THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'INSTITUT NATIONAL DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT – AGROPARISTECH

N° : 2024 AGPT 0010

En Sciences de la Terre

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau – n°584
Portée par l'Université de Montpellier

Unité mixte de recherche TETIS – Territoire Environnement Télédétection Information Spatiale

Potential and Challenges of Spaceborne LiDAR for Multi-Source Canopy Height Mapping in Tropical Forests

Potentiel et défis du LiDAR spatial pour la cartographie multisource de la hauteur de la canopée dans les forêts tropicales

Présentée par Kamel LAHSSINI
Le 2 décembre 2024

Sous la direction de Nicolas BAGHDADI et Guerric LE MAIRE

Devant le jury composé de

Frédéric FRAPPART, Directeur de recherche, INRAE (centre Nouvelle-Aquitaine-Bordeaux)	Rapporteur
Mehrez ZRIBI, Directeur de recherche, CNRS	Rapporteur
Sandra LUQUE, Directrice de recherche, INRAE (centre Occitanie-Montpellier)	Examinatrice
Pierre COUTERON, Directeur de recherche, IRD	Président

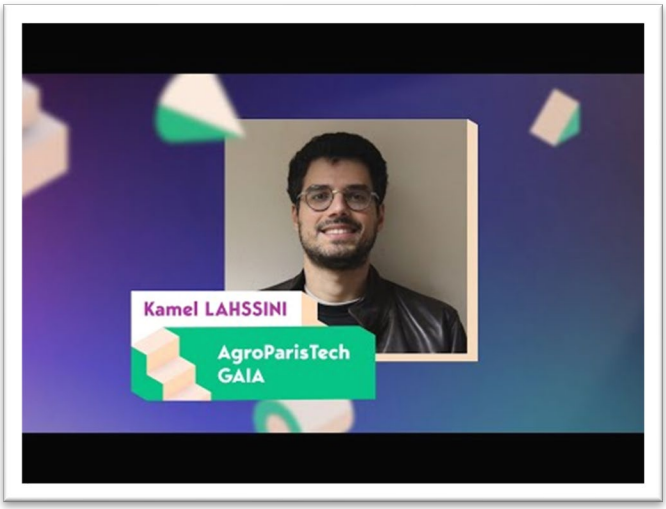
AgroParisTech

UNIVERSITÉ DE MONTPELLIER

Prologue

This doctoral thesis is the result of three years of work co-financed by the *Centre national d'études spatiales* (CNES) and the *Institut national de recherche pour l'agriculture, l'alimentation et l'environnement* (INRAE). This three-year research journey was encapsulated into a three-minute speech shared with a broad audience during the participation in *Ma thèse en 180 secondes* regional competition.

The presentation video can be viewed at the following link:



Below, the transcript of the presentation is given to serve as prelude to the scientific exploration that follows.

Vous sentez ? Les rayons du soleil sur votre visage… Le souffle du vent dans vos cheveux… Le calme plat, troublé uniquement par le bruit léger et continu de la flamme d'un brûleur… Oui, vous êtes dans une montgolfière et vous volez au-dessus de l'Amazonie. Mais que faites-vous si loin de chez vous ? Vous êtes à la recherche d'une cité perdue, enfouie au plus profond de la forêt, et qui abrite en son sein un trésor fabuleux que vous convoitez. Avec le peu d'indices dont vous disposez, trouver la cité au sol serait impossible sur un territoire aussi vaste et inaccessible.

Alors vous avez décidé de prendre de la hauteur. Mais là-haut, un problème de taille se pose pour vous. Depuis le ciel, aussi loin que porte votre regard, vous ne voyez rien d'autre que du vert, du vert, du vert à perte de vue. Ça, c'est la couleur verte moyenne d'une forêt en Amazonie. Impossible de distinguer quoi que ce soit. Ce plafond de vert que vos yeux ne peuvent pas traverser, un outil fantastique peut alors le faire à votre place. Lequel ? Le laser !

Le laser voit à travers la végétation. Il est capable de pénétrer les canopées les plus denses jusqu'au sol pour voir ce que l'œil humain ne peut pas. Et si on monte très haut, très très haut, là où vous ne pouvez pas aller avec votre montgolfière, dans l'espace, on peut voir des pays et des continents entiers. Dans ma thèse j'utilise un laser ultrapuissant, monté sur la Station Spatiale Internationale, pour mesurer précisément la biomasse de la forêt tropicale et ainsi mieux comprendre le cycle de la Terre. Le nom de ce laser ? GEDI, avec un G. GEDI est une sorte de pistoler laser géant qui tourne autour de la Terre et qui tire 250 coups par seconde. Suivons ensemble un seul de ces coups de laser : il voyage dans l'espace à la vitesse de la lumière, pénètre dans l'atmosphère, traverse tous les nuages sur son passage, parvient au sommet de la forêt, se diffuse à travers la végétation et rebondit sur le sol. Et ensuite ? La même chose, en sens inverse. Un infime rayonnement parvient ainsi à revenir à son point de départ et est capté par la force de GEDI. Et cette goutte de lumière qui revient contient un concentré pur d'informations sur la forêt.

Maintenant, imaginez ça 250 fois par seconde et sur toute l'Amazonie. Ça fait beaucoup de données, un amas touffu et dense qu'il faut débroussailler. Mais attention ! Pas question d'y aller sauvagement à la machette pour se frayer un passage, non non non. J'utilise des outils très fins d'intelligence artificielle, basés sur des réseaux de neurones, pour transformer ces millions de gouttes de lumière en information forestière. On est donc plutôt sur de la taille minutieuse de haie façon jardinage au château de Versailles. Un travail qui nécessite à la fois délicatesse et précision, sans oublier un peu d'amour aussi. Car mon trésor à moi je l'ai trouvé : c'est la forêt elle-même.



"Forest X" by Vassilis Theocharakis (1930).

This painting was viewed at the National Gallery, Athens, Greece, during my attendance at the 2024 IEEE International Geoscience and Remote Sensing Symposium.

Abstract

The REDD+ program aims at reducing carbon emissions resulting from forest degradation and deforestation. To meet its objectives, better understanding and characterization of forests world-wide, especially tropical forests, are required. This doctoral thesis explores the potential and challenges of spaceborne LiDAR for multi-source canopy height mapping through the integration of complementary remote sensing technologies to enhance the estimation of canopy height in tropical forests. Canopy height is a critical parameter for quantifying the biomass and the carbon stocks of forests. This research is structured in four studies, each corresponding to a journal article, which collectively address the use of spaceborne LiDAR and other remote sensing data for canopy height mapping. The first study investigates the accuracy of canopy height estimates derived from the Global Ecosystem Dynamics Investigation (GEDI) LiDAR data over tropical forests in French Guiana (South America) and Gabon (Africa). The results reveal that regression models incorporating multiple GEDI metrics significantly improve the estimation accuracy compared to using a single GEDI height metric. Signal physical parameters such as beam type and sensitivity, which affect laser penetration, are identified as major influences on the quality of GEDI data. The second study extends the analysis of GEDI-derived canopy height estimates to tropical forests of Mayotte Island, characterized by moderate canopy heights and steep terrain. Terrain slope proves to significantly impact the measurements retrieved by GEDI and needs to be accounted for when dealing with data that was acquired over steep areas. Accounting for slope can be done through the integration of terrain information in the regression models or through the direct correction of slope effects in the GEDI waveforms. Moreover, LiDAR beam penetration capability appears to be strongly dependent on forest characteristics, as signal penetration depth differs for different forest types. The third study presents a comprehensive canopy height map of French Guiana at a 10-m spatial resolution, produced through an operational data fusion approach that integrates optical, radar, and ancillary environmental data sources. The study employs a U-Net neural network model trained and validated using GEDI data as reference canopy heights. The integration of hydrological and geomorphological descriptors, such as the height above nearest drainage and forest landscape types, significantly enhances the model's accuracy. Moreover, addressing the uncertainties of the GEDI database through the filtering of relevant data and the correction of geolocation errors further improves the performance of the canopy height estimation through U-Net. In the fourth and final study, the spatial correlation of this canopy height map is analyzed, and geostatistical approaches are implemented to improve the canopy height predictions. This analysis shows that the characteristics of the GEDI sensor, particularly laser beam type and the ground sampling pattern, introduce sensor-induced anisotropies in

the measurements that are not representative of the actual spatial variability of the canopy height. The residual kriging spatial interpolation technique addresses the spatial autocorrelation of canopy heights and improves the accuracy of the estimates.

Keywords: Remote Sensing, LiDAR, GEDI, Satellite Imagery, Data Fusion, Tropical Forest

Résumé

Le programme REDD+ vise à réduire les émissions de carbone résultant de la dégradation des forêts et de la déforestation. Pour atteindre ses objectifs, une meilleure compréhension et caractérisation des forêts du monde entier, en particulier des forêts tropicales, sont nécessaires. Cette thèse de doctorat explore le potentiel et les défis du LiDAR spatial pour la cartographie multisource de la hauteur de la canopée à travers l'intégration de technologies de télédétection complémentaires pour améliorer l'estimation de la hauteur de la canopée dans les forêts tropicales. La hauteur de la canopée est un paramètre essentiel pour quantifier la biomasse et les stocks de carbone des forêts. Cette recherche est structurée en quatre études, chacune correspondant à un article scientifique, qui traitent collectivement de l'utilisation du LiDAR spatial et d'autres données de télédétection pour la cartographie de la hauteur de la canopée. La première étude porte sur la précision des estimations de la hauteur de la canopée obtenues à partir des données LiDAR Global Ecosystem Dynamics Investigation (GEDI) sur des forêts tropicales en Guyane française (Amérique du Sud) et au Gabon (Afrique). Les résultats révèlent que des modèles de régression incorporant plusieurs métriques GEDI améliorent de manière significative la précision de l'estimation par rapport à l'utilisation d'une seule métrique de hauteur GEDI. Les paramètres physiques du signal tels que le type de faisceau et la sensibilité, qui affectent la pénétration du laser, sont identifiés comme des influences majeures sur la qualité des données GEDI. La deuxième étude étend l'analyse des estimations de la hauteur de la canopée à partir de GEDI aux forêts tropicales de l'île de Mayotte, caractérisées par des hauteurs de canopée modérées et un terrain escarpé. La pente du terrain a un impact significatif sur les mesures effectuées par GEDI et doit être prise en compte lorsque les données ont été acquises dans des zones escarpés. La prise en compte de la pente peut se faire par l'intégration d'informations sur le terrain dans les modèles de régression ou par la correction directe des effets de pente dans les formes d'onde GEDI. De plus, il apparaît que la capacité de pénétration du faisceau LiDAR dépend fortement des caractéristiques de la forêt, puisque la profondeur de pénétration du signal diffère selon le type de forêt. La troisième étude présente une carte complète de la hauteur de la canopée en Guyane française à une résolution spatiale de 10 m, produite à partir d'une approche opérationnelle de fusion de données qui intègre des données optiques, radar et des sources de données environnementales auxiliaires. L'étude utilise un modèle de réseau de neurones U-Net calibré et validé à partir de données GEDI comme hauteurs de canopée de référence. L'intégration de descripteurs hydrologiques et géomorphologiques, tels que la hauteur au-dessus du drainage le plus proche et les types de paysages forestiers, améliore considérablement la précision du modèle. En outre, la prise en compte des incertitudes de la

base de données GEDI par le filtrage des données pertinentes et la correction des erreurs de géolocalisation améliore davantage la performance de l'estimation de la hauteur de la canopée par U-Net. Dans la quatrième et dernière étude, la corrélation spatiale de cette carte de la hauteur de la canopée est analysée et des approches géostatistiques sont mises en œuvre pour améliorer les prédictions de la hauteur de la canopée. Cette analyse montre que les caractéristiques du capteur GEDI, en particulier le type de faisceau laser et le schéma d'échantillonnage au sol, introduisent dans les mesures des anisotropies induites par le capteur qui ne sont pas représentatives de la variabilité spatiale de la hauteur de canopée. La technique d'interpolation spatiale par krigeage des résidus prend en compte l'autocorrélation spatiale des hauteurs de la canopée et améliore la précision des estimations.

Mots-clés : Télédétection, LiDAR, GEDI, Imagerie satellitaire, Fusion de données, Forêt tropicale

Résumé détaillé

Objectifs

Dans le contexte du changement climatique, la communauté scientifique a largement relayé la nécessité de mieux comprendre le cycle du carbone et de développer de nouvelles approches pour estimer avec précision les stocks de carbone à l'échelle régionale et mondiale. Ces dernières années, les chercheurs ont noté le besoin urgent de nouvelles technologies et méthodes pour mesurer les stocks de biomasse forestière à l'échelle mondiale, ce qui permettra de fournir des informations qui aideront non seulement à mieux comprendre le cycle du carbone mondial, mais aussi plus directement à réduire les émissions de gaz à effet de serre en limitant la déforestation et la dégradation des forêts. La télédétection spatiale, et en particulier la télédétection LiDAR, représente une solution intéressante et complémentaire aux autres techniques de télédétection pour cartographier la biomasse forestière à grande échelle. Pour ce faire, la hauteur de la canopée est une variable structurelle mesurable qui sert d'indicateur principal pour estimer la biomasse aérienne.

Au cours des dernières années, de nombreuses études ont exploité la complémentarité des données radar et LiDAR. Les données optiques sont également utilisées pour estimer la hauteur de la canopée, souvent en complément des données radar et LiDAR. Cependant, deux limitations principales ont été identifiées dans ces travaux, qui utilisent principalement des données satellitaires radar en bande L et des données LiDAR d'ICESat-1. D'une part, les anciennes missions LiDAR satellitaires, telles que le capteur GLAS, fournissent des données sur de vastes zones mais ne sont pas en mesure d'acquérir des données avec une densité spatiale suffisante. Elles présentent en effet une faible densité spatiale d'empreintes LiDAR, par exemple environ 0,5 points/km² sur un territoire comme la Guyane française. D'autre part, les signaux radar en bande L présentent une faible sensibilité aux niveaux de biomasse élevés, avec une saturation à partir d'environ 150 mégatonnes (Mt) par hectare, un seuil qui est bien en-dessous des niveaux de biomasse normaux dans les écosystèmes tropicaux, et en particulier en Guyane française. Les données optiques sont également concernées par les problèmes de saturation, dans la mesure où il a été observé que la réflectance de la canopée des forêts tropicales sature à une certaine hauteur et n'évolue plus au-delà de ce seuil.

Pour relever ces défis, la mission LiDAR spatial GEDI, lancée en 2018, a été conçue pour produire des données permettant de cartographier la biomasse forestière mondiale. D'autres missions radar sont prévues à l'avenir pour remédier aux limitations observées, notamment la

mission radar en bande P BIOMASS de l'ESA (dont le lancement est prévu en 2024) et la mission radar en bande L NISAR (dont le lancement est également prévu en 2024) soutenue par la NASA (Etats-Unis) et l'ISRO (Inde).

L'objectif scientifique principal de cette thèse de doctorat est d'exploiter les nouvelles données GEDI pour la cartographie de la hauteur de la canopée à l'échelle de la Guyane française. Les données fournies par la mission GEDI ont une densité spatiale plus élevée que celles fournies précédemment par ICESat-1. De plus, elles sont caractérisées par une résolution spatiale plus fine, avec des empreintes au sol de 25 m de diamètre contre 70 m pour les données de ICESat-1. Ces paramètres permettent de générer des produits de hauteur de canopée avec une plus grande précision et une résolution spatiale plus fine (de l'ordre de 10 m). Les données GEDI sont utilisées en complément de données radar existantes en bande C et en bande L, ainsi que des données optiques multispectrales. Dans ce contexte, de nouvelles approches pour l'estimation de la hauteur de la canopée, basées sur la complémentarité multi-capteurs, sont étudiées.

De cet objectif principal découlent plusieurs questions de recherche essentielles concernant l'utilisation des données GEDI pour la cartographie de la hauteur de la canopée, en particulier dans le contexte des écosystèmes tropicaux tels que la Guyane française. La disponibilité des données GEDI soulève des questions importantes sur leur qualité et leur potentiel pour améliorer la cartographie de la hauteur de la canopée. L'évaluation de l'utilité et de la fiabilité des mesures GEDI est une étape obligatoire pour comprendre leur intérêt et cette évaluation est menée en considérant différents niveaux et caractéristiques des données :

1. Comment les propriétés physiques du signal GEDI et les paramètres de post-traitement du signal influencent-ils les mesures ? Les mesures GEDI, comme celles de tout instrument, sont influencées par les interactions physiques du signal. Les propriétés du signal laser et la chaîne de traitement des données impactent directement les formes d'onde et les métriques extraites. Il est donc essentiel de comprendre leur impact sur les mesures de hauteur de la canopée pour pouvoir interpréter et utiliser ces données avec précision.
2. Comment les facteurs environnementaux, tels que le relief et la structure de la végétation, affectent-ils les mesures de GEDI ? Dans les écosystèmes tropicaux, des facteurs comme la configuration du terrain et la structure de la végétation peuvent influencer de manière significative les mesures obtenues par GEDI. Il est donc essentiel de comprendre et quantifier ces impacts pour développer des stratégies de correction et améliorer la précision des cartes de hauteur de la canopée.

En s'appuyant sur les enseignements tirés de l'évaluation des données GEDI, la prochaine étape est consacrée au développement de méthodologies efficaces pour la cartographie de la hauteur de la canopée, afin d'exploiter pleinement le potentiel des données GEDI, en combinaison avec d'autres sources de données :

3. Comment développer des modèles efficaces de fusion de données afin d'exploiter plusieurs sources de données pour la cartographie de la hauteur de la canopée ? L'intégration de plusieurs sources de données de télédétection (optique, radar, LiDAR) est essentielle pour exploiter le potentiel des données GEDI pour la cartographie de la hauteur de la canopée à l'échelle régionale de la Guyane française. La manière dont ces différentes données sont intégrées dans les modèles de prédiction est également d'une importance capitale pour maximiser leur potentiel et extraire l'information liée à la hauteur de la canopée. Cette question concerne également la sélection de descripteurs environnementaux pertinents qui peuvent enrichir les modèles et améliorer les prédictions, ainsi que la compréhension de comment et pourquoi ces données contribuent positivement aux résultats.
4. Comment les paramètres d'acquisition des données GEDI influencent-t-ils les résultats de la cartographie de la hauteur de la canopée ? Les données GEDI sont caractérisées par des points de mesure dispersés et inégalement répartis. Malgré l'existence de méthodes capables de gérer ce type de données, la prise en compte des dépendances spatiales dans le processus de cartographie de la hauteur de la canopée reste essentielle, puisque la hauteur de la canopée est une variable continue et ses dynamiques spatiales résultent de nombreux facteurs environnementaux complexes. La compréhension et la prise en compte de cette variabilité et de l'autocorrélation spatiale sont des étapes importantes pour générer des produits plus précis et plus fiables sur la hauteur de la canopée.

