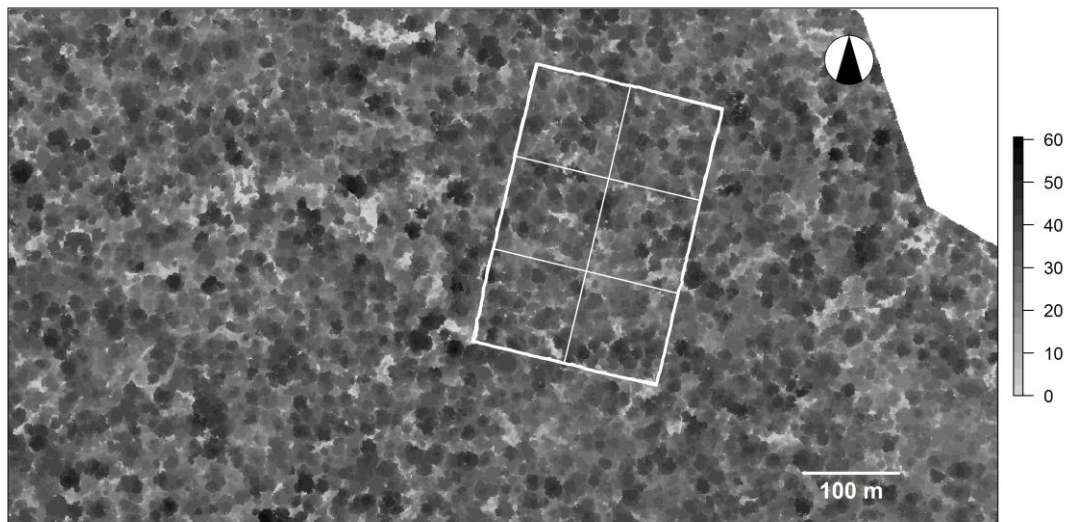


## ANNEXE 6 : SITE DE MONTAGNE PLOMB



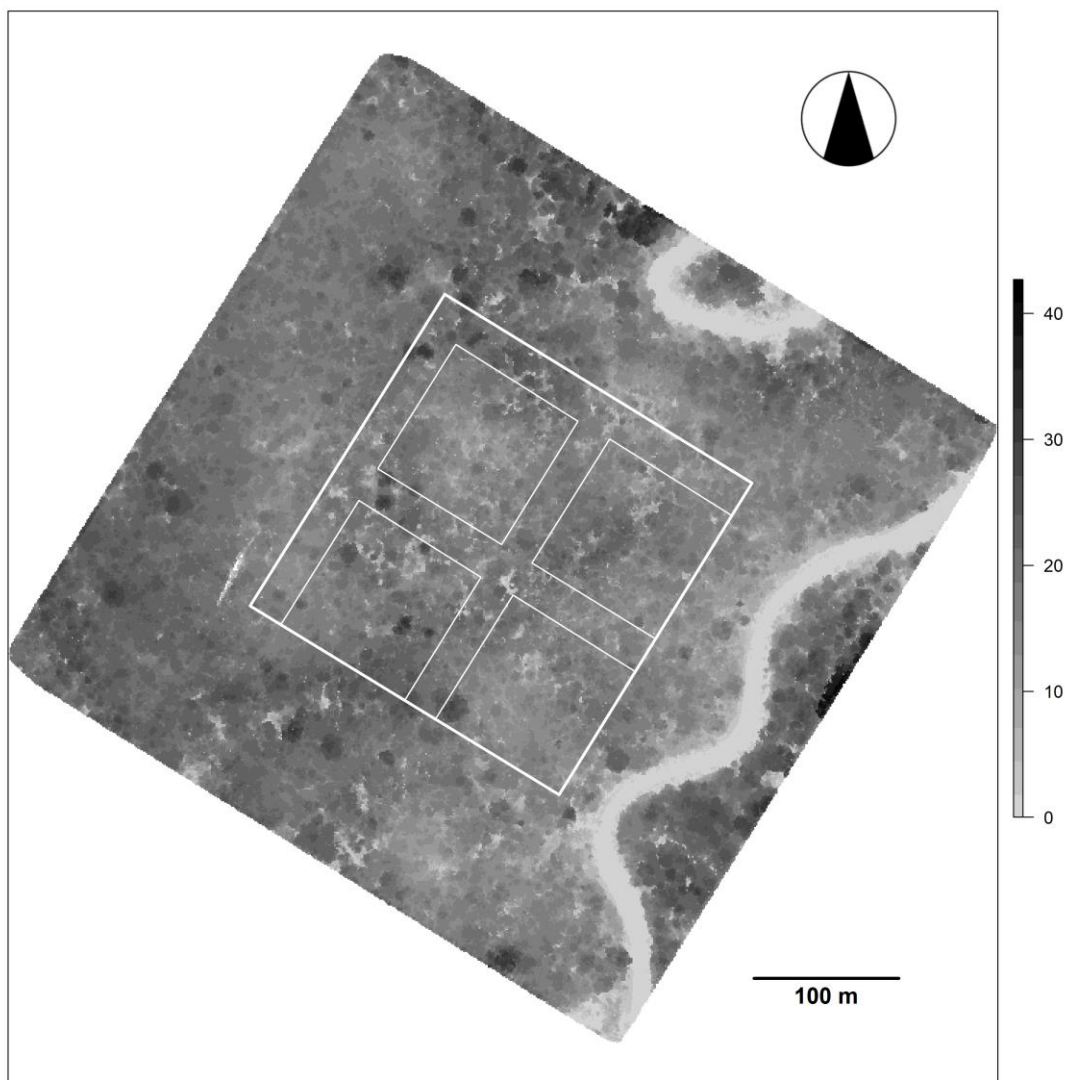
MNC de Montagne Plomb (hauteur en m)

## ANNEXE 7 : SITE DE SAUT LAVILLETTE



MNC de Saut Lavillette (hauteur en m)

## ANNEXE 8 : SITE D'ARBOCEL



MNC d'Arbocel (hauteur en m)

En trait épais, la parcelle d'inventaire Guyafor et en trait fin, les 4 placettes de 1 ha ayant servi à l'étude.