Ces questions guident la recherche effectuée dans cette thèse de doctorat intitulée « Potentiel et défis du LiDAR spatial pour la cartographie multisource de la hauteur de la canopée dans les forêts tropicales ». Une carte de la hauteur de la canopée en Guyane française à une résolution spatiale de 10 m est présentée.

Influence des paramètres d'acquisition et de traitement de GEDI sur les estimations de la hauteur de la canopée dans les forêts tropicales

La première étude de la thèse vise à répondre à la première question de recherche : comment les propriétés physiques du signal GEDI et les paramètres de post-traitement du signal influencent-ils les mesures ? L'évaluation des données GEDI dans un contexte tropical est

essentielle pour l'objectif plus large de cette thèse de doctorat qui vise à utiliser ces données comme référence pour la cartographie de la hauteur de la canopée.

GEDI est un capteur LiDAR spatial spécialement conçu pour mesurer la structure verticale de la végétation dans le monde et pour surmonter les limites des missions LiDAR précédentes. En tant que système LiDAR à retour d'onde complète, la structure verticale de la végétation est mesurée grâce à des formes d'onde, qui sont des signaux temporels unidimensionnels résultant des interactions entre les impulsions laser émises et la végétation. Le produit GEDI de niveau L1B (GEDI01_B) fournit des formes d'onde corrigées, lissées et géolocalisées qui décrivent les profils végétation à l'intérieur d'empreintes au sol et qui contiennent donc des informations sur la structure verticale des forêts. Il est important d'identifier les paramètres qui influencent les formes d'onde du produit L1B pour interpréter correctement les valeurs de hauteur de canopée dérivées de ces données. La qualité et la forme des signaux sont influencées par plusieurs facteurs principalement liés aux propriétés physiques du signal laser, à savoir la puissance du laser et la sensibilité du faisceau. La puissance du laser correspond à l'énergie des impulsions émises et GEDI fonctionne avec deux niveaux de puissance (pleine puissance et demi-puissance). La sensibilité du faisceau est définie comme la couverture maximal du couvert végétal qui peut être entièrement pénétrée par le signal laser. Ces propriétés influencent la façon dont le laser interagit avec la végétation, ce qui a un impact sur les profils verticaux mesurés par GEDI. En effet, pour décrire avec précision la structure verticale complète de la végétation, il est nécessaire que le signal GEDI atteigne le sol. Dans les écosystèmes tropicaux, qui sont généralement caractérisés par des niveaux de biomasse élevés, la capacité du signal à pénétrer la végétation dense jusqu'au sol pour fournir des informations détaillées sur les couches profondes de la canopée est un défi majeur pour mesurer la hauteur de la canopée avec précision.

L'information contenue dans les formes d'onde L1B peut être synthétisée sous la forme de métriques extraites des signaux. Dans cette perspective, le produit GEDI de niveau L2A fournit une interprétation des formes d'onde à travers l'extraction de métriques descriptives liées à l'élévation du sol, la hauteur du sommet de la canopée et des métriques de hauteur relative. Ces diverses mesures sont utilisées pour estimer la hauteur de la canopée et sont produites à partir des formes d'onde en utilisant six configurations possibles de traitement du signal. Par conséquent, les valeurs obtenues sont sensibles à la configuration utilisée. Les paramètres de chaque configuration, tels que le niveau de lissage, le filtrage du bruit et la sélection de seuils, influencent de manière significative les métriques extraites et les estimations de hauteur de la canopée. Il est donc important de comprendre et d'évaluer ces configurations de traitement du

signal afin de sélectionner la plus appropriée dans un environnement complexe comme les forêts tropicales denses.

Pour répondre à ces objectifs, ce chapitre étudie la précision des estimations de la hauteur de la canopée à partir de données LiDAR GEDI acquises sur la Guyane française (Amérique du Sud) et le Gabon (Afrique). L'étude associée évalue l'influence de plusieurs paramètres d'acquisition et de traitement du signal GEDI sur l'estimation de la hauteur de la canopée. En utilisant comme référence des modèles de hauteur dérivés de données LiDAR aéroporté, plusieurs approches linéaires et non-linéaires sont testées pour estimer la hauteur de la canopée à partir de données GEDI. Les résultats révèlent que les modèles de régression incorporant plusieurs métriques GEDI améliorent de manière significative la précision de l'estimation par rapport à l'utilisation d'une seule métrique de hauteur. En particulier, les modèles Random Forest améliorent la précision de l'estimation de la hauteur de la canopée de près de 80% en termes d'erreur par rapport à l'utilisation d'une unique métrique GEDI comme indicateur direct de la hauteur de la canopée. L'approche par réseaux de neurones convolutifs calibrés directement sur les formes d'onde GEDI présente également des résultats intéressants, mais néanmoins légèrement inférieurs à ceux des modèles Random Forest basés sur plusieurs métriques. Les facteurs importants liés à la physique du signal, comme la puissance du faisceau et la sensibilité, qui affectent la pénétration du laser, sont identifiés comme des influences majeures sur la qualité des données GEDI. Il est ainsi recommandé de n'utiliser que les faisceaux de pleine puissance et de haute sensibilité lorsque les données disponibles sont en quantité suffisante. Par ailleurs, le groupe d'algorithme numéro 5 apparaît comme la configuration de traitement du signal la mieux adaptée pour un contexte tropical dense. Enfin, les modèles de régression calibrés sur des données de référence s'avèrent transférables sur des sites d'étude présentant des conditions de structure de la végétation similaires.

Utilisation du signal GEDI et des paramètres environnementaux pour améliorer l'estimation de la hauteur de la canopée dans les écosystèmes forestiers tropicaux à Mayotte

À la suite de l'évaluation de plusieurs approches pour l'estimation de la hauteur de la canopée à partir de données GEDI, avec un accent particulier mis sur l'impact des paramètres d'acquisition et de traitement sur la précision des estimations, la deuxième étude de la thèse vise à répondre à la deuxième question de recherche : comment les facteurs environnementaux, tels que le relief et la structure de la végétation, affectent-ils les mesures de GEDI ? Dans les applications de télédétection, le capteur ainsi que l'environnement

observé influencent les données acquises par les instruments. Après avoir acquis une meilleure compréhension des paramètres d'acquisition et de traitement de GEDI, la prochaine étape importante consiste à évaluer l'impact des caractéristiques des environnements tropicaux sur les mesures de hauteur de GEDI. Cette évaluation fait partie de l'objectif plus large de cette thèse de doctorat qui vise à développer une méthodologie robuste pour la cartographie de la hauteur de la canopée à partir de données GEDI.

Les signaux rétrodiffusés mesurés par GEDI et les variables forestières et métriques de hauteur qui en découlent peuvent contenir des incertitudes dues à la distorsion du signal causée par les conditions topographiques à l'intérieur des empreintes au sol. Comprendre et corriger les effets de pente est une étape obligatoire pour les études effectuées sur des zones escarpées où la pente a un impact important sur les mesures de GEDI. En outre, la structure de la végétation exerce également une influence directe sur les données collectées par GEDI. Dans les écosystèmes tropicaux, la capacité de pénétration du faisceau LiDAR dépend fortement des caractéristiques de la forêt. La structure horizontale de la végétation, selon qu'elle soit plus ou moins hétérogène, peut aussi exacerber les erreurs liées aux incertitudes de géolocalisation des empreintes GEDI.

Afin de caractériser et de corriger ces effets, ce chapitre étend l'analyse des estimations de la hauteur de la canopée à partir de données GEDI aux forêts tropicales de l'île de Mayotte (Archipel des Comores), caractérisées par des hauteurs de canopée et des niveaux de biomasse modérés, ainsi que par un terrain escarpé. L'étude associée évalue l'impact du signal GEDI (sensibilité du faisceau) et de paramètres environnementaux (pente et hauteur de la végétation) sur les estimations de la hauteur de la canopée dans un environnement forestier hétérogène. Des modèles de régression linéaires et non-linéaires sont développés à partir de métriques GEDI et calibrés sur des données de référence issues d'acquisitions LiDAR aéroporté. Les résultats indiquent à nouveau que les modèles de régression utilisant plusieurs métriques GEDI sont plus précis que les estimations à partir d'une seule métrique, avec une erreur passant de 6,6 m à 5,7 m. Les facteurs significatifs qui affectent les estimations de hauteur à partir de GEDI sont la pente du terrain, la hauteur de la canopée et la sensibilité du faisceau. En effet, la pente du terrain s'avère avoir un impact significatif sur les mesures effectuées par GEDI et doit être considérée dans le cadre d'un contexte escarpé. La prise en compte de la pente peut se faire soit par l'intégration d'informations sur le terrain dans les modèles de régression, soit par la correction directe des effets de pente dans les formes d'onde de GEDI. De plus, la structure de la végétation est déterminante dans les mesures effectuées par GEDI. Dans le cas des forêts de Mayotte, la réduction de la profondeur de pénétration du faisceau laser apparaît dès une hauteur de canopée de 15 m, contrairement

aux résultats d'autres études sur d'autres zones tropicales qui font état d'une pénétration totale du signal GEDI jusqu'à des hauteurs d'environ 30 m. Cela implique que la capacité de pénétration du faisceau LiDAR dépend fortement des caractéristiques de la forêt, et la profondeur de pénétration du signal peut ainsi varier entre des types de forêts différents ayant pourtant des niveaux de hauteur et de biomasse similaires. L'erreur de géolocalisation des données GEDI représente également un défi majeur dans les environnements forestiers hétérogènes, dans la mesure où environ la moitié des erreurs d'estimation de la hauteur de la canopée peut être attribuée aux incertitudes de géolocalisation.

Cartographie de la hauteur de la canopée en Guyane française à partir de données satellitaires multisources et d'informations environnementales dans une architecture U-Net

En s'appuyant sur les conclusions des analyses précédentes qui se sont concentrées sur l'évaluation des performances de GEDI en milieu tropical et sur les différentes influences sur les mesures de hauteur de la canopée, la troisième étude de la thèse vise à répondre à la troisième question de recherche : comment développer des modèles efficaces de fusion de données afin d'exploiter plusieurs sources de données pour la cartographie de la hauteur de la canopée ? Une carte à haute résolution de la hauteur de la canopée à l'échelle régionale de la Guyane française est présentée et discutée. Cette carte a été générée grâce à l'intégration optimale des données GEDI avec d'autres sources de télédétection et des descripteurs environnementaux.

Après l'évaluation exhaustive des capacités et des limites de GEDI dans les écosystèmes forestiers tropicaux, l'objectif opérationnel de cette thèse de doctorat s'oriente vers la production d'une carte continue de la hauteur de la canopée en Guyane française à une résolution spatiale de 10 m. L'étude associée présente une approche d'apprentissage profond qui exploite la complémentarité entre les mesures dispersées de GEDI et d'autres données de télédétection en libre accès (données optiques et radar), ainsi que des variables environnementales liées à la structure de la canopée et au contexte écologique de la Guyane française (hauteur au-dessus du drainage le plus proche et types de paysages forestiers). L'intégration de ces diverses sources de données est réalisée grâce à une architecture U-Net qui est particulièrement adaptée au traitement d'informations spatiales complexes et hétérogènes. Le modèle, appelé *CHNET*, est entraîné en utilisant une métrique GEDI comme hauteur de canopée de référence. Les conclusions tirées lors des analyses précédentes permettent une utilisation optimale de ces données et guident un processus d'amélioration

itératif à travers des scénarios d'entraînement successifs, et ce afin d'améliorer la précision des prédictions finales de hauteur de la canopée.

Une carte complète de la hauteur de la canopée en Guyane française à une résolution spatiale de 10 m est produite grâce à une approche opérationnelle de fusion de données qui intègre des données optiques (Sentinel-2), radar (Sentinel-1 et ALOS) et des sources auxiliaires de données environnementales dans une architecture U-Net. Le modèle est calibré et validé à l'aide de données GEDI de référence. La précision de la carte de hauteur est validée grâce à des données LiDAR aéroporté à haute résolution de l'Office national des forêts (ONF). L'intégration de descripteurs hydrologiques et géomorphologiques (hauteur au-dessus du drainage le plus proche et types de paysages forestiers) améliore considérablement la précision du modèle. Ces descripteurs sont liés à la structure de la canopée et apportent des informations que les données de télédétection classiques ne peuvent pas révéler à elles seules. En outre, la prise en compte des incertitudes de la base de données GEDI par le filtrage des données pertinentes et la correction des erreurs de géolocalisation améliorent encore plus les performances de l'estimation de la hauteur de la canopée à partir de U-Net.

Les résultats montrent d'une part que les conséquences des imprécisions des données GEDI peuvent être atténuées par une approche opérationnelle, et d'autre part que l'intégration de descripteurs environnementaux pertinents pour caractériser la hauteur de la canopée apporte une valeur ajoutée significative. La carte de hauteur de la canopée ainsi produite peut servir de base utile pour l'estimation de la biomasse en Guyane française et pour l'évaluation des stocks de carbone de la région. De manière plus générale, cette étude fournit un cadre opérationnel et reproductible pour la cartographie de la hauteur de la canopée à partir de données librement disponibles, cadre qui peut être adapté et utilisé dans différentes régions tropicales.

Traitement des anisotropies liées au capteur dans les mesures GEDI par krigeage des résidus pour la cartographie de la hauteur de la canopée en Guyane française

A partir de cette carte à haute résolution de la hauteur de la canopée en Guyane française, la quatrième étude de la thèse aborde la quatrième et dernière question de recherche : comment les paramètres d'acquisition des données GEDI influencent-ils les résultats de la cartographie de la hauteur de la canopée ? Les objectifs sont d'examiner l'impact des caractéristiques de l'acquisition GEDI sur les estimations de la hauteur de la canopée et d'explorer des méthodes spatiales pour améliorer ces résultats.

Dans cette étude, deux cartes différentes construites à partir d'approches de modélisation distinctes et avec les mêmes données d'entrée sont utilisées comme base pour l'analyse : l'architecture U-Net présentée précédemment (*CHNET*) et un algorithme Random Forest (*RFH*). L'approche U-Net étant conçue pour traiter des données de référence discrètes et incorporer des informations spatiales dans sa conception, elle a été choisie comme cadre principal pour produire une carte de hauteur de la canopée. En revanche, le modèle Random Forest est une méthode non spatiale qui traite chaque échantillon de données de manière indépendante sans tenir compte des relations spatiales. En comparant ces deux méthodologies, l'étude évalue l'importance de l'intégration de l'autocorrélation spatiale dans le processus de cartographie d'une variable continue comme la hauteur de la canopée. L'objectif final est d'améliorer les estimations par krigeage des résidus pour prendre en compte l'autocorrélation spatiale des hauteurs de la canopée dans le processus de modélisation.

L'analyse spatiale des résidus des deux modèles révèle la présence d'anisotropies qui ne représentent pas la structure horizontale réelle de la canopée. Ces anisotropies s'expliquent par les caractéristiques d'acquisition du capteur GEDI, notamment les variations de la puissance du laser et l'agencement spatial des traces au sol de GEDI. Les différences de pénétration du laser dans la végétation causées par les différents niveaux de puissance entraînent des incohérences dans les mesures sur la zone d'étude. Ces effets associés à la configuration spatiale de l'acquisition GEDI (transects au sol régulièrement espacés et orientés dans deux directions azimutales données) introduisent systématiquement des anisotropies périodiques dans les mesures. Ces artefacts dus au capteur doivent être pris en compte pour obtenir des estimations plus fiables de la hauteur de la canopée.

Pour améliorer les estimations de hauteur de la canopée, une procédure de krigeage des résidus est intégrée dans le processus de modélisation. A partir de l'analyse spatiale préliminaire des résidus, les informations spatiales sont incorporées d'une manière qui atténue et traite les anisotropies liées au capteur. Pour ce faire, des sous-ensembles de la base de données GEDI originale sont analysés afin de produire une représentation précise et fidèle de l'autocorrélation spatiale réelle. Ainsi, les faisceaux GEDI de pleine puissance et alignés sur un seul azimut sont utilisés pour déterminer la structure spatiale réelle des hauteurs de la canopée telles que mesurées par GEDI. La structure spatiale calculée sur ces données dans la direction de l'azimut permet d'éliminer les sources d'anisotropies liées au capteur et est considérée comme une représentation fiable de la variabilité spatiale des hauteurs de la canopée de manière isotrope. L'intégration de composantes spatiales dans le processus de modélisation par le biais du krigeage des résidus permet d'améliorer la précision finale des estimations de hauteur de la canopée. Les deux modèles *CHNET* et *RFH* bénéficient de ces

corrections, avec des gains plus importants observés pour *RFH*. Les corrections apportées par le krigeage des résidus sont localisées autour des empreintes GEDI. Par conséquent, la densité de points de mesure utilisables est un facteur clé pour l'efficacité du krigeage des résidus. Enfin, les résultats révèlent qu'un modèle Random Forest associé à une interpolation spatiale peut fournir des performances comparables à celles d'un modèle U-Net seul.

Conclusions

En conclusion, ce travail de thèse met en évidence le potentiel de GEDI pour mesurer avec précision la hauteur de la canopée, en particulier dans les forêts tropicales où les relevés terrain sont difficiles à mettre en œuvre. En utilisant les faisceaux laser de pleine puissance et les paramètres de traitement adaptés, les données GEDI peuvent fournir des informations fiables sur la structure verticale des forêts tropicales denses. L'intégration de ces données avec des modèles statistiques et des approches d'apprentissage profond a permis de créer des cartes de la hauteur de la canopée à haute résolution (10 m) utiles pour la gestion des forêts et l'évaluation des stocks de carbone mondiaux. Toutefois, des défis et difficultés persistent. Les erreurs de géolocalisation et la pénétration limitée du laser dans les forêts extrêmement denses affectent la précision des mesures.

Cette thèse ouvre des perspectives intéressantes, notamment sur l'utilisation des estimations de hauteur de la canopée pour obtenir des prédictions de la biomasse aérienne (AGB). Bien que la hauteur soit un bon indicateur de l'AGB, l'intégration d'autres paramètres structuraux et environnementaux est nécessaire pour améliorer les modèles d'estimation. Par ailleurs, il est suggéré que les recherches à venir sur ces questions se concentrent sur la réduction des incertitudes liées à l'estimation de l'AGB dans les forêts tropicales denses, notamment grâce au déploiement et à l'utilisation de nouveaux capteurs spatiaux moins sensibles aux problématiques de saturation des mesures.