## ANNEXE 9 : COMPOSITION FLORISTIQUE

Famille	Fréquence (%)
Lecythidaceae	16,7
Fabaceae	15,5
Chrysobalanaceae	10,1
Sapotaceae	5,7
Annonaceae	5,5
Clusiaceae	4,3
Melastomataceae	4,1
Euphorbiaceae	3,9
Myristicaceae	3,2
Malvaceae	3,2

Familles dominantes à Paracou

Famille	Fréquence (%)
Fabaceae	12,4
Lecythidaceae	11,6
Sapotaceae	9,3
Burseraceae	7,0
Malvaceae	6,1
Chrysobalanaceae	5,8
Meliaceae	3,7
Moraceae	3,0
Annonaceae	2,7
Lauraceae	2,7

Familles dominantes aux Nouragues

Famille	Fréquence (%)
Fabaceae	15,5
Sapotaceae	10,1
Chrysobalanaceae	6,5
Burseraceae	6,3
Urticaceae	6,2
Lauraceae	5,4
Lecythidaceae	5,1
Malvaceae	4,4
Apocynaceae	3,2
Anacardiaceae	2,8

Familles dominantes à Mont. Tortue

Famille	Fréquence (%)
Lecythidaceae	25,3
Fabaceae	17,4
Sapotaceae	6,0
Annonaceae	4,2
Malvaceae	2,8
Euphorbiaceae	2,7
Melastomataceae	2,5
Chrysobalanaceae	2,4
Putranjivaceae	2,0
Violaceae	1,9

Familles dominantes à Mont. Plomb

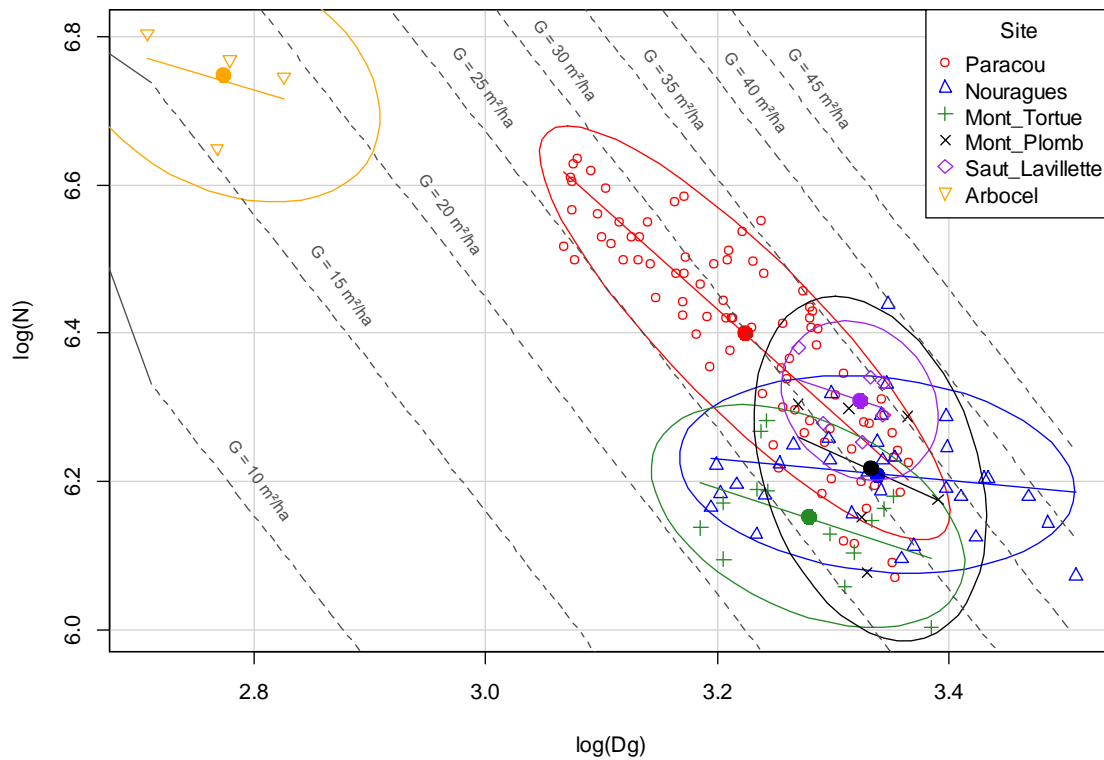
Famille	Fréquence (%)
Lecythidaceae	13,4
Fabaceae	13,0
Sapotaceae	12,3
Chrysobalanaceae	11,7
Burseraceae	9,6
Annonaceae	8,3
Arecaceae	4,2
Lauraceae	3,4
Malvaceae	2,2
Apocynaceae	2,0

Familles dominantes à Saut Lavillette

Famille	Fréquence (%)
Annonaceae	21,3
Flacourtiaceae	18,4
Celastraceae	16,4
Melastomataceae	7,8
Mimosaceae	5,4
Clusiaceae	4,6
Moraceae	2,2
Malpighiaceae	2,2
Bignoniaceae	1,9
Boraginaceae	1,9