Table of Contents

Chapter 1: Introduction **22**

1.1 Carbon and forests 22
1.1.1 The carbon cycle *22*
1.1.2 Human activities and carbon cycle disruption *23*
1.1.3 Forest preservation: a major stake *24*
1.2 Tropical forests 26
1.2.1 Tropical forests in the world *26*
1.2.2 A threatened carbon stock *27*
1.2.3 Forest biomass: a key driver of global carbon stocks *28*
1.2.4 Canopy height: a key indicator of forest biomass *29*
1.3 Contribution of remote sensing 30
1.3.1 Remote sensing and forest structure *30*
1.3.2 LiDAR technology for canopy height estimation *31*
1.3.3 Canopy height mapping at regional and global scales *33*
1.4 Scientific framework 36
1.4.1 Objectives and research questions *36*
1.4.2 Dissertation plan *38*

Chapter 2: Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests **40**

Chapter 3: Use of GEDI Signal and Environmental Parameters to Improve Canopy Height Estimation over Tropical Forest Ecosystems in Mayotte Island **68**

Chapter 4: Canopy Height Mapping in French Guiana Using Multi-Source Satellite Data and Environmental Information in a U-Net Architecture **92**

Chapter 5: Addressing GEDI-Induced Anisotropies in Residual Kriging to Improve Multi-Source Canopy Height Estimates in French Guiana **116**

Chapter 6: Conclusions and Perspectives **136**

6.1 Conclusions 136
6.1.1 Potential *136*
6.1.2 Challenges *137*
6.2 Perspectives 138
6.2.1 From canopy height to AGB *138*
6.2.2 Better characterizing dense forest ecosystems *139*

References **142**

Chapter 1

Introduction

1.1 Carbon and forests

1.1.1 The carbon cycle

The global carbon cycle represents the complex exchange of carbon among the Earth's four major carbon reservoirs: the atmosphere, biosphere, hydrosphere, and lithosphere. This natural biogeochemical cycle is essential for balancing carbon, a fundamental component of life, across these different compartments. The carbon cycle's complexity arises from the interactions between life, which is composed of carbon-based molecules, and the diverse ecosystems that store and exchange carbon in many ways (Falkowski et al., 2000).

Global carbon is found in varying quantities and forms across the four primary reservoirs (Sabine et al., 2004). The lithosphere, which consists of carbonaceous and carbonate rocks, holds most of Earth's carbon, with an estimated amount of 30,000,000 gigatons (Gt). However, this carbon is mostly trapped in rocks and deep ocean sediments, and thus plays a little part in the active carbon cycle. The hydrosphere, which includes the world's oceans and seas, contains approximately 38,000 Gt of dissolved carbon as well as 3 Gt of carbon in marine organisms. The biosphere, which includes all living species, deceased organic matter, and soil, stores around 2,300 Gt of carbon. The atmosphere, despite containing only about 700 Gt of carbon in the form of carbon dioxide (CO_2), plays an important role in the carbon cycle by influencing climate and interacting directly with the biosphere and hydrosphere.

The exchanges between these reservoirs drive the global carbon cycle. The lithosphere contains the majority of Earth's carbon, but it has a very slow natural exchange of carbon with other reservoirs. This limited interaction is due to long processes like fossilization and sedimentation that span over millions of years. Conversely, there is a balance between the hydrosphere and the atmosphere because CO_2 is easily absorbed in water. Atmospheric CO_2 dissolves in ocean surface waters, which maintains a balance between the hydrosphere and the atmosphere. Some of this dissolved CO_2 is absorbed by marine organisms, contributing to carbon storage in the hydrosphere. On land, the terrestrial biosphere sequesters atmospheric carbon through photosynthesis, with vegetated ecosystems serving as significant carbon sinks. The carbon stored in the terrestrial biosphere gets released into the atmosphere through

the respiration of living organisms and the decomposition of organic matter. In the end, the carbon exchanges among the four compartments naturally maintain balance, ensuring that carbon levels remain stable in the absence of human interference (Prentice et al., 2001).

1.1.2 Human activities and carbon cycle disruption

Over the last two centuries, since the beginning of the industrialized period, human activities have significantly raised atmospheric carbon concentrations, creating an imbalance in the natural cycle of carbon across the four primary carbon sinks (Crowley, 2000).

One of the most significant human impacts comes from the exploitation and combustion of fossil fuels, which have accumulated over millions of years within the rocks of the lithosphere. When these fuels are burned for energy, they release huge amounts of CO_2 into the atmosphere. This increase in atmospheric CO_2 concentrations disrupts the natural carbon fluxes and contributes to the enhanced greenhouse effect, which is one of the main drivers of global climate change (Calvin et al., 2023).

The oceans, as part of the hydrosphere, absorb a large fraction of CO_2 emissions from human activities and serve as an extremely important buffer to increased atmospheric CO_2 concentrations. However, as CO_2 levels increase in oceanic waters, the water becomes more acidic. This acidification can have detrimental effects on marine ecosystems, especially on marine organisms that rely on calcium carbonate for their shells and skeletons (Feely et al., 2004). Moreover, the capacity of the oceans to absorb CO_2 may decrease over time, further aggravating the accumulation of greenhouse gases in the atmosphere.

Human activities also impact the living organisms and ecosystems within the biosphere. Deforestation and land-use changes, driven by agriculture, urbanization, and other forms of development, reduce the amount of carbon stored in forests and soils (Houghton et al., 2012). These activities not only release large amounts of stored carbon into the atmosphere but also diminish the biosphere's ability to act as a carbon sink. Forests, in particular, play a critical role in carbon sequestration through the process of photosynthesis, and their degradation weakens one of the Earth's most effective tools for mitigating climate change (Pan et al., 2011).

Overall, human activities have profound consequences on the global carbon cycle, particularly through the release of CO_2 from fossil fuels and land-use changes (Friedlingstein et al., 2023; Houghton et al., 2012). While these activities have greatly increased the concentrations of CO_2 in the atmosphere, the biosphere remains the major active component of the carbon cycle that can be managed to restore some balance. Reforestation, conservation efforts, and sustainable

land management practices can enhance the biosphere's potential for carbon storage and thus mitigate some of the impacts of human activities on the carbon cycle (Griscom et al., 2017).

1.1.3 Forest preservation: a major stake

In this perspective, forest preservation has become a priority of global environmental policies. The international efforts towards sustainable management of forests are motivated by their key role in maintaining ecological balance, supporting biodiversity, and mitigating climate change.

Forests cover approximately 31% of the Earth's land surface and are home to many terrestrial species (Food and Agriculture Organization, 2020; Watson et al., 2018). They are a major contributor to the global carbon cycle, acting as both carbon sinks and sources of greenhouse gases. Forest ecosystems sequester around 2.4 Gt of carbon annually (Pan et al., 2011) and therefore significantly influence atmospheric carbon concentrations. Consequently, the preservation of forests is not only a regional or national concern, but a global imperative given their impact on the Earth's climate.

The importance of forest preservation has been recognized at various international forums and through numerous initiatives over the past several decades, reflecting a growing international awareness of the need to protect and preserve these ecosystems. The global response to deforestation and forest degradation has evolved significantly over time, from early efforts focused on conservation to more recent comprehensive approaches that integrate forest preservation with sustainable development goals.

One of the earliest international initiatives aimed at forest conservation was the United Nations Conference on the Human Environment held in Stockholm (Sweden) in 1972. The conclusions of this event highlighted the environmental significance of forests and called for coordinated global action (United Nations, 1973). However, it was not until the late 1980s and 1990s that forest preservation became a major focus of international environmental agreements. The 1992 Earth Summit in Rio de Janeiro marked a turning point with the adoption of Forest Principles and Agenda 21, which set out a global framework for the sustainable management of forests (United Nations, 1992).

Building on these efforts, the United Nations Forum on Forests (UNFF) was established in 2000 to promote the management, conservation and sustainable development of forests. The UNFF provided a platform for international cooperation and dialogue on global forest-related issues. It also allowed the adoption of the Non-Legally Binding Instrument on All Types of

Forests in 2007, which aimed at strengthening political commitment to forest protection (Kunzmann, 2008).

In recent years, the focus on forest preservation has intensified, especially in the context of climate change mitigation. The recognition of forests as a critical component of climate action led to the inclusion of the Reducing Emissions from Deforestation and Forest Degradation (REDD+) mechanism in the 2015 Paris Agreement. REDD+ is a comprehensive framework designed to encourage developing countries to reduce carbon emissions from deforestation and forest degradation while promoting sustainable management practices to enhance forest carbon stocks (Bayrak & Marafa, 2016; Duchelle et al., 2018). The mechanism also insists on the importance of protecting the rights of indigenous people and local communities that depend on forests for their livelihoods.

The United Nations 2030 Agenda for Sustainable Development also underlines the importance of forests to achieve several Sustainable Development Goals (SDGs), particularly SDG number 15, which focuses on life on land (United Nations, 2015). Indeed, this goal calls for urgent and significant actions to reduce deforestation, restore degraded forests, and substantially increase afforestation and reforestation at the global level. Reforestation is the process of planting trees in a forest where the number of trees has been decreasing while afforestation is when new trees are planted or seeds are sown in an area when there were no trees before, thus creating a new forest. The SDGs reflect an extensive approach to forest preservation, associating it with other socio-environmental objectives such as reducing poverty, ensuring food security, and resilience to climate.

Besides these international frameworks, many regional and national initiatives have been taken to limit deforestation and promote forest conservation. For example, in 2003, the European Union adopted the Forest Law Enforcement, Governance and Trade Action Plan (FLEGT) to fight illegal logging by encouraging trade in legally produced timber (Brack, 2012). Similarly, the Amazon Fund was created in 2008 by Brazil to reduce deforestation in the Amazon Rainforest (J. C. Ferraz et al., 2023), which is one of the largest carbon sinks in the world.

Despite all these efforts, challenges remain in effectively preserving forests on a global scale. Deforestation rates are still high, particularly in tropical areas, mainly due to agricultural growth and infrastructure development (Hansen et al., 2013). To address these challenges, there is still a need for continued international cooperation, increased financial support for forest

preservation, and better incorporation of forest conservation strategies into development policies.

1.2 Tropical forests

1.2.1 Tropical forests in the world

Forests cover an area of over 4 billion hectares globally, which amounts to nearly one-third of the Earth's land surface. Specifically, tropical forests account for 45% of the global forest cover (Food and Agriculture Organization, 2020). They represent vital and complex ecosystems that play a crucial role in global biodiversity and climate regulation, as well as in the provision of ecosystem services. The Food and Agriculture Organization (FAO) defines a tropical forest as a forest where the average monthly temperature consistently exceeds 18°C. Tropical forests can be classified into three main types based on their annual rainfall: dry, i.e., less than 1,500 mm/year and with a distinct dry season; moist, i.e., 1,500-3,500 mm/year; and wet, i.e., more than 3,500 mm/year and with evapotranspiration exceeding rainfall for less than a month (Holdridge & Grenke, 1970). Because of their location near the equator and their climatic conditions, they are characterized by dense vegetation and a great diversity of species. Three main regions host the majority of the world's tropical forests: the Amazon Basin in South America, the Congo Basin in Central Africa, and Southeast Asia. The Amazon Basin is the largest, covering approximately 800 million hectares, with about 520 million hectares located within Brazil alone. The Congo Basin contains roughly 300 million hectares, while Southeast Asia (mainly in Indonesia and Malaysia) comprises around 240 million hectares (Food and Agriculture Organization, 2011).

Tropical forests are characterized by a rich yet vulnerable biodiversity. They are home to 62% of global terrestrial vertebrate species, which is more than twice the number of any other terrestrial habitat, and despite covering only 10% of the Earth's land surface (Pillay et al., 2022). In the Amazon alone, approximately 16,000 tree species coexist, with a density ranging from 120 to more than 200 species per hectare (Baraloto et al., 2013). Additionally, they are a crucial carbon reservoir that stores more than half (55%) of the world's forest carbon stock (Pan et al., 2011). Tropical forests can also be defined by their history of disturbances. Old-growth forests, also known as primary forests, have reached a stable stage of their growth and have not experienced any major disturbance in their development. Because of their maturity, they usually store significant amounts of carbon and exhibit a diverse biological composition (Molina-Valero et al., 2021). Conversely, secondary forests are forested areas that have regenerated through natural processes after strong disturbances, either human-caused or not (Chokkalingam & De Jong, 2001). Since their structure and composition have been modified compared to the preceding primary forests, they are usually not as diverse as old-growth

forests. Additionally, during natural regeneration, the carbon stocks of secondary forests take several decades to return to their previous levels (Foody et al., 1996).

1.2.2 A threatened carbon stock

The goods and services that tropical forests provide place these ecosystems at the center of critical global environmental and socio-economic issues (Seymour & Busch, 2016). One of the main socio-economic challenges associated with tropical forests is the pressure from a rapidly growing global population, which is projected to exceed 9 billion by 2050 according to the United Nations. To meet the resulting increase in demand for food, between 6.4 and 8.8 million hectares of tropical forest are converted into agricultural land each year (Pendrill et al., 2022). In the period between 1990 and 2015, 62% of deforestation in Africa and 78% in Latin America were driven by agriculture (Carter et al., 2018). When focusing on the 2011-2015 period, agriculture accounted for more than 90% of global tropical deforestation, yet only about half of these deforested lands were ultimately converted into productive agricultural land (Pendrill et al., 2022). In Southeast Asia alone, about 8 million hectares of forest have been transformed into oil palm plantations over the past 50 years (Koh & Wilcove, 2008). Alongside these considerations, forested areas are also of significant economic interest because of the harvesting and exploitation of wood. The natural resources extracted contribute to the socio-economic development of timber industries that involve many direct and indirect stakeholders (Dauber et al., 2005).

When tropical forests are cleared or degraded, the carbon stored within them can be released back into the atmosphere, mainly through the burning of vegetation and the decomposition of organic matter (Mitchard, 2018). Furthermore, in addition to exacerbating global warming through carbon release, the loss of trees also reduces the carbon sequestration capabilities of the forests. Therefore, human economic activities are not only major sources of carbon emissions, but also an increasing threat to carbon stocks of tropical forests through deforestation and forest degradation. These effects are even more intensified by the fact tropical forests are mainly located in fast developing countries, which are currently experiencing rapid economic and demographic growth. According to the key findings of the Global Forest Resources Assessment 2020, long-term management plans are implemented for less than 25% of forests in Africa and less than 20% in South America (Food and Agriculture Organization, 2020). As reported by Global Forest Watch, Brazil alone lost 30.7 million hectares of primary forests between 2002 and 2023, representing a 9% decrease in its primary forest area over this period.

Globally, the total carbon stock in forests decreased from 668 Gt in 1990 to 662 Gt in 2020 (Food and Agriculture Organization, 2020). There were notable regional differences in this trend: while the carbon stocks increased significantly in Asia, Europe, and North America, they decreased considerably in South America and Africa. These degradations and losses of tropical forests have profound implications on the global carbon cycle and climate regulation: the resulting carbon dioxide emissions in the atmosphere have indeed caused a rise in temperatures on a global scale, as highlighted by the Intergovernmental Panel on Climate Change (IPCC) in their latest and previous reports.

1.2.3 Forest biomass: a key driver of global carbon stocks

Over 40% of the world's vegetation carbon stocks are found in tropical forests (Mills et al., 2023). The carbon stored in a tropical forest is contained in its biomass, and biomass is defined as the mass of the organic matter of a living organism. A widely assumed value for the carbon content of dry wood is 50% (Matthews, 1993). The biomass levels of tropical forests are therefore the key driver to understanding and quantifying the carbon sequestration capabilities of these ecosystems.

However, forests are complex ecosystems that contain carbon not only in trees but in several other compartments. Of the total carbon contained in the world's forests (i.e., 662 Gt in 2020), 300 Gt are found in soil, 295 Gt in living biomass, and 68 Gt in dead wood and litter (Food and Agriculture Organization, 2020). Most forest carbon is thus contained in the living biomass (44%) and soil organic matter (45%). The living biomass can itself be divided into two mains compartments: the aboveground biomass (AGB) and the belowground biomass (BGB). In general, living AGB is the most investigated because it is openly exposed and therefore easier to access. BGB can later be derived from AGB using existing models (Cheng & Niklas, 2007).

AGB is defined as the quantity of aerial living vegetation, including stem, stump, branches, bark, seed, and foliage. It is the reference variable for characterizing carbon stored in forests because it is the most directly impacted by forest destruction. However, converting AGB in equivalent carbon mass remains a challenging scientific question, as the relationships between these two parameters depend on tree species and biomes. For example, in tropical species, wood carbon content ranges from 41.9% to 51.6%, while other tree tissues also contribute to carbon storage (Martin & Thomas, 2011; Thomas & Martin, 2012).

The most accurate method to measure AGB is through field campaigns that involve cutting down trees to weight them (Huynh et al., 2021). However, this approach is both destructive, as it requires cutting down trees, and restrictive, because it is labor-intensive and unsuitable

for large-scale studies. Therefore, it is necessary to find ways to estimate AGB rather than directly measuring it. Other variables can be collected in a non-destructive manner to characterize forests. In this perspective, many studies have established allometric equations to estimate the AGB of tropical forests from their measurable characteristics (Asner & Mascaro, 2014; Chave et al., 2005, 2014). Among various factors such as diameter at breast height (DBH) and wood density, canopy height stands out as a key parameter for estimating AGB.

1.2.4 Canopy height: a key indicator of forest biomass

The canopy represents the upper layer of the forest formed by the crowns of trees. Since AGB represents a quantity of aerial organic matter, forest structural parameters are generally well-suited descriptors to derive AGB. In particular, canopy height, which reflects the vertical extent of the forest from the ground to the top-of-canopy, is closely related to AGB. It is important to distinguish between tree height and canopy height, as they correspond to different measurements and serve different purposes.

Tree height refers to the vertical measurement of an individual tree from its base to the highest point of its crown. Tree height is used with other tree parameters for the characterization of individual trees, for example to estimate the AGB of single trees through geometric approaches (Chave et al., 2005). These individual estimates can then be grouped to produce AGB estimations for entire stands. It has been widely demonstrated that tree height is a crucial allometric factor that must be included in forest biomass estimates to reduce the estimation error of tropical carbon stocks and emissions due to forest loss (Feldpausch et al., 2012). Methodologies based on tree height are well-suited for describing inventory plots and characterizing forest stands at the scale of a plot. However, tree-level approaches are limited when it comes to studying large areas with huge numbers of individuals.

In contrast, canopy height is defined as the height of the highest vegetation components above ground level (Nakai et al., 2008). This measure reflects the collective height of a stand within a given area and is not attached to a given tree. Therefore, it provides a broader perspective on forest height and is not used in the same way as individual tree height. Canopy height has a direct and well understood relationship to aboveground carbon storage (Lefsky et al., 2002, 2005) and its inclusion in allometric models to derive AGB results in reduced estimation errors and improved accuracies (Lima et al., 2012). Moreover, allometric relationships based exclusively on canopy height may even be used in some cases to estimate AGB (Lefsky et al., 2005). Characterizing tropical forest canopy height not only presents an interest for AGB estimation, but also for addressing various other ecological questions. For example, canopy height can also provide insights into the factors influencing climatic patterns (Hunter, 1999).

Indeed, vegetation structure, especially vegetation height, can influence light availability, local rainfall, and air flows, which in turn impact local climatic and environmental conditions. Ultimately, these conditions affect the ecological habitat of different species, such as birds for example (Maurer & Whitmore, 1981; Robinson & Holmes, 1982).