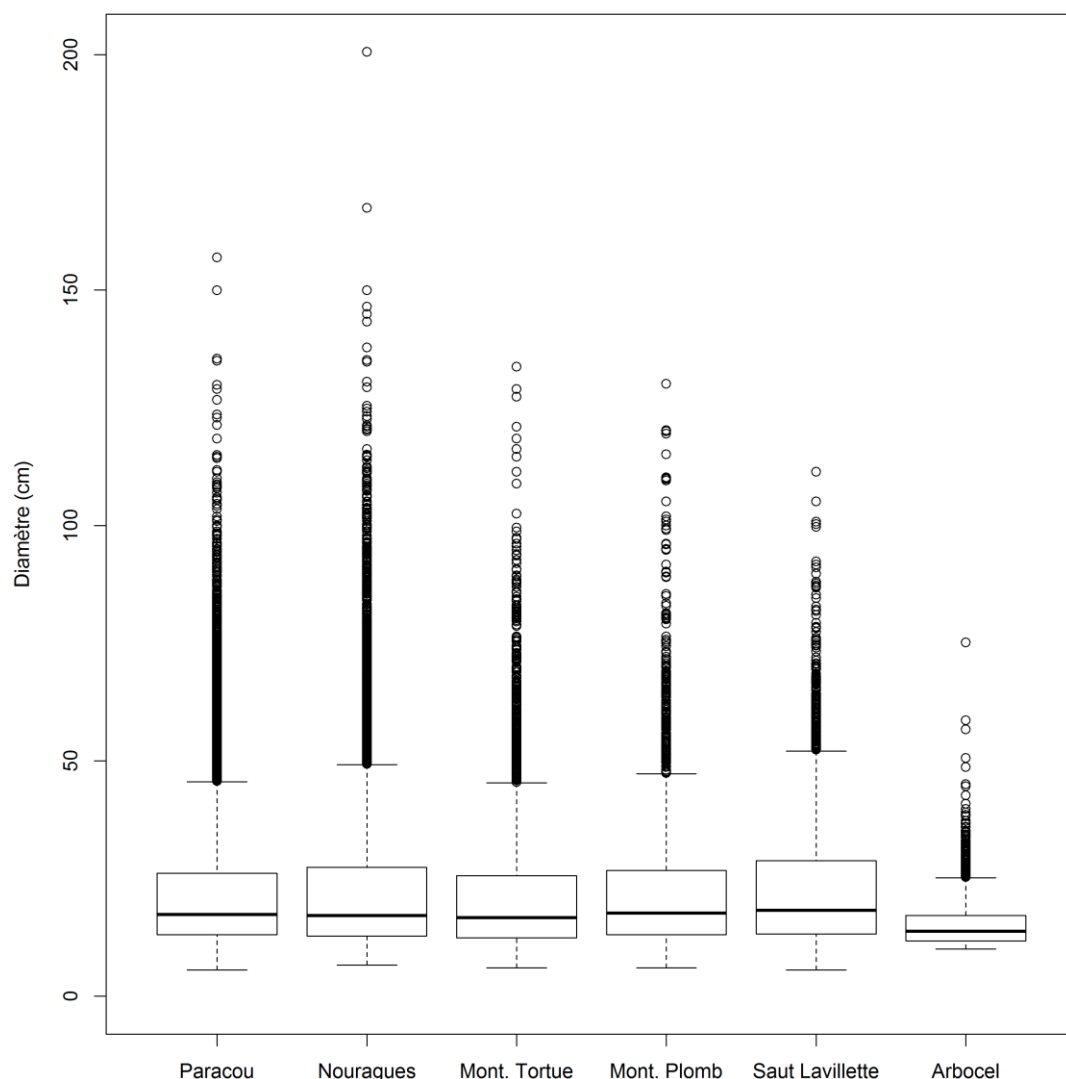
Familles dominantes à Arbocel

## ANNEXE 10 : STRUCTURE DES PEUPELEMENTS



La figure ci-dessus montre comment le peuplement de chaque site est structuré. A surface terrière égale ( $G$ ), elle permet d'observer s'il s'agit d'un ensemble de gros arbres espacés les uns par rapport aux autres ( $\log(N)$  faible,  $\log(Dg)$  élevé) ou d'un ensemble de tiges de petits diamètres densément agencées ( $\log(N)$  élevé,  $\log(Dg)$  faible). Le site d'Arbocel, en cours de régénération, est clairement dans le second cas.

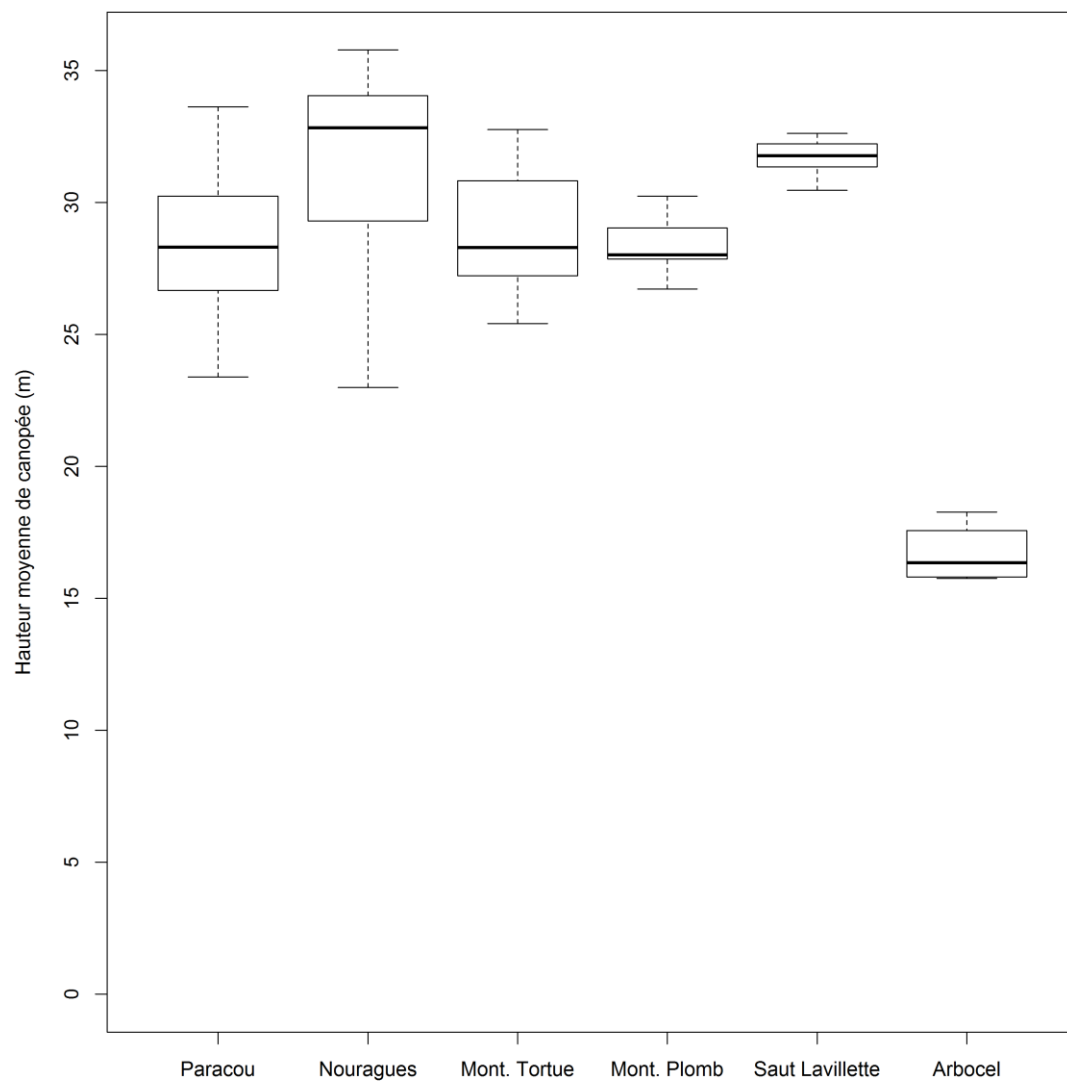
## ANNEXE 11 : DIAMETRES DES ARBRES PAR SITE



Pour chaque boîte à moustache, les côtés inférieur et supérieur du rectangle représentent respectivement les premier et le troisième quartile  $Q1$  et  $Q3$ , alors que le segment intermédiaire représente la médiane.

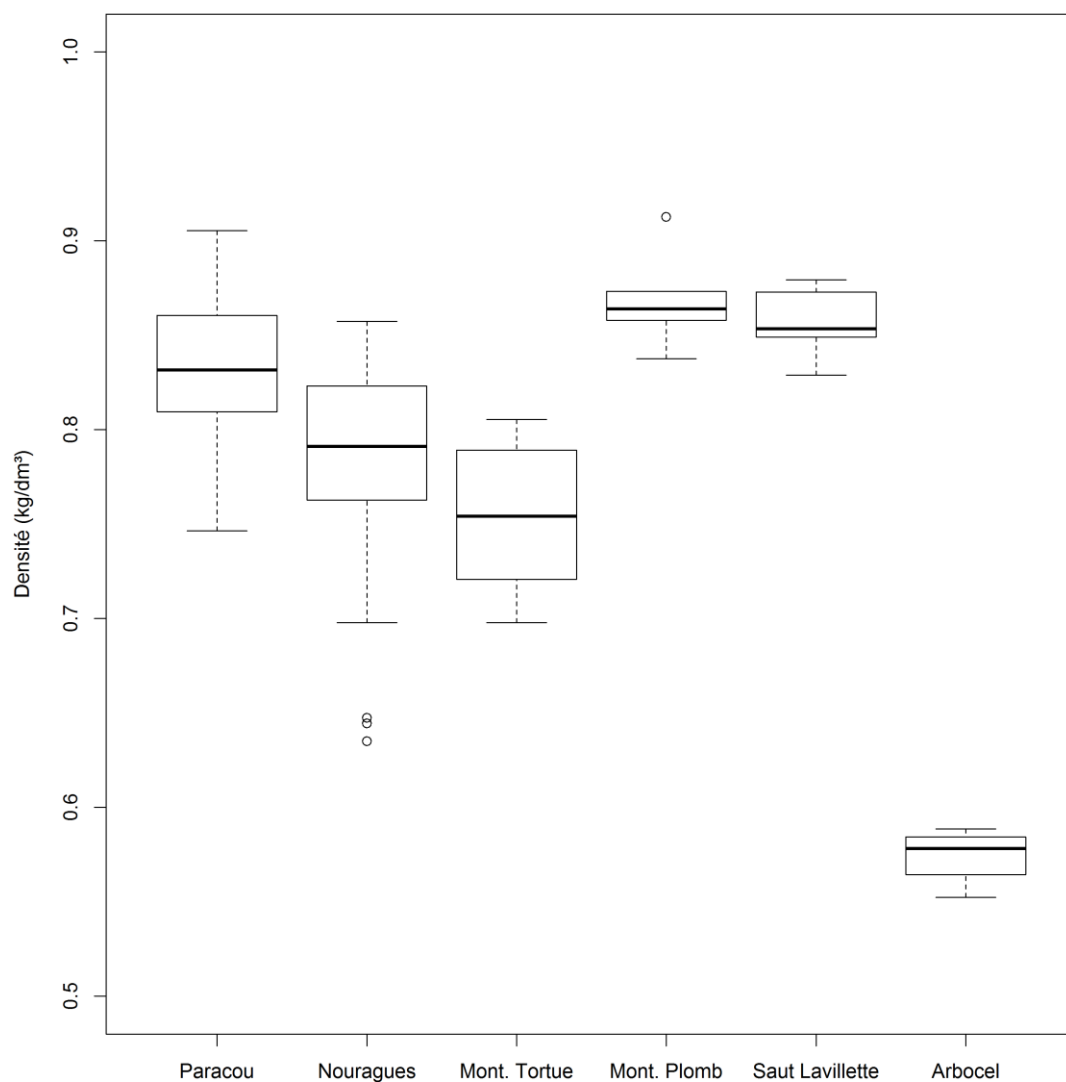
L'extrémité de la moustache inférieure est égale à  $\max[\min(x), Q1 - 1,5 \times (Q3 - Q1)]$  et celle de la moustache supérieure à  $\min[\max(x), Q3 + 1,5 \times (Q3 - Q1)]$  où  $x$  est l'ensemble des valeurs de la distribution.