In order to quantify AGB levels of tropical forests around the world at large scales (regional and global), and with the ultimate goal of understanding their carbon storage capabilities, there is an important need for accurate measurements and estimates of the canopy height of tropical forests. Given the vast and hardly accessible nature of tropical forests, traditional field-based methods are insufficient for capturing large-scale data. In this perspective, remote sensing technologies offer the ability to obtain high-resolution, consistent, and comprehensive data over extensive areas. They have become essential means for monitoring, modeling, and managing the dynamics of tropical forests worldwide.

1.3 Contribution of remote sensing

1.3.1 Remote sensing and forest structure

Remote sensing has become an indispensable tool for addressing tropical forest monitoring issues and providing policymakers with relevant information (Fuller, 2006; Trisasongko & Paull, 2020). The range of available data has considerably expanded over the last few decades and a great variety of sensors are now used to study forest ecosystems around the world (Boyd & Danson, 2005; Sanchez-Azofeifa et al., 2017; Wulder et al., 2018). These sensors feature a wide range of technologies and are embedded in various dedicated airborne and satellite vehicles. The information collected through remote sensing systems cover extensive areas, sometimes difficult to access if not completely inaccessible, and provide detailed information on all kinds of ecosystems. The proliferation of available data has also led to the development of new methodologies for processing them. Remote sensing data are now essential for studying and monitoring Earth's natural ecosystems, and forests in particular. These data combined with field measurements are used to extract quantitative and descriptive information on tropical forest ecosystems at different time and space scales, enabling efficient monitoring (Murrins Misiukas et al., 2021; Portillo-Quintero et al., 2021) and a better understanding of their dynamics (Da Ponte et al., 2015).

In particular, information on forest structure is a key descriptor of forest ecosystems (McElhinny et al., 2005). The term structure refers to the three-dimensional spatial distribution of vegetation elements such as trunks, crowns, and leaves. These structural characteristics of forest ecosystems are directly related to variables of interest such as AGB (Chave et al., 2005). Canopy height is one of the most important structural parameters of tropical forest ecosystems

as it is linked to many ecological processes, and particularly AGB levels. It can be measured directly using specialized remote sensing techniques or estimated through models based on remote sensing data. On one hand, direct measurements, while highly accurate, are often costly and limited in spatial coverage, making them insufficient for extensive mapping applications. On the other hand, models that estimate canopy height from remote sensing data require ground-truth reference data to build reliable empirical predictions. These reference data can originate from either remote sensing measurements (with the measurement errors associated with the data acquisition and processing parameters), or in-situ field measurements collected over inventory plots. The development of empirical models for predicting canopy height relies on the integration of remote sensing data with sampled field measurements of canopy height. Once implemented, these models can be applied across the entire study area. The precision of field measurements, no matter how they were obtained, is therefore critical for the relevance of the models and the accuracy of their results.

Light Detection and Ranging (LiDAR) technology is particularly well-suited for characterizing the structure of tropical forests (Lim et al., 2003) and measuring canopy height over large areas (Simard et al., 2011). For this reason, airborne and spaceborne LiDAR data are commonly used for the calibration, implementation, and validation of canopy height prediction models.

1.3.2 LiDAR technology for canopy height estimation

LiDAR is an active remote sensing technique based on the emission and reception of laser pulses. It is used to measure the distance between the emitter/receiver and the target, based on the propagation speed of the laser pulse (speed of light) and its travel time (i.e., the time elapsed between emission of the pulse and reception of an echo).

For the characterization of forest vertical structure and especially canopy height, LiDAR is particularly well-suited because it measures the geometrical structure of the environment through the emission and reception of laser pulses. The acquisition process happens as follows: a vehicle equipped with a three-dimensional positioning system (GPS for position and inertial unit for attitude) carries a LiDAR sensor and a laser beam deflection mechanism (scanning mechanism). The sensor emits a laser pulse, usually in the near-infrared wavelengths, which travels through the atmosphere and interacts with targets on the ground. This interaction generates a backscattering signal which is recorded by the sensor and digitally processed. The shape of the return signals depends on several factors, such as the radiometric and geometric properties of the objects in the acquisition environment, the specificities of the acquisition environment itself, and the physical characteristics of the laser beam. Digital

processing of this signal can result in either a set of return pulses (discrete-return LiDAR), or a continuous signal (full-waveform LiDAR).

Discrete-return LiDAR systems capture multiple echoes for each emitted pulse and generate three-dimensional point clouds that detail the vegetation structure from the top-of-canopy down to the ground. Each point in the georeferenced 3D point clouds corresponds to an echo resulting from the interaction between the laser pulse and the environment. A distinction is made between single echoes (single returns) and multiple echoes (several backscattering returns recorded for the same laser pulse). By analyzing the spatial distribution of these 3D point clouds, various approaches can be employed to extract forest biophysical variables (Véga et al., 2016). Full-waveform systems capture the complete profiles of return signals by sampling them at fixed time intervals. The resulting waveforms are geolocated one-dimensional temporal signals that represent the vertical structure of the vegetation at specific locations. From these waveforms, a variety of signal metrics and forest parameters can be extracted (Pirotti, 2011).

Airborne and spaceborne LiDAR systems are the primary tools used to characterize forest vertical structures. Airborne sensors offer the advantage of extremely high-resolution data and allow producing a large number of return points over a given area, and therefore accurate measurements of canopy height. However, this comes with significant financial costs for data users, and their use is restricted to limited areas compared to the size of tropical forest areas around the world. In contrast, spaceborne LiDAR provides freely available data with global coverage, but with a lower data density. The data are usually provided in the form of sparsely and unevenly distributed footprints on the ground, with various footprint sizes depending on the sensor. Additionally, the high operational altitudes of spaceborne LiDAR make it more vulnerable to difficult environmental conditions (Fayad et al., 2021). Currently, airborne LiDAR is still the most precise remote sensing system for acquiring detailed site-level data on canopy height (A. Ferraz et al., 2016; Leitold et al., 2015), but the use of spaceborne LiDAR has proven increasingly viable to produce accurate estimates of canopy height.

The Geoscience Laser Altimeter System (GLAS) aboard the Ice, Cloud, and land Elevation Satellite (ICESat) was the first satellite system to deliver data over extensive areas at a relatively low cost to users. GLAS data have been used effectively to produce accurate estimates of canopy height and AGB, particularly in regions with smooth topography and uniform forest conditions (Fayad et al., 2014; Lefsky, 2010; Lefsky et al., 2005; Pourrahmati et al., 2015). However, in areas with dense vegetation, steep terrain, or heterogeneous vegetation cover, GLAS data are more difficult to leverage for canopy height estimation

(Duncanson et al., 2010). The Advanced Topographic Laser Altimeter System (ATLAS), embarked on ICESat-2, the successor of the original ICESat, is a photon counting system that is still operating and that has high potential for retrieval of forest canopy height at global and regional scale, even though its primary focus is not on detailed forest structure. Nevertheless, it was successfully used to estimate canopy height in a wide variety of biomes and in complex terrain conditions (Jiang et al., 2021; Li et al., 2020; Neuenschwander & Magruder, 2019; Zhu et al., 2020). The most recent operating spaceborne system providing LiDAR data globally is the Global Ecosystem Dynamics Investigation (GEDI) attached to the International Space Station (ISS). GEDI is a full-waveform LiDAR sensor that was specifically designed to measure the vertical structure of the Earth's forest ecosystems (Dubayah et al., 2020). GEDI data has been widely used to accurately estimate canopy height across a variety of forest ecosystems (Adam et al., 2020; Dorado-Roda et al., 2021; Rajab Pourrahmati et al., 2023), and even directly as ground truth reference data for calibrating canopy height prediction models (Lang et al., 2023; Potapov et al., 2021; Schwartz et al., 2023).

Spaceborne LiDAR systems are therefore efficient tools for the retrieval of canopy height in tropical biomes. However, their sparse measurements do not exhibit a sufficient spatial density for continuous mapping of canopy height. Continuous mapping is indeed essential to generate accurate AGB products and assess the carbon stocks of tropical forests. Consequently, there is a need to combine spaceborne LiDAR measurements with other complementary remote sensing sources to achieve comprehensive and precise canopy height mapping.

1.3.3 Canopy height mapping at regional and global scales

Ground truth data in the form of sampled inventory plots and reliable canopy height measurements over small areas or sparse footprints on the ground are valuable information, but in the end, they are not enough alone for the ultimate goal of continuous mapping of canopy height over extensive areas. LiDAR offers high accuracy in estimating canopy height, but it faces limitations in its horizontal coverage, with airborne LiDAR being constrained by its limited spatial extent and spaceborne LiDAR by its lower acquisition density. Continuous mapping of the biomass of tropical forests is necessary to quantify their carbon stocks. To achieve this goal, LiDAR information can be combined with other data sources that provide continuous information related to canopy height over the areas of interest.

In this perspective, several studies have demonstrated the benefits of combining LiDAR data with image information from optical (Wang et al., 2020) or radar sensors (Carcereri et al., 2023) to produce canopy height and AGB maps. Many studies have chosen to include all these remote sensing data together to benefit from their complementarity (Fagua et al., 2019), and

the additional use of other ancillary data related to environmental description has also shown promising results (Lahssini et al., 2022).

Optical sensors are widely employed to characterize state parameters of forest ecosystems. The radiometric data obtained from optical systems, known as reflectance, contains spectral information that can be used to assess stand compositions, tree species, and vegetation biophysical parameters (Grabska et al., 2019; Lemaire et al., 2008). One of the key benefits of current satellite optical data is its ability to provide global continuous coverage and high-resolution multispectral imaging, both spatial and temporal. Many optical data products from a variety of satellite platforms are publicly and easily available to data users, allowing for long-term and large-scale studies of forest ecosystems. However, there are also certain limits to consider. Cloud cover, particularly in tropical regions, may restrict data collection as optical signals cannot penetrate through clouds. Moreover, optical sensors are passive systems that are impacted by solar illumination conditions, limiting data collection to certain hours of the day. More specifically in the context of tropical forests, optical sensors are not able to penetrate canopy cover, making it difficult to investigate variables like canopy height or AGB that are related to the whole vertical structure of the vegetation. Consequently, combining optical data with other remote sensing techniques is necessary when it comes to the characterization of forest structural parameters.

Synthetic aperture radar (SAR) employs microwaves, which have a lower frequency than LiDAR, and presents many advantages for forest monitoring. Its main advantage is its capacity to go through cloud cover and acquire data in the night, making data collection possible in any weather conditions and at any time of the day. Radar signals are characterized by longer wavelengths that allow a moderate penetration of the forest canopy, with longer wavelengths showing deeper penetration. SAR data is also extremely sensitive to changes in canopy structure, which makes it ideal for identifying disturbances, deforestation, and degradation (Bouvet et al., 2018; Rahman & Sumantyo, 2010; Watanabe et al., 2018). Using different frequency bands is interesting for forest applications because they complement each other (Sothe et al., 2022). The L-band has a longer wavelength, which gives it better penetration capabilities (Baghdadi et al., 2015). The information on the lower layers of the canopy is particularly useful when predicting forest structural variables (Mermoz et al., 2014). The C-band has less penetration capabilities due to its shorter wavelength, but it provides information about the upper layers of the canopy (Fagua et al., 2019). In the coming years, the European Space Agency (ESA) plans to launch the BIOMASS mission which will collect SAR data in the P-band to derive worldwide estimates of forest biomass and height (Quegan et al., 2019). The wavelength of P-band is much longer than C-band or L-band, allowing for even more

penetration capabilities, especially in the dense context of tropical forests. Additionally, specific radar techniques such as polarimetric interferometric SAR (PolInSAR) and tomographic SAR (TomoSAR) have been successfully used to improve canopy height estimates (Garestier et al., 2008; Ho Tong Minh et al., 2016).

While LiDAR is the best technology to measure or estimate canopy height, optical and radar can provide complementary information to overcome the inherent limitations of LiDAR data acquisition. The interplay between optical, radar, and LiDAR is therefore leveraged to extrapolate LiDAR measurements and produce continuous canopy height maps. In these multimodal approaches to estimating canopy height, various techniques can be used to merge data from different sources and build canopy height maps, ranging from simple statistical methods to more advanced deep learning models. The most basic approach involves linear regression, where LiDAR-derived canopy height measurements are used as the dependent variable (i.e., the target variable), and data from other remote sensing sources and ancillary datasets serve as predictors. This method can provide reasonable estimates when the relationships between the variables are straightforward and linear (Lefsky, 2010). However, it may be limited in capturing the complexity of forest ecosystems, as interactions between variables are often non-linear. To improve accuracy, more sophisticated models like Random Forest (RF) are frequently used. RF is an ensemble learning method that builds multiple decision trees based on different subsets of the data. This design offers a robust way to model complex relationships, both linear and non-linear. This approach has been widely used to produce canopy height maps with good accuracies (Simard et al., 2011). While these approaches are effective, they do not necessarily account for spatial dependencies, which are crucial to consider when mapping a continuous variable like canopy height. To address this, spatial interpolation techniques such as kriging are often employed alongside canopy height regression models, in order to integrate the spatial autocorrelation (Fayad et al., 2016; Hudak et al., 2002).

As the volume of remote sensing data continues to grow, deep learning approaches have been developed to facilitate the analysis of large and complex remote sensing datasets such as time series (Masolele et al., 2021), multi-source data (Schwartz et al., 2023), and multi-resolution analyses (Yu et al., 2021). Studies have shown that deep learning models can provide improvements in the accuracy of canopy height and AGB estimates compared to more common methods (Shah et al., 2020). Neural networks, particularly Convolutional Neural Networks (CNNs), can automatically learn complex patterns from large and diverse datasets. They can process data from multiple sensors simultaneously and inherently integrate the local and global spatial context in the regression process. They are a powerful tool for generating

continuous and high-resolution canopy height maps, which has enabled the production of several global data products in the recent years (Lang et al., 2023; Potapov et al., 2021).

1.4 Scientific framework

1.4.1 Objectives and research questions

In the context of climate change, a strong need to understand the carbon cycle and develop new approaches to accurately estimate carbon stocks at regional and global scale has been widely echoed by the scientific community. In recent years, scientists have noted the urgent need for new technologies and methods to observe forest biomass stocks on a global scale, providing data that will not only help to better understand the global carbon cycle, but also more directly reduce greenhouse gas emissions by reducing deforestation and forest degradation. Spatial remote sensing, and particularly LiDAR remote sensing, represents an interesting and complementary solution to other remote sensing techniques for mapping forest biomass at large scales. To achieve this, canopy height is a measurable structural variable that has emerged as the primary proxy for deriving AGB information.

In recent years, numerous studies have exploited the complementary nature of radar and LiDAR data. Optical data are also used to estimate canopy height, often as complement to radar and LiDAR data. However, two main limitations were identified in these works, which mostly used L-band radar satellite data and ICESat-1 LiDAR data. Firstly, past satellite LiDAR missions such as the GLAS sensor provide data over large areas, but they are unable to acquire data with sufficient spatial density. They indeed exhibit a low density of spaceborne LiDAR footprints, for example around 0.5 points/km² over a territory like French Guiana. Moreover, L-band radar signals display a low sensitivity to high biomass levels, with a saturation at around 150 megatons (Mt) per hectare, a threshold that is well below normal biomass levels in tropical biomes, and particularly in French Guiana. Optical data is also concerned by saturation issues, as it has been observed that the canopy reflectance of tropical forests saturates at a certain height and then no longer changes.

To meet these challenges, the GEDI spaceborne LiDAR mission, launched in 2018, has been designed to produce data for mapping global forest biomass. Other radar missions are planned for the future to address the observed limitations, notably ESA's BIOMASS P-band radar mission (planned for launch in 2024) and the NISAR L-band radar mission (also planned for launch in 2024) supported by the National Aeronautics and Space Administration (NASA) and the Indian Space Research Organization (ISRO).

The main scientific objective of this doctoral thesis is to exploit the new GEDI data for canopy height mapping. The data provided by the GEDI mission have a higher spatial density than those previously provided by ICESat-1. In addition, they are characterized by a finer spatial resolution, with ground footprints of 25 m in diameter compared to 70 m for ICESat-1 data. These parameters allow producing canopy height products with greater precision and finer spatial resolution (of the order of 10 m). GEDI data are used in complement to existing C-band and L-band radar data, as well as multispectral optical data. In this context, new approaches to estimating canopy height, based on multi-sensor complementarity, are studied.

From this primary objective, several key research questions arise concerning the use of GEDI data for canopy height mapping, particularly in the context of tropical biomes such as French Guiana. The availability of GEDI data raises important questions about their quality and their potential for enhanced canopy height mapping. Assessing the utility and reliability of GEDI measurements is a mandatory step to understand their interest and this assessment is done considering various levels and characteristics of the data:

1. How do the physical properties of the GEDI signal and the subsequent processing parameters impact the measurements? GEDI measurements, like those of any instrument, are influenced by the physical interactions of the signal. The laser signal's properties and the data processing chain directly shape the resulting waveforms and metrics extracted. Understanding their impact on canopy height measurements is essential for accurate interpretation and application of these data.
2. How do environmental factors, such as terrain configuration and vegetation structure, affect GEDI measurements? In tropical biomes, factors like terrain and vegetation structure can significantly influence the measurements retrieved by GEDI. Quantifying and understanding these impacts are crucial for developing correction strategies and improving the accuracy of canopy height maps.

Building on the insights gained from assessing GEDI data, the next step is to focus on developing effective methodologies for canopy height mapping that can fully exploit the potential of GEDI data, in combination with other data sources:

3. How can effective data fusion models be developed to leverage multiple data sources for canopy height mapping? The integration of multiple remote sensing data sources (optical, radar, LiDAR) is key to exploit the potential of GEDI data for canopy height mapping across extensive territories like French Guiana. The way these various data products are integrated into prediction models is also of paramount importance to maximize their potential and to extract their features related to canopy height. This question also addresses the selection of relevant additional environmental descriptors

that can enrich the models and improve predictions, as well as the understanding of how and why these data contribute positively to the results.

4. How do the acquisition parameters of GEDI data influence the results of canopy height mapping? GEDI data are characterized by sparsely and unevenly distributed measured points. Despite the availability of methods to manage such data, accounting for the spatial dependencies in the canopy height mapping process remains essential as canopy height is a continuous variable and its spatial dynamics arise from many complex environmental factors. Understanding and taking into account this spatial variability and autocorrelation are important steps to generate more accurate and reliable canopy height products.

These questions guide the research performed in this doctoral thesis entitled ‘Potential and Challenges of Spaceborne LiDAR for Multi-Source Canopy Height Mapping in Tropical Forests’. A canopy height map of French Guiana at a spatial resolution of 10 m is presented.

1.4.2 Dissertation plan

This dissertation is organized in six chapters and is presented in the form of journal articles each covering a research question. After a general introduction and presentation of the research objectives in Chapter 1, Chapter 2 focuses on studying the impact of GEDI signal properties and data processing on measurements. It is a peer-reviewed article published on 10 December 2022 in the international journal ‘Remote Sensing’ (MDPI). Chapter 3 covers the environmental impact on GEDI measurements. It is a peer-reviewed article published on 14 May 2024 in the international journal ‘Canadian Journal of Remote Sensing’ (Taylor & Francis). Chapter 4 presents data fusion and model building for canopy height mapping in French Guiana. It is a peer-reviewed article published on 25 November 2024 in the international journal ‘Frontiers in Remote Sensing’ (Frontiers). Chapter 5 explores the impact of GEDI acquisition characteristics on canopy height maps and addresses the spatial autocorrelation of the measurements. It is structured as a research article intended for submission. Chapter 6 summarizes the main conclusions and outlines the perspectives of this doctoral thesis.