## ANNEXE 12 : HAUTEUR MOYENNE DE CANOPEE PAR SITE





## ANNEXE 13 : DENSITE DE BOIS PAR SITE

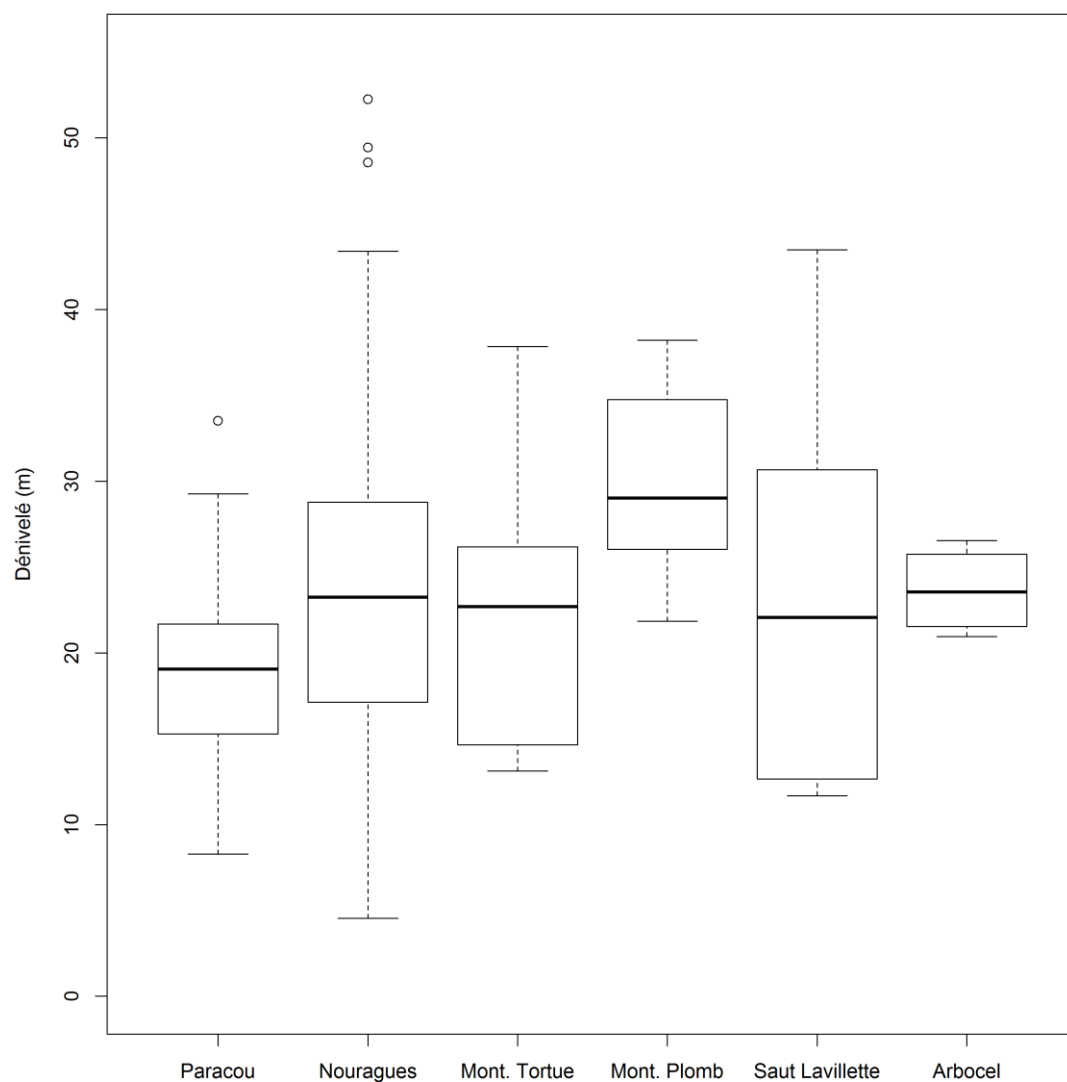


Il s'agit de la distribution de densité de bois par placette de 1 ha pondérée par le biovolume calculé selon l'Équation 2 :

$$\rho_{\text{site}} = \frac{1}{AGV_{\text{site}}} \sum_i \rho_i \cdot AGV_i = \frac{AGB_{\text{site}}}{AGV_{\text{site}}}$$

Le facteur i représente chaque arbre pris individuellement.

## ANNEXE 14 : DENIVELE PAR SITE



Il s'agit de la différence pour chaque placette entre le point le plus haut et le point de plus bas calculé à partir du MNT du site.

## ANNEXE 15 : CORRECTION DES MESURES DE DIAMETRE

Seules les parcelles ayant été suivies durant un minimum de 3 campagnes sont prises en compte. Les codes suivants correspondent aux corrections appliquées aux mesures de circonférence des arbres.

### Code 0

La mesure n'a subi aucune correction.

### Si un arbre montre une croissance annuelle en diamètre supérieure à 5 cm :

#### Code 1

Cette croissance ne concerne que quelques points suivis d'un retour à une croissance normale. Ces points sont réalignés par régression linéaire.

#### Code 2

Cette croissance concerne plusieurs points sans retour à une croissance normale. La série de points la plus abondante (ou la plus récente si égalité) sert de base pour réaligner l'autre série par ajout/retrait de la différence de DBH, corrigée par la croissance attendue (calculée comme la croissance moyenne des points environnants).

### Si un arbre montre une décroissance annuelle supérieure à 2 cm :

#### Code 3

Cette décroissance ne concerne que quelques points suivis d'un retour à une croissance normale. Ces points sont réalignés par régression linéaire.

#### Code 4

Cette décroissance concerne plusieurs points sans retour à une croissance normale. La série de points la plus abondante (ou la plus récente si égalité) sert de base pour réaligner l'autre série par ajout/retrait de la différence de DBH, corrigée par la croissance attendue (calculée comme la croissance moyenne des points environnants).

### Code 5

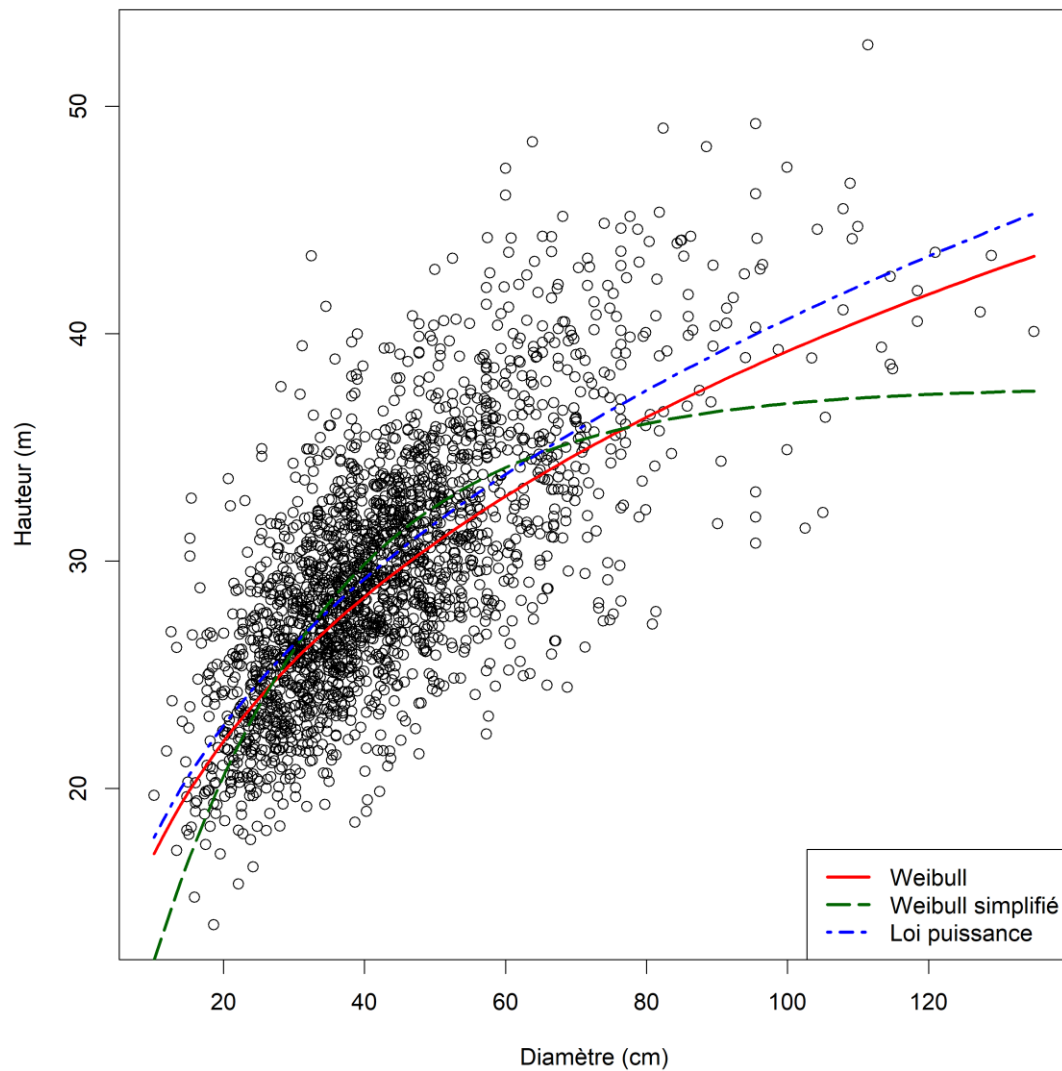
Cette correction concerne les bois cathédrale (Bugu Bugu) pour lesquels Nogueira (2006) propose de multiplier le DBH par 0,53 pour évaluer le diamètre si l'arbre était rond. Cette correction est appliquée sur les DBH corrigés, le code\_corr apparaît donc, le cas échéant, comme « x + 5 », x étant le code correspondant au type de correction éventuellement subie par la mesure concernée.

### Code 6

Cette correction concerne les très gros arbres mal conformés et donc non mesurés. Dans Guyafor, ils apparaissent avec une circonférence constante de 888 cm. Ces circonférences sont remplacées par des circonférences maximales empiriques :

- $0,80 \times \text{DBH} \times \pi$  pour les kimboto, maho cigare, diagidia, buguni, yayamadou marécage, supun udu ;
- $0,50 \times \text{DBH} \times \pi$  pour les weti udu, lebi koko, yayamadou montagne, wapa rivière, weti koko ;
- $0,55 \times \text{circ}$  pour les patawa.