Chapter 2

Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests

This chapter presents the first study in the scope of the dissertation and aims at addressing the first research question: how do the physical properties of the GEDI signal and the subsequent processing parameters impact the measurements? Assessing GEDI data in a tropical context is key to the broader objective of this doctoral thesis which aims at using these data for canopy height mapping.

GEDI is a spaceborne LiDAR sensor that was specifically designed to measure the vertical structure of the vegetation around the world and to overcome the limitations of previous LiDAR missions. As a full-waveform LiDAR system, the vertical structure of the vegetation is measured through return waveforms, which are one-dimensional temporal signals resulting from the interactions between the emitted laser pulses and the vegetation they encounter. The GEDI Level 1B (L1B) Geolocated Waveforms product (GEDI01_B) provides geolocated corrected and smoothed waveforms that describe vegetation profiles within footprints on the ground and therefore contain information about the vertical structure of forests. Understanding the parameters that impact the GEDI L1B return waveforms is important to correctly interpret the subsequent canopy height values derived from these data. The shape and quality of these waveforms are influenced by several factors mainly linked to the physical properties of the laser signal, namely laser power and beam sensitivity. Laser power is the energy output of the emitted pulses and GEDI operates with two power levels (full-power and half-power). Beam sensitivity is defined as the maximum canopy cover that can be fully penetrated by the laser signal. These properties affect how the laser interacts with the vegetation, which in turn impacts the retrieval of vertical profiles from GEDI data. Indeed, to accurately describe the complete vertical structure of the vegetation, it is necessary for the GEDI signal to reach the ground. In tropical ecosystems that are usually characterized by high AGB levels, the ability of the signal

to penetrate dense vegetation all the way to the ground and retrieve detailed information from deep canopy layers is a major challenge to measure canopy height with precision.

The received waveforms contained in the L1B product describe the vertical profiles of the vegetation, and this information can be synthesized in the form of metrics derived from the one-dimensional signals. In this perspective, the GEDI Level 2A Geolocated Elevation and Height Metrics product (GEDI02_A) provides waveform interpretation and extracted metrics from each L1B waveform, including ground elevation, canopy top height, and relative height metrics. These metrics are used to estimate canopy height, and they are produced from the waveforms using six possible signal processing configurations. Consequently, the values retrieved are sensitive to the processing configuration used. The settings of each configuration, such as smoothing level, noise filtering, and threshold selection, significantly influence the extracted metrics and the subsequent canopy height retrievals. Therefore, it is important to understand and assess these processing configurations to select the most appropriate in a challenging environment like dense tropical forests.

In summary, this chapter focuses on analyzing how the physical properties of the GEDI signal and the waveform processing configurations influence the canopy height estimates retrieved from GEDI data over dense forests in French Guiana (South America) and Gabon (Africa). It lays the groundwork for improving canopy height estimates by addressing the first research question of this dissertation. By assessing these factors, this study establishes guidelines for selecting optimal data that enhance the accuracy of canopy height estimates in tropical regions, with the goal of ensuring the reliability of the canopy height maps produced in this doctoral thesis.

remote sensing

MDPI

Article

Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests

Kamel Lahssini [1,*], Nicolas Baghdadi [1], Guerric le Maire [2,3] and Ibrahim Fayad [4,5]

1 UMR TETIS, INRAE, 34093 Montpellier, France
2 UMR Eco&Sols, CIRAD, 34398 Montpellier, France
3 Eco&Sols, Université de Montpellier, CIRAD, INRAE, IRD, Institut Agro, 34093 Montpellier, France
4 Kayrros SAS, 75009 Paris, France
5 Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS9 UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France
* Correspondence: kamel.lahssini@inrae.fr

Abstract: LiDAR technology has been widely used to characterize structural parameters of forest ecosystems, which in turn are valuable information for forest monitoring. GEDI is a spaceborne LiDAR system specifically designed to measure vegetation's vertical structure, and it has been acquiring waveforms on a global scale since April 2019. In particular, canopy height is an important descriptor of forest ecosystems, as it allows for quantifying biomass and other inventory information. This paper analyzes the accuracy of canopy height estimates from GEDI data over tropical forests in French Guiana and Gabon. The influence of various signal acquisition and processing parameters is assessed to highlight how they impact the estimation of canopy heights. Canopy height models derived from airborne LiDAR data are used as reference heights. Several linear and non-linear approaches are tested given the richness of the available GEDI information. The results show that the use of regression models built on multiple GEDI metrics allows for reaching improved accuracies compared to a direct estimation from a single GEDI height metric. In a notable way, random forest improves the canopy height estimation accuracy by almost 80% (in terms of RMSE) compared to the use of *rh_95* as a direct proxy of canopy height. Additionally, convolutional neural networks calibrated on GEDI waveforms exhibit similar results to the ones of other regression models. Beam type as well as beam sensitivity, which are related to laser penetration, appear as parameters of major influence on the data derived from GEDI waveforms and used as input for canopy height estimation. Therefore, we recommend the use of only power and high-sensitivity beams when sufficient data are available. Finally, we note that regression models trained on reference data can be transferred across study sites that share identical environmental conditions.

Keywords: LiDAR; GEDI; French Guiana; Gabon; canopy height; tropical forests

check for updates

Citation: Lahssini, K.; Baghdadi, N.; le Maire, G.; Fayad, I. Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests. *Remote Sens.* **2022**, *14*, 6264. https://doi.org/10.3390/rs14246264

Academic Editor: Yanjun Su

Received: 4 November 2022
Accepted: 6 December 2022
Published: 10 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Tropical forest ecosystems play a major role in natural balances as they contribute to the global carbon storage. They act as natural carbon dioxide sinks and contain more than 40% of the terrestrial carbon stock [1]. In order to understand their capability for carbon sequestration, standing aboveground biomass (AGB) is a fundamental factor that allows for quantifying the amount of carbon dioxide that they can absorb. AGB is defined as the quantity of living vegetation above the soil, including stem, stump, branches, bark, seed and foliage. Several studies have defined allometric relationships linking the measurable characteristics of a forest to its AGB levels [2,3]. In particular, canopy height is a key parameter for biomass quantification, as most allometric equations rely on canopy heights for AGB estimation [4]. Many authors demonstrated that the inclusion of canopy heights in AGB prediction models allows for reducing the estimation errors and for reaching improved accuracies [5,6].

To estimate canopy heights at regional and global scales, Earth observation data has proven to be a valuable source of information, especially since it provides informative indicators over large areas that cannot be studied otherwise [7]. For the characterization of forest vertical structures, light detection and ranging (LiDAR) is the most suited technology, as it measures the geometrical structure of the environment by emitting laser pulses and by retrieving the return signals resulting from the interaction between the laser and the forest individual elements. Discrete return (DR) LiDAR systems record multiple echoes for each emitted pulse and provide 3D point clouds describing the vegetation structure from the top of the canopy to the ground. The statistical analysis of the 3D point clouds' spatial distribution then allows for the extraction of forest biophysical variables through different approaches [8,9]. Full waveform (FW) systems record the full profiles of the return signals by sampling them in fixed time intervals. The resulting received waveforms consist of geolocated one-dimensional temporal signals describing the vertical structure of the vegetation at given geolocations, from which diverse forest parameters can then be extracted [10]. In another perspective, optical systems are used to study state parameters of forest stands since they only measure the reflectance from the canopy's upper layer and cannot penetrate through the vegetation. For example, the radiometric data retrieved from optical sensors allows for characterizing stand compositions and tree species by analyzing the spectral information contained in the images [11,12]. In contrast, other active sensors such as synthetic aperture radar (SAR) use lower frequency electromagnetic waves than those used by LiDARs but still present advantages for the study of forest ecosystems. Because of their longer wavelengths, radar signals tend to penetrate the forest canopy and reach the ground. SAR polarimetry and tomography techniques can provide information on the vertical structure of forested areas. In particular, starting from 2023, the European Space Agency (ESA) BIOMASS mission will collect P-band SAR data and deliver global estimates of forest biomass and height [13]. In the context of canopy height estimation, optical and radar data can provide complementary information and enrich prediction models [14,15] even though LiDAR remains the most adapted technology to derive accurate structural indicators of forest ecosystems. However, unlike SAR, spaceborne LiDARs do not provide measurements with sufficient spatial density for accurate mapping of tree height and forest biomass. Thus, in the future, the complementarity between optical, radar and LiDAR data will be exploited to produce canopy height and forest AGB maps.

Airborne and spaceborne LiDARs are the two main systems that are of interest for the description of forest vertical structures. The main advantage of airborne sensors is the very high resolutions at which they operate (availability of a high number of returned points over a given surface), but in return, it implies high financial costs that restrict their usage to limited areas. On the contrary, spaceborne LiDAR data are freely accessible and provide information with global coverage. Their counterparts are a low density of available information (spatial coverage of about 4% of the Earth's surface for the latest system) [16] as well as their high operational altitudes that make them more sensitive to unfavorable environmental conditions [17,18]. There are currently two operating spaceborne systems providing LiDAR data globally: the Advanced Topographic Laser Altimeter System (ATLAS) onboard the Ice, Cloud, and land Elevation Satellite (IceSAT-2) and the Global Ecosystem Dynamics Investigation (GEDI) attached to the International Space Station (ISS). These systems emit laser pulses that pass through the atmosphere and interact with the objects on the Earth surface, thus producing a return signal by backscattering. They measure the profiles of the return signals by sampling them in fixed time intervals (e.g., 1 ns in the case of GEDI, which is equivalent to a 15 cm sampling distance) and differ by the technology they use. ATLAS is a photon counting system equipped with a 532 nm laser producing six distinct beams. Its main scientific objectives are, on one hand, to finely characterize ice sheets in the context of recent sea level change and, on the other hand, to collect measurements over forests to facilitate canopy height determination as a basis for estimating global biomass and biomass change [19]. GEDI is a FW LiDAR sensor that was specifically designed to measure the vertical structure of the Earth's forest ecosystems [16]. It has been acquiring

data since April 2019 through the use of three identical 1064 nm lasers. One laser is divided into two beams of half power (coverage beams) while the other two remain at full power (power beams). These coverage and power beams are slightly dithered to finally give eight parallel tracks of observations across a 4.2 km swath on the ground (four coverage beams and four power beams). Each laser emits 242 pulses per second and illuminates a 25 m circular footprint on the Earth's surface over which the vertical structure of the vegetation is measured through the recording and processing of received waveforms.

The characterization of forest parameters from GEDI data is particularly challenging in tropical biomes since these ecosystems are usually very dense and developed. The main factor impacting canopy height estimates is the signal's capability to penetrate through the vegetation to accurately describe the vertical structure of the vegetation all the way from the top-of-canopy to the ground [20]. Additionally, the extraction of relevant descriptive metrics from the GEDI-received waveforms requires the use of extraction algorithms that can be tuned differently depending on factors such as the forest type or the study site's location. Given the characteristics of the GEDI system and its processing chain, the objective of this paper is to assess the influence of various signal acquisition and processing parameters on the estimation of canopy heights over tropical forests in French Guiana and Gabon. GEDI-derived heights are compared to reference values from a canopy height model (CHM) derived from airborne LiDAR (ALS) data. Firstly, a comparative analysis of the performances of GEDI's predefined extraction algorithms is performed to understand how the algorithm tuning parameters can impact the waveform metrics used for canopy heights retrieval. Through this analysis, we aim at identifying the most adapted algorithm setting group for the processing of GEDI data in a tropical context characterized by dense forests and high AGB values. Then, in the same spirit, the influence of GEDI's beam type is studied to evaluate which laser configuration is most suited for the estimation of canopy heights. Coverage and power beams correspond to different energy levels for the emitted laser pulses, and therefore, the beam type has a major influence on the resulting return waveforms. Other signal parameters such as beam sensitivity and signal-to-noise ratio (SNR) are also taken into account to better evaluate how the signal physical parameters can affect the information contained in the received waveforms. Our goal is to analyze and understand the link between the accuracy of canopy height estimates and the aforementioned signal parameters that are linked to the laser penetration through dense vegetation. Finally, we test several regression approaches based on GEDI data to assess how the accuracies of canopy height prediction models can be improved through the use of multiple metrics derived from GEDI waveforms. A stepwise multilinear regression (called hereafter "stepwise regression for canopy height estimation", SRH) and a random forest (RF) are implemented to take advantage of the information contained in GEDI metrics. Moreover, to overcome the possible issues related to the generation of relevant metrics from the GEDI data [21], we test a metric-free approach through the use of convolutional neural networks (CNN) built directly onto the GEDI waveforms. The viability of model transferability across study sites is also analyzed to understand how reference data can be valued to build general canopy height estimators. Given the results of this comparative analysis, we aim at providing GEDI users with key inputs and recommendations on how to take advantage of their data depending on the specific configurations of their studies. GEDI information is quite rich and can be used in diverse manners to derive structural indicators describing forest ecosystems.

This paper is organized as follows. A description of Materials and Methods is given in Section 2. The Results are presented in Section 3, followed by the Discussion and Conclusions in Sections 4 and 5, respectively.

2. Materials and Methods

2.1. Study Areas

The study was conducted on four areas of interest over equatorial eco-climatic zones, specifically, two sites in French Guiana (South America) and two sites in Gabon (Africa).

Regarding French Guiana, the first site, the Nouragues Ecological Research Station, is located 120 km south of Cayenne (Figure 1a). It is a protected natural reserve characterized by a lowland moist tropical rainforest and a humid climate. AGB levels range from 200 to 600 Mg/ha [22]. The topography is heterogeneous, with a succession of hills ranging from 27 to 280 m above sea level (asl) and a granitic outcrop reaching 430 m asl. One hectare of forest can contain up to 200 different tree species [22]. Top-of-canopy height reaches a maximum value of 60 m with the average height around 40 m [23,24]. The second site is the Paracou Station and is located near the city of Sinnamary (Figure 1a). It is also characterized by a lowland moist tropical rainforest and a humid climate. The topography is quite flat (ground elevation ranging from 5 to 50 m asl) and homogenous. The number of tree species estimated through forest inventories is approximately between 140 and 160 species per hectare [22]. Top-of-canopy height reaches up to 45 m with the average value around 35 m [23,24].

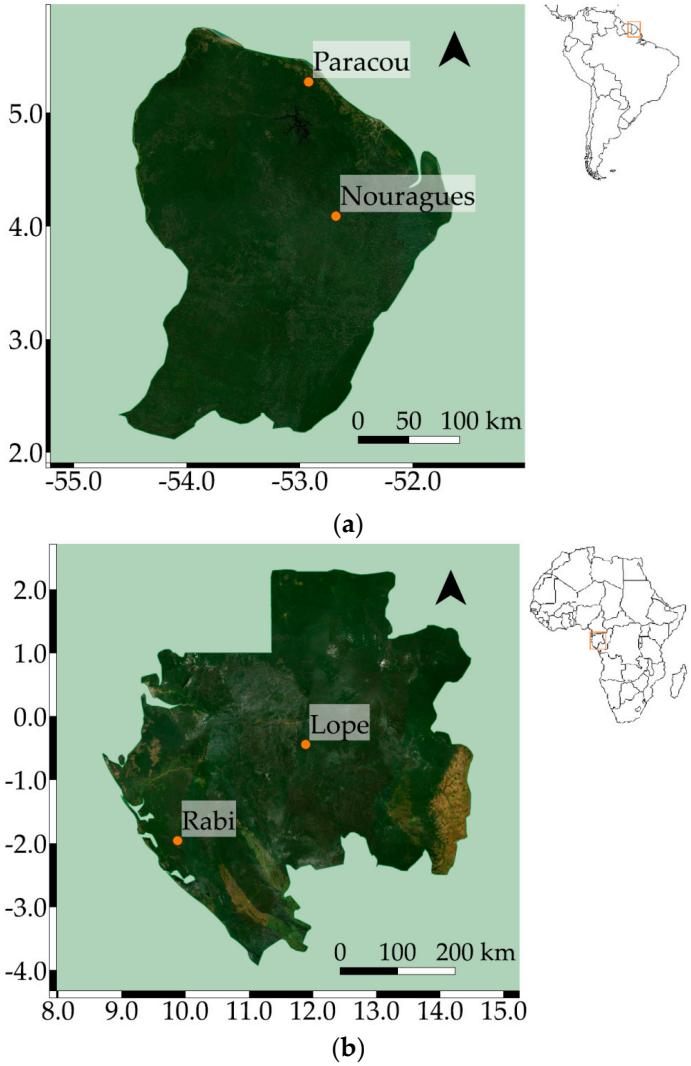


Figure 1. French Guiana (**a**) and Gabon (**b**) study sites (ESRI Satellite®).

As for Gabon, the first site is located in Lope National Park, one of the largest national parks in central Gabon (Figure 1b). The site is mostly characterized by compact and humid tropical forests with complex structure and high AGB values (superior to 400 Mg/ha) [25]. The topography of Lope National Park is diverse, varying between large flat plains and steep slopes. The ground elevation ranges between 200 and 600 m asl, and tree heights reach up to 60 m with an average value of about 40 m [25,26]. The second site is the Rabi

Forest Monitoring Plot, located near the Rabi Oil Concession (Figure 1b). It is characterized by dense humid rainforests and a fairly flat topography (ground elevation ranging from 31 to 82 m asl). Top-of-canopy height reaches up to 50 m, with the average value around 35 m [26,27].

The areas of interest considered in this study have been widely used in other similar works [22,25–27] due to the fact that they are specifically protected for research activities and conservation purposes. It can be fairly said that no major impacts or changes caused by anthropogenic activities have occurred recently. Nouragues is an ecological research station operated by the French National Center for Scientific Research (CNRS), and Paracou has been used as a research station to study Amazonian forest ecosystems since 1982. Similarly, Lope is a protected National Park in Gabon and was added to the World Heritage List by UNESCO in 2007. Finally, Rabi Forest Monitoring Plot is part of the Gabon Biodiversity Program and provides data for studies of carbon dynamics and biodiversity.

The four study sites correspond to zones with similar structural properties in terms of average heights and AGB levels. Nevertheless, they also differ in climate, biodiversity, elevation and slope, among other things. For the purpose of this study, all the data related to each site were merged together to create a single database describing a tropical context. The total area of natural equatorial forest covered by the merged study sites amounts to 100 km^2.

2.2. Data

2.2.1. GEDI Data

GEDI-processed data are produced by NASA's Land Processes Distributed Active Archive Center (LP DAAC). Datasets used in this study consist of both the L1B (Level 1) and L2A (Level 2) processing level. The L1B data product includes geolocated corrected and smoothed waveforms, as well as their ancillary parameters. The L2A data product provides waveform interpretation through footprint-level elevation and height metrics, including ground elevation, canopy top height and relative height (RH) metrics. The L2A product is derived from the received waveforms contained in the L1B product using six possible signal processing configurations (referred to as algorithm setting groups). Algorithm setting groups define thresholds and smoothing settings used to interpret the received waveform and therefore affect the height metrics produced in the L2A data product. Six distinct set of values are used as shown in Table 1. Therefore, for each GEDI footprint, six different canopy height values are available depending on the algorithm setting group used to process the received waveform.

Table 1. GEDI L2A product algorithm setting groups. σ is the standard deviation of the background noise level of the received waveform.

Algorithm Setting Group	Smoothing Width (Noise)	Smoothing Width (Signal)	Waveform Signal Start Threshold	Waveform Signal End Threshold
1	6.5 σ	6.5 σ	3 σ	6 σ
2	6.5 σ	3.5 σ	3 σ	3 σ
3	6.5 σ	3.5 σ	3 σ	6 σ
4	6.5 σ	6.5 σ	6 σ	6 σ
5	6.5 σ	3.5 σ	3 σ	2 σ
6	6.5 σ	3.5 σ	3 σ	4 σ

Over the four sites considered in this study, GEDI shots (i.e., a received waveform associated with a footprint on the ground) acquired in the period between April 2019 and August 2021 were collected. Since atmospheric conditions can negatively affect the measurements carried out by the GEDI sensor, several preliminary filters were applied in order to only keep usable and relevant shots for the analysis:

- Shots with no mode detected (*num_detectedmodes* = 0) were removed. A signal with no mode is pure noise and does not contain any information related to the vertical structure of the forest.
- Shots with a null SNR (*SNR* = 0) were removed. These shots are also pure noise. Information about the computation of SNR can be found in [28].
- Shots with an erroneous detection of the ground were removed. The quality assessment of the ground detection provided by GEDI data is performed using the corresponding Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM). If the absolute difference between the elevation of the lowest mode (*elev_lowestmode*) and the SRTM DEM (*digital_elevation_model_srtm*) is bigger than 100 m, then the shot is discarded.
- Shots with an incomplete waveform were removed. An incomplete waveform does not have a sufficient number of bins to be interpretable. If the end location of the usable portion of the waveform (*search_end*) is equal to the total number of bins in the waveform (*rx_sample_count*), then the waveform is considered incomplete.
- Shots with a distance between the canopy top and the ground return (*rh_100*) lower than 3 m were removed. These shots most likely correspond to bare soil or low vegetation and therefore are of no interest in this study.

After the filtering steps, and considering the availability of ground truth data, a total of 3864 GEDI footprints were retained for the combined analysis of the four sites described in Section 2.1 (Figure 2).

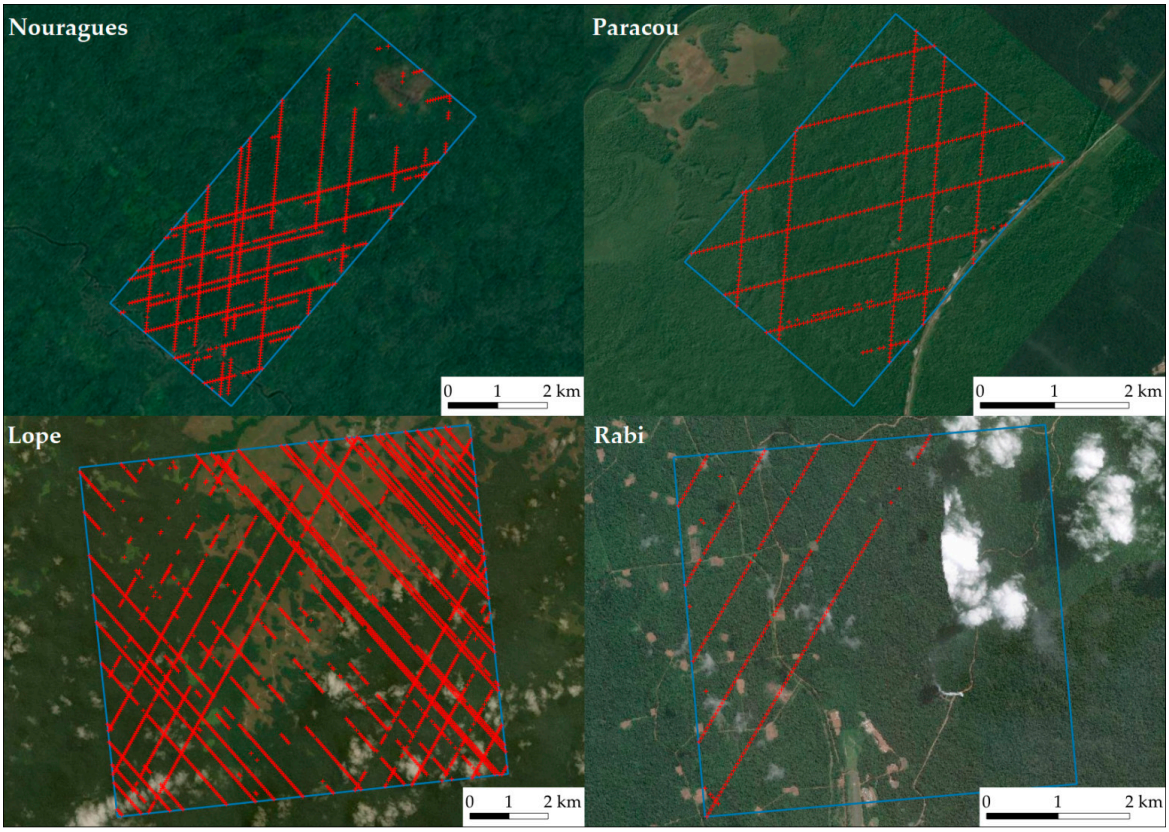


Figure 2. ALS acquisition areas (blue) and filtered GEDI footprints (red) over the four study sites (ESRI Satellite®).

2.2.2. Reference Data

In order to assess the accuracy of GEDI data to estimate forest heights, CHMs derived from ALS data were used as ground truth. Regarding the French Guiana study sites, two airborne LiDAR surveys were conducted by private contractors in 2008 and 2012.

Considering Gabon sites, the CHMs were produced in the framework of the AfriSAR mission, a collaboration between NASA and ESA conducted in the year 2016.

The four study sites are characterized by relatively high AGB levels (see Section 2.1) and tree heights. More than 80% of the GEDI footprints retained in this study correspond to heights superior to 30 m, and almost 90% of the footprints present a reference height superior to 20 m. In the context of old-growth tropical rainforest landscapes, some studies have assessed the canopy dynamics over large time spans. Based on two LiDAR datasets from 1998 and 2005 acquired over La Selva (a tropical biome in Costa Rica, Central America), Dubayah et al. [29] reported a net loss in height of −0.33 m for old-growth forests and a net gain of 2.08 m for secondary forests over a 7-year period. Kellner et al. [30] also analyzed two LiDAR datasets obtained nearly in the same periods over La Selva (1997 and 2006, corresponding to a 8.5-year time span) and found a mean change of height of −0.32 m for old-growth landscapes. Although the height does not change on average, it can vary locally by increasing or decreasing. In the same study, Dubayah et al. [29] reported local variations of about +/− 3 m over a 7-year period. Nevertheless, the dispersion that will be observed between GEDI-derived heights and airborne LiDAR CHMs is linked to many factors such as GEDI geolocation accuracy, measurement uncertainties and local variations of heights in the time span between acquisitions.

Specifically in French Guiana, according to calculation methods based on atmospheric inversion, AGB increases of about 0.5% each year [13]. If the totality of this increase were linked to tree heights, it would imply a growth of 5% in heights in a 10-year time span, resulting in a 1.5 m augmentation in this time interval (5% of 30 m, if we consider 30 m as the reference height of mature forest environments in French Guiana). In reality, in old-growth forest environments, the increase in AGB levels is mostly due to the increase in basal area rather than tree height [31,32]. Additionally, secondary forests account for a part of this AGB increase. Therefore, the actual height increase in old-growth forest ecosystems such as the ones in Paracou and Nouragues should be significantly smaller than 1.5 m in the period between the airborne LiDAR surveys and the GEDI acquisitions (approximately 10 years).

The CHM for each study site was obtained by subtracting the digital terrain model (DTM) from the digital surface model (DSM), which are both derived from the ALS data. Reference CHMs consist of raster datasets with a grid size (spatial resolution) of 1 m. Details on canopy model construction can be found in [33,34]. Considering all four sites, the total forest area covered by reference data for tree heights equals 100 km^2.

2.3. Methodology

The scope of the work presented in this study can be divided into two main tasks. Firstly, assuming that a GEDI height metric can be directly used to estimate canopy heights, we assessed the influence of GEDI acquisition and processing parameters on the estimation accuracy by comparing GEDI-derived heights with the reference CHM. We focused on the impact of the algorithm setting group (six possible configurations) and of the laser beam type (two beam types). The influence of other signal parameters such as beam sensitivity and SNR was also studied. Secondly, in light of the results obtained in the first task, we developed and improved the canopy height estimation task by implementing regression models based on either GEDI metrics (L2A product) or GEDI waveforms (L1B product). Several linear and non-linear regression models were evaluated in order to estimate canopy heights from GEDI data. Finally, the generalization of these models was also tested: they were built considering various training strategies depending on the data from the available study sites.

2.3.1. Data Processing

The GEDI L2A product provides information on each acquired footprint and allows for describing them by a unique ID and a set of values corresponding to signal metrics and

acquisition parameters. Regarding reference data, four raster datasets representing canopy heights for each study site are available for the analysis.

Using the geolocation of the received waveforms, with an expected geolocation accuracy of about 10 m, GEDI footprints are overlaid with the CHM rasters. For every footprint that is deemed usable in the GEDI L2A product (see Section 2.2.1), the maximum value of the CHM raster cells contained within the extent of the footprint (a 25 m circle) was extracted. This value is considered the reference canopy height for a given footprint and is referred to as ALS-CHM in this study. The maximum value of the CHM raster cells was chosen because it is theoretically closer to the top-of-canopy signal obtained from GEDI waveforms. Adam et al. [35] justified using the maximum value by the fact that return waveforms start at the highest point of vegetation within the footprints. In a similar study where they compared ALS reference heights with spaceborne LiDAR-derived height estimates, Hilbert and Schmullius [36] obtained better correlations when using the maximum rather than the mean value. Therefore, in this study, we assume that the maximum CHM value within the extent of the footprint is relevant to compare with GEDI canopy heights (referred to as GEDI-CHM).

2.3.2. Direct Estimation

In order to assess the potential of GEDI for the estimation of canopy heights, we firstly decided to use one metric from the L2A product as GEDI-CHM. Among all the available values, we identified two relevant metrics to choose between: *rh_100_an* or *rh_95_an*, with n identifying the algorithm setting group (1 to 6). Relative height metrics give the height at which a certain percentile returned energy is reached relative to the ground. For example, *rh_100_a5* informs on the height at which 100% of the waveform energy is reached, in the signal processing configuration of algorithm setting group number 5. Relative heights of high percentiles are thus good indicators of canopy heights. Regarding the two specific relative heights that were identified, they are strongly correlated as shown in Figure 3.

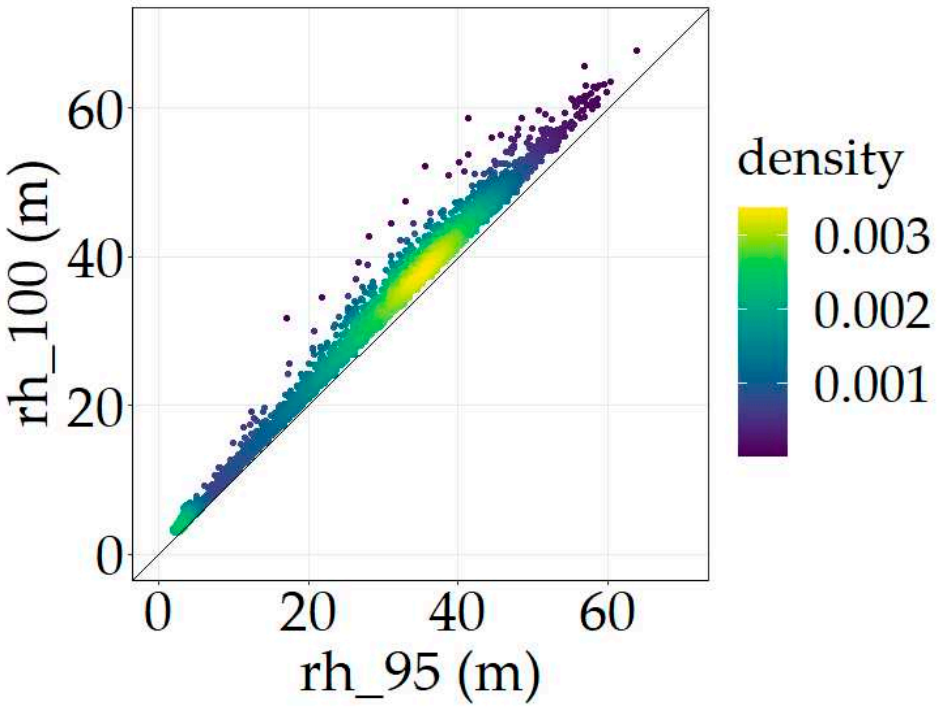


Figure 3. Correlation between GEDI relative heights of percentiles 95 and 100 (R^2 = 0.99).

In the case of direct estimation of canopy heights, we used the *rh_95_an* metric as the value for GEDI-CHM, in accordance with assumptions taken in previous works in the literature [37–40]. The relative height at which 95% of the signal energy is contained is

directly interpreted as the estimated canopy height for a given footprint (GEDI-CHM). Since the study sites are characterized by relatively low slopes, topographic factors were not used to account for slope effects. Regarding the impact of topography, Fayad et al. [41] observed that the estimation of canopy heights from uncorrected GEDI waveforms degraded by a maximum of 1 m for slopes between 11° and 24°. For values below 11°, Fayad et al. [41] noted that the effect of slopes was minimal, with comparable accuracies over the 0–6° and 6–11° ranges. More than 90% of the footprints considered in this study present a mean slope value inferior to 24°, and 60% of the data correspond to slope values in the 0–11° range. In this perspective, given their limited impact on canopy height estimates in the context of low-sloped terrains, topography factors were not taken into account when extracting height metrics from GEDI data.

The accuracy of GEDI estimates is achieved by comparing this value to the reference canopy height derived from the reference raster datasets (ALS-CHM). For the statistical analysis and quantifying the performances of this approach, footprints with ALS-CHM values over 70 m were removed since they most likely do not represent realistic vegetation heights. In order to quantify the accuracy, the root mean square error (RMSE) and the differences between GEDI-CHM and ALS-CHM (GEDI-CHM–ALS-CHM, henceforth referred to as CHM-Differences) are calculated. To account for the fact that CHM-Differences are not normally distributed, two robust metrics used in similar studies [35,42] were chosen for their statistical description: the median and the median absolute deviation (MAD).

For each algorithm setting group (six possible configurations), GEDI-CHM was estimated using the corresponding GEDI height metric (*rh_95_an*, with n being the algorithm setting group) and compared to ALS-CHM, allowing for the evaluation of performances for each configuration. Additionally, the L2A product provides a variable named *selected_algorithm*, which corresponds to the algorithm setting group with the lowest non-noise ground return for a given footprint. With this information, GEDI-CHM was computed using the *rh_95* value associated with the GEDI-selected algorithm setting group, thus creating a new mixed configuration that was also assessed in our investigations.

Regarding the beam type, a GEDI return waveform is acquired from either a coverage or a power beam, depending on the on-board laser that was used to generate the emitted pulse. Coverage and power beams differ in terms of laser power, with the former being produced at an energy level of 5 mJ while the latter are twice as strong at 10 mJ. The intensity of the beam affects its ability to penetrate the vegetation canopy to detect the ground and thus has a major influence on GEDI canopy height metrics. To quantify this influence, the GEDI L2A dataset was split according to beam type, and GEDI-CHM was compared to the corresponding ALS-CHM, allowing for highlighting the differences in performances between each beam type for the estimation of canopy heights.

GEDI's canopy height estimation accuracy is directly linked to the laser ability to penetrate through the vegetation and to detect the ground. Identifying the ground peak in the waveform signal is of paramount importance to estimate canopy heights from a relative height metric such as *rh_95*. An incorrect ground peak detection in a waveform signal can be explained either by the fact that the laser could never reach the ground or because the ground mode is too hard to isolate from the noise that characterizes every numeric signal. In this prospective, sensitivity and SNR are also two fundamental signal parameters that allow for understanding and characterizing these scenarios. Firstly, beam sensitivity is defined as the maximum canopy cover that can be fully penetrated all the way to the ground by the laser. A high sensitivity allows for the signal to penetrate denser canopies and thus to reach the ground. Secondly, SNR is a commonly used measure that compares the level of a desired signal to the level of background noise [28]. A high SNR implies a better extraction of the useful parts of the signal to perform the metrics computation. The influence of these parameters on GEDI-CHM was assessed to understand how they affect canopy heights estimates.

2.3.3. Regression Models

GEDI relative height metrics are good indicators of forest heights. The simplest method to estimate the canopy height from a GEDI waveform is to use a single metric derived from the waveform and to consider that it directly represents the canopy height. In the case of this study, we first used the relative height at which 95% of the waveform energy is reached (*rh_95_an*). However, over densely forested areas with high AGB levels, it is harder for the laser signal to penetrate the vegetation and to detect the ground, making the identification of the ground peak in the received waveform difficult and the estimation of forest height inaccurate [20]. In this prospective, statistical approaches have been developed and used in several studies to predict canopy heights from GEDI data [4,23,43]. These methods proposed regression models based on waveform metrics. Rather than using a single metric, GEDI-CHM is therefore the result of a model built on several pre-selected GEDI L2A waveform metrics (see Table 2). In addition, in the same spirit, deep learning approaches allowed for estimating canopy heights directly from the geolocated waveforms of the GEDI L1B data product [21,44]. In that case, GEDI-CHM is the result of a model trained and validated on the GEDI L1B received waveforms.

Table 2. GEDI L2A variables to be used as predictors in the canopy height regression models.

Variable	Unit	Description
Waveform extent	m	Distance between the highest and lowest detectable returns in the waveform
Trailing edge extent	m	Distance between the lowest detectable return and the ground return
Leading edge extent	m	Distance between the highest detectable return and the first mode
Relative height at n% of cumulative energy	m	Height at which n% of the waveform energy is reachedn = 10, 15, . . . , 95, 100
Beam type		Type of beam associated with the waveform (power or coverage)
Sensitivity	%	Maximum canopy cover that can be penetrated considering the SNR of the waveform

In this study, three types of models were implemented and evaluated using ALS-CHM reference data: a stepwise multilinear regression (SRH), a random forest (RF) regressor and a convolutional neural network (CNN) encoder. To assess how the models perform and how they can be generalized to an independent dataset, a ten-fold cross validation was used, and the RMSE as well as the CHM-Differences were computed. This process was either applied on the whole available dataset (i.e., the four sites merged into a single database) or on sub-samples corresponding to the data of a given country (Nouragues and Paracou altogether on one hand, Lope and Rabi on the other). The purpose is to highlight if the availability of reference data is of interest to create general estimators that can be applied at a larger scale to other areas with similar heights and AGB levels. In addition, for the sake of completeness and to understand how these approaches could be generalized to other study sites, the models were also built on the data available for each country and were tested directly on the other one.