## ANNEXE 16 : JEU DE DONNEES DE PARACOU ET MONTAGNE TORTUE



## ANNEXE 17 : EFFET DE LA DENSITE DE TIR SUR LES PREDICTEURS

Tirs conservés	Hauteur moyenne du MNC (m)					Hauteur médiane du MNC (m)				
	100%	50%	20%	10%	5%	100%	50%	20%	10%	5%
Placette 1	26,4	26,2	25,8	25,2	24,2	26,6	26,4	26,1	25,7	24,8
Placette 2	26,9	26,7	26,3	25,7	24,6	26,9	26,7	26,5	26,0	25,1
Placette 3	25,4	25,3	24,9	24,4	23,6	25,2	25,0	24,7	24,4	23,8
Placette 4	28,1	27,9	27,5	27,0	26,1	28,2	28,0	27,7	27,3	26,5
Placette 5	29,1	28,9	28,5	27,9	26,9	29,2	29,0	28,6	28,1	27,2
Placette 6	26,1	25,9	25,6	25,1	24,2	25,4	25,2	24,8	24,4	23,7
Placette 7	30,0	29,8	29,4	28,7	27,6	30,9	30,7	30,4	29,8	28,9
Placette 8	27,5	27,3	26,8	26,3	25,2	28,6	28,4	28,1	27,7	26,7
Placette 9	29,0	28,8	28,5	28,0	27,1	29,0	28,9	28,5	28,2	27,6
Moyenne	27,6	27,4	27,0	26,5	25,5	27,8	27,6	27,3	26,8	26,0
Ecart-type	1,5	1,4	1,4	1,4	1,4	1,8	1,8	1,8	1,7	1,7

Tirs conservés	Hauteur moyenne du NDP (m)					Hauteur médiane du NDP (m)				
	100%	50%	20%	10%	5%	100%	50%	20%	10%	5%
Placette 1	21,3	21,3	21,3	21,3	21,3	22,2	22,2	22,2	22,2	22,2
Placette 2	22,2	22,2	22,2	22,2	22,1	23,2	23,2	23,2	23,2	23,1
Placette 3	21,4	21,4	21,4	21,4	21,5	22,0	22,0	22,0	22,0	22,1
Placette 4	23,2	23,2	23,2	23,2	23,2	24,4	24,4	24,4	24,4	24,4
Placette 5	23,8	23,8	23,9	23,9	23,9	25,0	25,0	25,0	25,0	25,0
Placette 6	21,6	21,6	21,6	21,7	21,6	21,9	21,9	21,9	21,9	21,9
Placette 7	24,5	24,5	24,5	24,6	24,6	26,4	26,4	26,3	26,4	26,4
Placette 8	22,0	22,0	22,0	22,1	22,1	24,0	24,0	24,0	24,0	24,0
Placette 9	24,8	24,8	24,8	24,8	24,8	26,0	26,0	26,0	26,0	26,0
Moyenne	22,7	22,8	22,8	22,8	22,8	23,9	23,9	23,9	23,9	23,9
Ecart-type	1,3	1,3	1,3	1,3	1,3	1,6	1,6	1,6	1,6	1,6

## ANNEXE 18 : ECHELLE SPATIALE DE PREDICTION

L = 50 m A = 0,25 ha		plsr				autopls				RandomForest			
		R <sup>2</sup>	CCL	RMSE	RMSE / $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE / $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE / $\bar{x}$
AGV (m <sup>3</sup> /ha)	$\sum \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,53	1,00	76,9	0,14	0,53	1,00	77,4	0,14	0,46	0,96	82,6	0,15
	$\sum 0,0509 \cdot D^2 \cdot H_D$	0,51	1,00	70,3	0,13	0,47	1,00	72,5	0,14	0,43	0,95	75,2	0,14
	$\sum 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,67	1,00	70,4	0,13	0,64	1,00	72,9	0,13	0,61	0,98	76,0	0,14
AGB (t/ha)	$\sum \rho \cdot \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,50	1,00	69,5	0,15	0,49	1,00	70,0	0,15	0,44	0,98	73,2	0,16
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_D$	0,48	1,00	64,1	0,14	0,47	1,00	64,6	0,14	0,42	0,98	67,7	0,15
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,62	1,00	65,0	0,14	0,62	1,00	65,5	0,14	0,59	1,00	68,2	0,15

L = 100 m A = 1,00 ha		plsr				autopls				RandomForest			
		R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$
AGV (m <sup>3</sup> /ha)	$\sum \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,63	1,00	47,1	0,08	0,60	1,00	48,9	0,09	0,42	0,88	58,8	0,11
	$\sum 0,0509 \cdot D^2 \cdot H_D$	0,59	1,00	42,3	0,08	0,58	1,00	42,9	0,08	0,37	0,86	52,1	0,10
	$\sum 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,76	1,00	43,1	0,08	0,77	1,00	42,6	0,08	0,61	0,94	55,6	0,10
AGB (t/ha)	$\sum \rho \cdot \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,52	1,00	46,1	0,10	0,59	1,00	42,8	0,09	0,37	0,94	52,8	0,11
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_D$	0,48	1,00	42,1	0,09	0,49	1,00	41,8	0,09	0,32	0,89	48,1	0,11
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,71	1,00	40,9	0,09	0,71	1,00	40,8	0,09	0,54	0,96	51,1	0,11