Firstly, a SRH was implemented to estimate canopy heights by automatically selecting a set of predictive explanatory variables among all possible predictors presented in Table 2. The choice of adding or removing a variable from the SRH linear model is based on the variation of the mean squared error (MSE). In this approach, the target variable is estimated through a linear relationship between the most relevant predictors. Canopy heights were also estimated through non-linear non-parametric regressions using a random forest (RF) regressor. The RF method is a machine learning algorithm that leverages ensemble of trees via a bagging strategy to estimate the target variable. Compared to SRH, RF has the advantage of being able to model non-linear relationships between the predictors. In this analysis, we used the explanatory variables described in Table 2, and the number of trees was set to

500. The importance of the predictors used as input for the regression can be quantified in order to understand the most contributing factors in the estimation of the target variable. Finally, CNN encoders such as those described in [21] were developed to predict canopy heights from GEDI waveforms. CNN architectures implement the convolutional operation and allow for dealing with the temporal autocorrelation that characterizes numeric signals such as the L1B waveforms. Compared to the previous models, these approaches rely solely and directly on the received GEDI waveform signals. In order to extract an extensive amount of signal features and to capture localized information in the input waveforms, we set a small kernel size equal to 3 for the convolution operations.

3. Results

3.1. Using rh_95

3.1.1. GEDI-CHM Accuracy

The results presented in Figure 4 are scatter plots of GEDI-CHM and ALS-CHM values for each algorithm setting group. They exhibit a quite weak linear correlation ranging from 0 to 0.30 (R^2 coefficient). The most notable feature that characterizes the distributions is a high number of footprints where GEDI significantly underestimates canopy height. GEDI's ability to correctly estimate tree heights is linked to its ability to detect the ground below and is therefore directly correlated to the laser penetration through the vegetation.

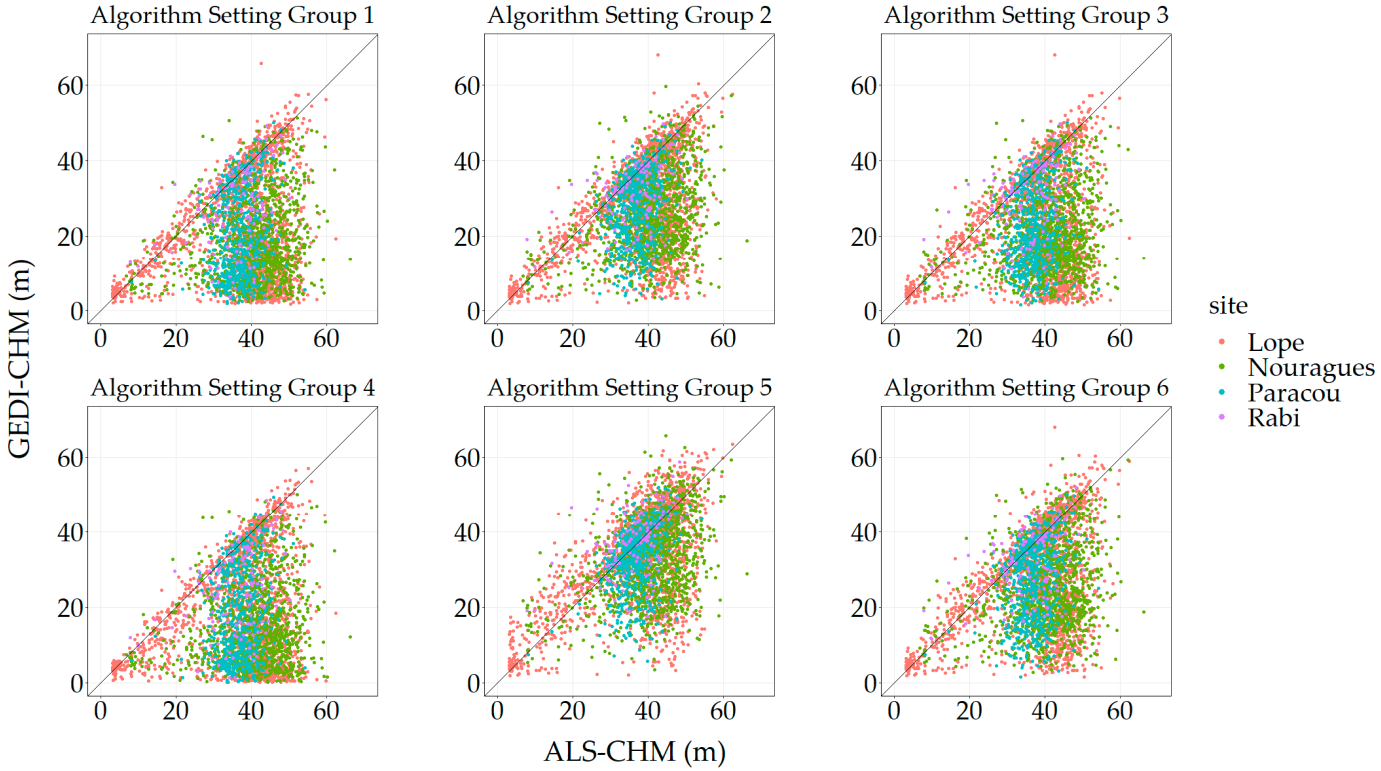


Figure 4. GEDI-CHM estimated from *rh_95* as a function of ALS-CHM for each algorithm setting group.

The accuracy metrics displayed in Table 3 confirm the results implied by the scatter plots. Algorithm setting group number 5 performs the best to estimate CHM with a median CHM-Differences of −1.5 m and a RMSE of 11.6 m. Other algorithms show a relatively low accuracy, with setting group numbers 1, 3 and 4 being less accurate. Algorithm setting group numbers 2 and 6 exhibit similar performances since they have similar signal processing settings. Finally, the GEDI-selected algorithm provided in the L2A product outperforms all the algorithms except for the number 5.

Table 3. Accuracy of GEDI-CHM estimates using *rh_95* height metric derived with different algorithm setting groups.

Algorithm Setting Group	Median CHM-Differences (m)	MAD CHM-Differences (m)	RMSE (m)
1	−16.6	19.8	21.9
2	−7.3	12.2	15.7
3	−14.6	18.8	20.4
4	−20.1	20.3	24.6
5	**−1.5**	**9.0**	**11.6**
6	−7.7	13.7	16.7
selected	**−4.7**	**10.3**	**14.4**

3.1.2. Influence of GEDI Beam Type

Focusing on the influence of beam type, the same trends can be observed, and the same conclusions can be drawn regardless of the algorithm setting group that was used. Therefore, in order to assess and understand the influence of beam type on GEDI's ability to estimate canopy heights, we focus on algorithm setting group number 5 since it performs better than all its competitors.

Boxplots in Figure 5a describe the distributions of CHM-Differences depending on beam type. A clear shift in accuracy can be observed with respect to beam type, and it is confirmed by the values reported in Table 4. The median of CHM-Differences from coverage beams is greater in absolute value than the median of CHM-Differences from power beams, the former being at −9.1 m while the latter is at 1.2 m. In addition, CHM-Differences are in general higher in absolute value for footprints acquired through coverage beams (Figure 5a). In a notable way, GEDI coverage beams tend to underestimate CHM, whereas no such observation can be made for power beams.

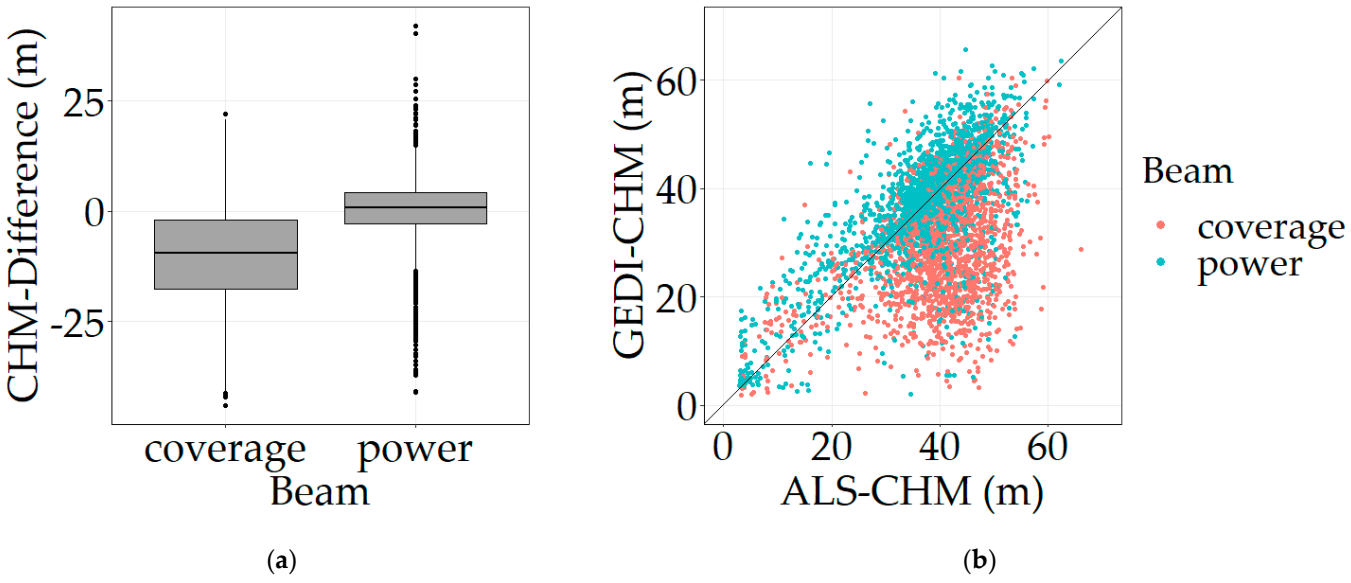


Figure 5. Boxplots of the CHM-Differences (**a**) and GEDI-CHM as a function of ALS-CHM (**b**) depending on beam type (algorithm setting group number 5).

Figure 5b provides a visual representation to appreciate the difference in CHM estimation accuracy between coverage and power beams. The linear correlation is stronger for footprints acquired through power beams, with an R^2 value of 0.70 instead of 0.25 for coverage beams.

Table 4. Accuracy of GEDI-CHM estimates using *rh_95* height metric depending on beam type (algorithm setting group number 5).

Beam Type	Median of CHM-Differences (m)	MAD of CHM-Differences (m)	RMSE (m)
coverage	−9.1	11.7	14.8
power	1.2	6.7	8.1

3.1.3. Influence of Sensitivity and SNR

Figure 6 exhibits the relation between CHM-Differences and sensitivity on one hand, and between CHM-Differences and SNR on the other. As highlighted by the trendlines, GEDI footprints with a higher sensitivity and SNR show a better accuracy for the estimation of canopy heights. A low sensitivity implies a bad ground height estimation because the signal is not able to penetrate dense vegetation all the way through, resulting in an overestimation of ground heights and therefore an underestimation of canopy heights. Similarly, a low SNR makes the ground mode harder to identify and extract, either because it is too weak or too noisy, and this results in an underestimation of canopy heights.

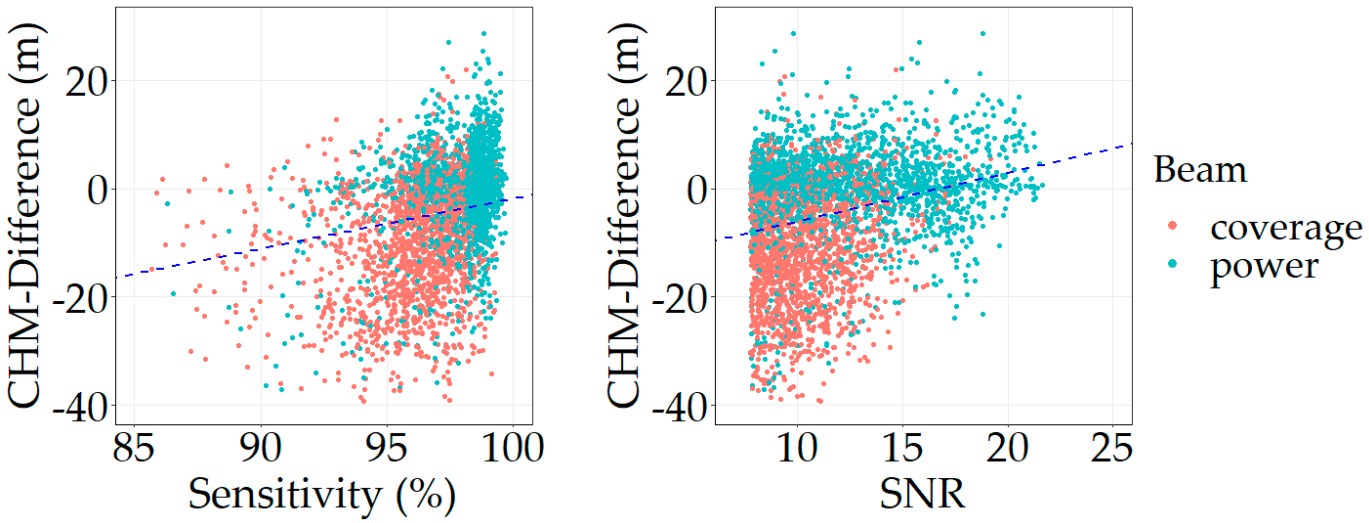


Figure 6. CHM-Differences as a function of sensitivity (**a**) and SNR (**b**) for algorithm setting group number 5. The blue dashed lines represent the trendlines (linear).

Additionally, when considering beam type in this analysis, coverage beams tend to be associated with lower sensitivities and SNR (Figure 6), whereas power beams do not show a tendency to underestimate canopy heights (Figures 5 and 6). The laser power has a direct influence on signal parameters such as sensitivity and SNR, which in turn are decisive for accurate canopy height estimates.

3.2. Using Regression Models

3.2.1. GEDI-CHM Accuracy with Models Based on Waveform Metrics

In this section, we evaluate the performances of SRH and RF regressors built on the metrics defined in Table 2. Figure 7 shows the RMSE and MAD of GEDI-CHM for each estimator based on the GEDI L2A product. The results highlight the fact that using regression models built on waveform metrics allows for significantly improving the results of the estimation task compared to the direct method. Algorithm setting group number 5 exhibits the best performances for all three estimators, and RF appears to be the best approach regardless of which algorithm setting group is used. For example, when considering algorithm setting group number 5, the RMSE improved from 11.6 m with a direct estimation

of 8.0 and 6.7 m with the SRH and RF models, respectively. Moreover, these models show that they are less dependent on the algorithm used to compute the waveform metrics than the direct method, with RF being the most stable approach in this regard.

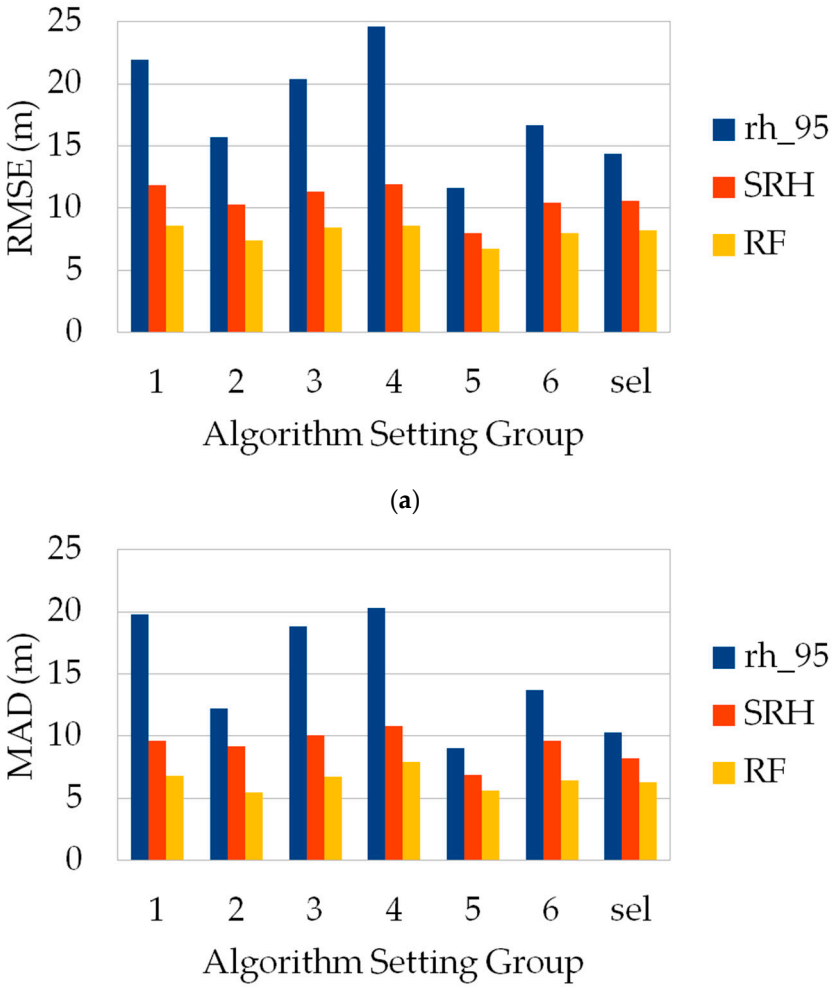


Figure 7. RMSE (**a**) and MAD (**b**) for GEDI-CHM estimators (10-fold cross validation, whole dataset) depending on algorithm setting group: direct estimation, SRH and RF. "sel" corresponds to GEDI-selected algorithm.

When the analysis is carried out according to beam type, a clear shift in performances is observed between the two laser types. Since algorithm setting group number 5 displays the best performances in GEDI-CHM estimations, further performance analyses in this paper are presented for this specific configuration. Figure 8 shows the RMSE and MAD of GEDI-CHM estimates for the competing estimators depending on the footprints' beam types. The main result that can be drawn is that power beams generally perform better than coverage for the estimation task regardless of the method used. Moreover, similarly to what was previously observed concerning the model stability in relation to the algorithm setting group, SRH and RF regressors exhibit relatively close performances between power and coverage beams, with RF being the most stable approach once again: the RMSE reaches values of 5.9 and 7.7 m for power and coverage beams, respectively (Figure 8a). Finally, power beams show a relative stability for GEDI-CHM estimations since the results obtained are of similar magnitude regardless of the method used, with RMSE values ranging from 5.9 to 8.1 m (Figure 8a).