L = 125 m A = 1,56 ha		pls				autopls				RandomForest			
		R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$
AGV (m <sup>3</sup> /ha)	$\sum \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,66	1,00	43,3	0,08	0,69	1,00	41,2	0,08	0,60	0,94	46,8	0,09
	$\sum 0,0509 \cdot D^2 \cdot H_D$	0,61	1,00	38,8	0,07	0,63	1,00	37,9	0,07	0,55	0,93	41,9	0,08
	$\sum 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,81	1,00	37,5	0,07	0,84	1,00	34,5	0,06	0,75	1,00	42,7	0,08
AGB (t/ha)	$\sum \rho \cdot \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,65	1,00	39,1	0,09	0,63	1,00	40,3	0,09	0,55	1,00	44,1	0,10
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_D$	0,62	1,00	35,4	0,08	0,60	1,00	36,5	0,08	0,51	0,96	40,5	0,09
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,78	1,00	34,8	0,08	0,77	1,00	36,0	0,08	0,68	1,01	42,5	0,09

L = 200 m A = 4,00 ha		plsr				autopls				RandomForest			
		R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$
AGV (m <sup>3</sup> /ha)	$\sum \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,86	1,00	24,5	0,04	0,87	1,00	23,4	0,04	0,74	1,22	33,6	0,06
	$\sum 0,0509 \cdot D^2 \cdot H_D$	0,85	1,00	20,8	0,04	0,86	1,00	20,0	0,04	0,68	1,22	30,3	0,06
	$\sum 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,94	1,00	19,3	0,03	0,94	1,00	19,1	0,03	0,83	1,27	32,9	0,06
AGB (t/ha)	$\sum \rho \cdot \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,95	1,00	12,8	0,03	0,86	1,00	20,8	0,05	0,69	1,27	30,9	0,07
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_D$	0,94	1,00	11,6	0,03	0,85	1,00	18,0	0,04	0,69	1,22	26,1	0,06
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,97	1,00	11,7	0,03	0,94	1,00	15,9	0,03	0,79	1,18	30,0	0,06

L = 250 m A = 6,25 ha		pls				autopls				RandomForest			
		R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$	R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$
AGV (m <sup>3</sup> /ha)	$\sum \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,92	1,00	18,6	0,03	0,92	1,00	17,8	0,03	0,74	1,19	33,0	0,06
	$\sum 0,0509 \cdot D^2 \cdot H_D$	0,91	1,00	16,2	0,03	0,94	1,00	12,5	0,02	0,67	1,15	30,4	0,06
	$\sum 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,95	1,00	16,4	0,03	0,97	1,00	14,0	0,03	0,82	1,27	32,6	0,06
AGB (t/ha)	$\sum \rho \cdot \exp[-1,499 + 2,148 \cdot \ln(D) + \dots]$	0,95	1,00	13,3	0,03	0,93	1,00	15,0	0,03	0,66	1,27	33,4	0,07
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_D$	0,94	1,00	11,7	0,03	0,93	1,00	12,6	0,03	0,64	1,19	29,1	0,06
	$\sum \rho \cdot 0,0509 \cdot D^2 \cdot H_{(D,Hp)}$	0,97	1,00	11,3	0,02	0,96	1,00	12,6	0,03	0,78	1,30	31,1	0,07

## ANNEXE 19 : PARAMETRAGE DES MODELES

		R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$
Variable prédite	AGV(D-H <sub>(D,Hp)</sub> )				
Origine des prédictors	NDP				
Modèle	Loi puissance	0,70	1,02	78	0,13
Domaine d'apprentissage	Global (147 pl.)				
Domaine d'application	Global (147 pl.)				
Variable prédite	AGV(D-H <sub>(D,Hp)</sub> )				
Origine des prédictors	MNC				
Modèle	Loi puissance	0,75	1,02	70	0,12
Domaine d'apprentissage	Global (147 pl.)				
Domaine d'application	Global (147 pl.)				
Variable prédite	AGV(D-H <sub>(D,Hp)</sub> )				
Origine des prédictors	NDP				
Modèle	autopl	0,74	1,00	73	0,12
Domaine d'apprentissage	Global (147 pl.)				
Domaine d'application	Global (147 pl.)				
Variable prédite	AGV(D-H <sub>(D,Hp)</sub> )				
Origine des prédictors	MNC				
Modèle	autopl	0,76	1,00	69	0,12
Domaine d'apprentissage	Global (147 pl.)				
Domaine d'application	Global (147 pl.)				
Variable prédite	AGB(D-H <sub>(D,Hp)</sub> )				
Origine des prédictors	MNC				
Modèle	Loi puissance	0,67	1,04	69	0,15
Domaine d'apprentissage	Global (147 pl.)				
Domaine d'application	Global (147 pl.)				
Variable prédite	AGB(D-H <sub>(D,Hp)</sub> )				
Origine des prédictors	MNC				
Modèle	autopl	0,71	1,00	65	0,14
Domaine d'apprentissage	Global (147 pl.)				
Domaine d'application	Global (147 pl.)				

		R <sup>2</sup>	CCL	RMSE	RMSE/ $\bar{x}$
Variable prédite	AGV(D-H <sub>(D,Hp)</sub> )				
Origine des prédicteurs	NDP				
Modèle	autopls	0,70	1,00	51	0,09
Domaine d'apprentissage	Paracou (85 pl.)				
Domaine d'application	Paracou (85 pl.)				
Variable prédite	AGV(D-H <sub>(D,Hp)</sub> )				
Origine des prédicteurs	MNC				
Modèle	autopls	0,75	1,00	47	0,08
Domaine d'apprentissage	Paracou (85 pl.)				
Domaine d'application	Paracou (85 pl.)				
Variable prédite	AGV(D-H <sub>(D,Hp)</sub> )				
Origine des prédicteurs	MNC				
Modèle	autopls	0,73	0,92	69	0,12
Domaine d'apprentissage	Global (147 pl.)				
Domaine d'application	Paracou (85 pl.)				