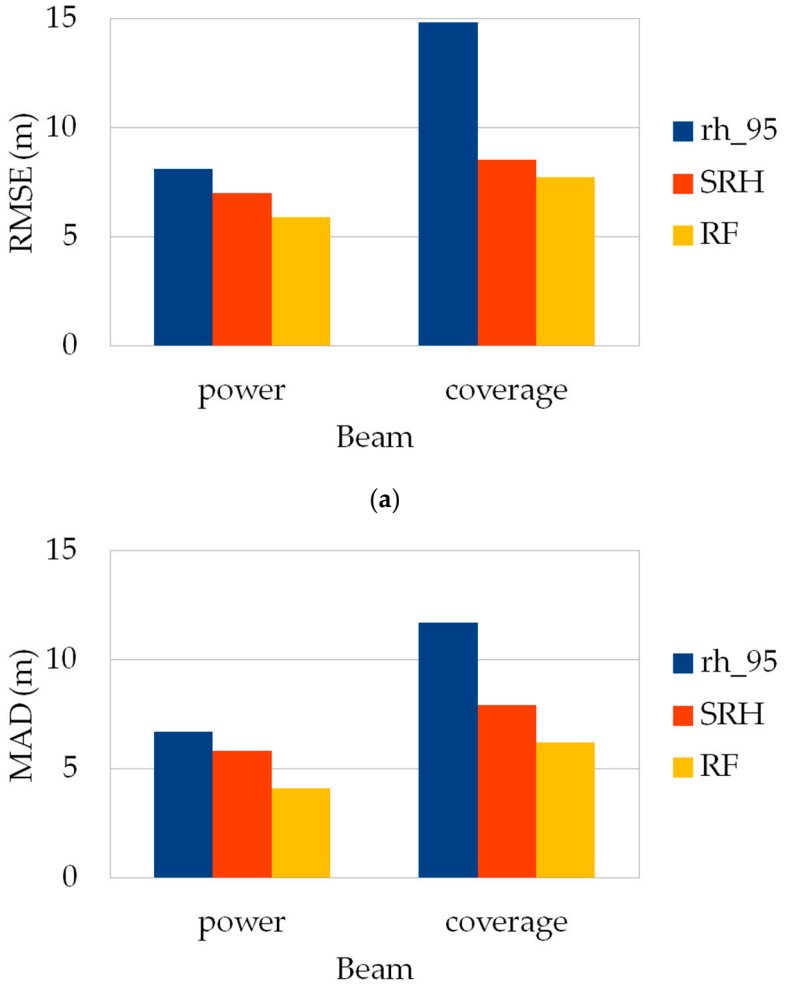


Figure 8. RMSE (**a**) and MAD (**b**) for GEDI-CHM estimators (10-fold cross validation, whole dataset) depending on beam type (algorithm setting group number 5): direct estimation, SRH and RF.

3.2.2. Variable Importance for Models Based on Waveform Metrics

In this section, we present the variable importance analysis of SRH and RF regressors. Firstly, the SRH model automatically selects the most relevant predictors among all the available inputs and produces a linear relation, allowing for linking CHM with the most explanatory variables. Before the implementation of the SRH model, sensitivity was converted to a binary parameter. Its value is defined depending on a certain threshold: for a given GEDI footprint, if the sensitivity is superior to 98%, then the value is set to 1 (reciprocally to 0 if the sensitivity is inferior to 98%). Similarly, beam type was also converted to a Boolean value: 0 for coverage beam and 1 for power. These binary parameters are used as input variables in the SRH model implementation. When considering all the available data (i.e., 3864 GEDI footprints), the relation built by the SRH model is the following:

$$\begin{gathered}\text{GEDI-CHM (m)} = \\ 22.5 - 0.6 \times \text{wext} - 0.3 \times \text{lead} + 1.3 \times \text{rh_95} + 1.2 \times \text{sensitivity} - 6.9 \times \text{beamtype}\end{gathered} \quad (1)$$

Waveform extent (m), leading edge extent (m), *rh_95* (m), sensitivity and beam type are automatically selected as the most contributing variables in CHM estimation.

Secondly, regarding RF, the analysis of the importance informs about which variables have the most predictive power (Figure 9). Beam type appears as the most important

variable for CHM estimation. In order to be able to compare all regressors on a fair basis, the categorical beam sensitivity values are also used as RF input variables, and sensitivity emerges as a key predictor. Waveform extent and leading edge extent still bring a significant contribution in a similar way to what has been observed for SRH. Concerning relative heights, *rh_10* and *rh_100* are the most contributing input variables when estimating CHM through a RF model.

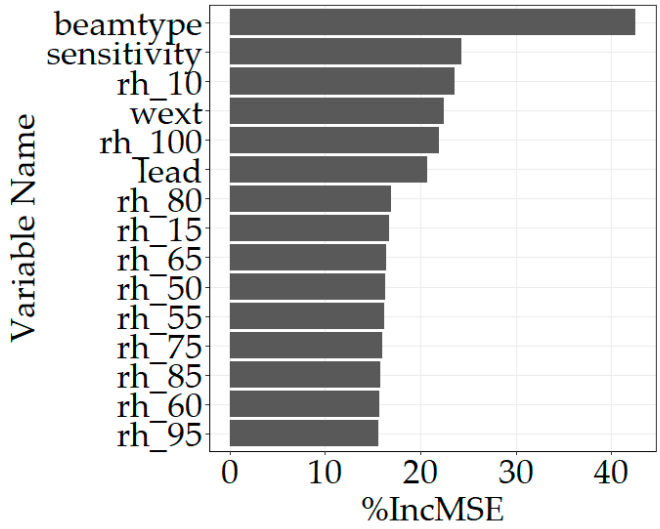


Figure 9. Importance of RF variables (fifteen most important) using mean decrease accuracy (%IncMSE).

3.2.3. Model Transferability

To study the viability of model transferability across study sites, several training options were considered given the variety of data in the dataset used in this study. In this section, we focus on the use of RMSE to evaluate performances. Regarding the strategies to train and validate models, a total of five approaches was adopted as reported in Table 5. These approaches rely on different partitions of the whole dataset for the training and validation of models.

Table 5. RMSE for GEDI-CHM SRH and RF estimators depending on training/validation strategies (algorithm setting group number 5).

Strategy	Training	Validation	SRH RMSE (m)	RF RMSE (m)
"Full"	Whole dataset (3864 footprints)	10-fold cross validation	8.0	6.7
"Guiana"	French Guiana dataset (1733 footprints)	10-fold cross validation	7.0	6.9
"Gabon"	Gabon dataset (2131 footprints)	10-fold cross validation	8.2	6.5
"Gabon > Guiana"	Gabon dataset (2131 footprints)	French Guiana dataset	8.4	7.3
"Guiana > Gabon"	French Guiana dataset (1733 footprints)	Gabon dataset	11.7	11.3

Figure 10 gives a graphical representation of the results reported in Table 5. When considering strategies relying on a 10-fold cross validation ("Full", "Guiana" and "Gabon"), RF shows significant stability in terms of performances, with identical values of RMSE for each configuration (6.7, 6.9 and 6.5 m, respectively). SRH is more sensitive to the dataset used to build the model but also remains quite stable, with RMSE values ranging from 7.0 to 8.2 m. Regarding strategies that are based on country training (training on one country, validation on the other), they exhibit degraded performances in terms of RMSE,

with "Guiana > Gabon" being by far the least interesting approach for both SRH and RF estimators.

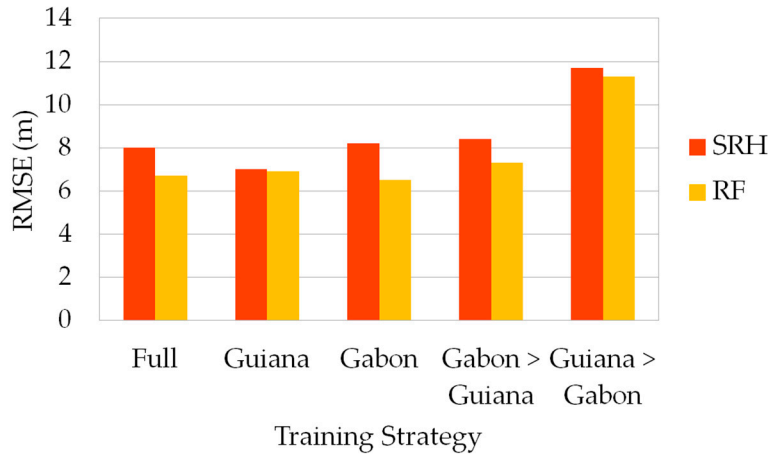
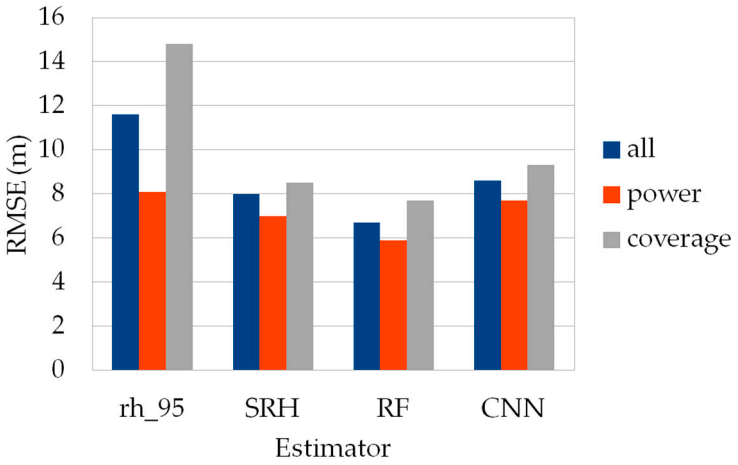


Figure 10. RMSE for GEDI-CHM SRH and RF estimators depending on training/validation strategies (algorithm setting group number 5).

3.2.4. GEDI-CHM Accuracy with Models Based on Waveforms

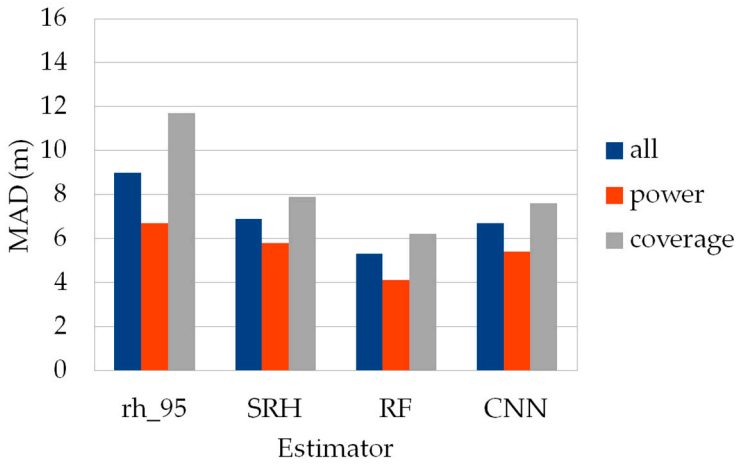
CNN encoders are directly built on the waveforms available in the GEDI L1B product and therefore are not dependent on algorithm setting group since the signal is directly processed as is, and no metrics extracted from it are used. In order to be able to compare the results of this method with its competitors, we considered the best performing algorithm setting group for the previous approaches, which is setting group number 5. Similarly to what was performed before, the CNN encoders were trained and validated following different strategies. Figure 11 shows the RMSE and MAD depending on beam type for all the methods considered in this study, trained with the "Full" strategy.

Regarding beam type, CNNs behave the same way as their competitors, with power beams also giving the best results in terms of RMSE. However, RF remains the best performing method to estimate CHM in all training scenarios, while CNNs are the second best. For example, when considering the common scenario of a "Full" training strategy and all beam types used, CNNs reach a RMSE value of 8.6 m compared to the best value of 6.7 m for RF.



(**a**)

Figure 11. *Cont.*



(**b**)

Figure 11. RMSE (**a**) and MAD (**b**) for all GEDI-CHM estimators depending on beam type: all beams, power only and coverage only.

4. Discussion

The different approaches tested in this paper to estimate canopy heights from GEDI data showed that prediction models based on waveform metrics allowed for obtaining good accuracies of heights estimates, with a best RMSE and MAD value of 5.9 and 4.1 m, respectively.

The simplest way to estimate canopy heights from GEDI data is to use a specific L2A level metric as a direct proxy of CHM. We chose to use *rh_95* for this purpose, corresponding to the 95th percentile of energy return height relative to the ground. Upper relative height metrics (i.e., *rh_90*, *rh_95* and *rh_100*) are related to the top of forest canopies. Potapov et al. [37] observed in their study that, in a tropical biome, *rh_95* showed the highest correlation with ALS-derived reference heights, while *rh_90* underestimated canopy height, and *rh_100* overestimated it compared to *rh_95*. Neuenschwander et al. [39] made a similar observation, although in a different context of sparse boreal forests in Alaska (USA).

GEDI metrics values are based on the L2A processing algorithm, which uses six different sets of parameters to perform the waveform processing. The analysis of the GEDI-CHM accuracy with *rh_95* derived from all available algorithm setting groups showed that algorithm setting group number 5 delivers a significantly better accuracy than other groups. In the tropical context of French Guiana and Gabon, forests present high AGB levels and dense canopies. The ability of GEDI to estimate canopy heights is directly dependent on its ability to detect the ground. Due to the density of the vegetation in tropical areas, the ground peak in GEDI waveforms is usually weaker in terms of intensity and is therefore harder to extract and separate from the background noise. Algorithm setting group number 5 is characterized by a lower waveform signal end threshold compared to the other setting groups. A low threshold allows for better discrimination of weak ground returns from the noise in the signal, while higher thresholds will lead in general to an error of detection. On the contrary, in less dense forest ecosystems, a low threshold can result in the interpretation of noise below the actual ground as the ground peak, leading to an overestimation of canopy heights [35,43]. GEDI waveforms acquired over tropical forests usually present weaker intensity ground peaks compared to other less dense forest environments, allowing for the conclusion that a lower signal end threshold such as the one characterizing algorithm setting group number 5 seems more adapted in this specific context. The results obtained with algorithm setting groups number 1, 3 and 4 are in the same direction as this observation: they are characterized by the highest signal end threshold among all possible configurations and are proven to be the least accurate when estimating canopy heights from *rh_95*.

The assessment of GEDI canopy height estimates is based on reference heights obtained from ground truth data. Each GEDI footprint is described by a unique reference height that is compared to the canopy height derived from the selected GEDI metric. In this prospective, the maximum CHM value in the footprint's extent appears to be better correlated to GEDI height metrics since the *toploc* of the waveform (i.e., the location of the highest detected return in the waveform) is associated with the first laser-environment interaction that occurs with the highest object within the footprint [35,36]. For the sake of completeness and in order to consider another proxy for ALS-CHM, the accuracy assessment of GEDI canopy height estimates was also performed using the mean CHM value within the footprint as the reference height value (Table 6).

Table 6. Accuracy of GEDI-CHM estimates using the mean height value for ALS-CHM.

Algorithm Setting Group	Median CHM-Differences (m)	MAD CHM-Differences (m)	RMSE (m)
1	−4.7	17.2	16
2	2.9	10.9	12.2
3	−2.3	16	15
4	−8.6	18.2	17.5
5	**6.7**	**9.5**	**12.6**
6	2.8	12.4	13.2
selected	**3.8**	**10.1**	**12.2**

When using the mean height value as ALS-CHM, the most notable observation that can be drawn is that the tendency to underestimate canopy heights is greatly lessened. For example, the least performing algorithm setting groups (groups number 1, 3 and 4, as stated previously) still present a negative median CHM-Differences but significantly lower in terms of absolute value. On the contrary, the other better performing algorithm setting groups tend to slightly overestimate canopy heights, especially algorithm setting group number 5, which is characterized by a lower waveform signal end threshold. The choice of either the mean or the maximum value to extract the ground truth greatly impacts the reference heights that are used for the accuracy assessment of GEDI canopy height estimates. Over the 3864 footprints that were retained in this study, the mean difference between mean ALS-CHM and maximum ALS-CHM amounts to −10.1 m (−9.4 m for the median), with a standard deviation of 5.3 m. The observed attenuation of the height underestimation for all algorithm setting groups could then be linked to the fact that reference heights given by mean ALS-CHM are significantly lower than the actual heights as seen by GEDI waveforms. Therefore, in accordance with assumptions taken in related works [35,36], the maximum ALS-CHM is deemed more relevant to compare with GEDI height metrics.

The signal capability to penetrate through the forest and reach the ground is closely linked to the laser's physical properties. GEDI's power beams are twice as powerful as coverage ones in terms of laser energy, which directly affects the penetration of dense canopies. Fayad et al. [20] observed that coverage lasers exhibited a significantly lower level of performance for tree height estimation in comparison to other full power configurations such as GEDI and NASA's Land Vegetation and Ice Sensor (LVIS). The analysis performed in our study delivered similar results, with power beams showing a clear advantage for CHM estimation compared to coverage beams. The main observation that was drawn is a tendency for the latter to underestimate tree heights, whereas the former exhibited a quite strong linear correlation with reference canopy heights. Pulses emitted with coverage lasers generally have more difficulty reaching the ground because of the density of the vegetation, and consequently, the mode extracted as the ground peak in the waveforms is located at a higher height than the actual ground elevation, thus resulting in an overestimation of the ground height and therefore an underestimation of canopy heights. On the contrary, power beams seem to be able to penetrate the vegetation all the way vertically and produce

received waveforms with a recognizable lowest mode corresponding to the actual ground. When using GEDI data for canopy height retrieval, Liu et al. [45] noted that beam strength had a strong impact on the obtained results and recommended the use of only strong/power beam data for better CHM estimates. We reached the same conclusion in light of our study, with an improvement of about 80% and 75% in terms of RMSE and MAD, respectively, when comparing coverage and power beams for CHM estimation through *rh_95*.

The waveform interpretation and metrics computation are also very dependent on the signal quality and shape. The physical properties of the received signal that is processed by GEDI have a major influence on the obtained L2A level metrics, and thus on the CHM estimation through *rh_95*. The results obtained in this study highlight the fact that beam sensitivity is a parameter of major importance for CHM estimation. The higher the sensitivity, the better the signal that can penetrate the vegetation and detect the ground. Using *rh_95* as a direct proxy of CHM requires a correct detection of both the ground peak and the canopy top in the waveform in order to be able to interpret this value as the actual canopy height. Over densely vegetated areas such as the ones considered in this study, the reflection off the ground that is recorded by GEDI sensors can be very weak, if not totally absent. Footprints characterized by a high sensitivity generally allow for a better detection of the ground return since these shots can penetrate denser canopy covers. In this prospective, Fayad et al. [20] concluded that footprints with a sensitivity higher than 98% showed a deeper penetration of the vegetation, as the *rh_100* value increased by 5 m on average for both power and coverage beams. Although they considered another height metric than the one we used, the strong correlation between *rh_95* and *rh_100* allows for drawing the same conclusion regardless of the metric used. Our results also highlight a connection between beam sensitivity and beam type (Figure 6). When exploring this connection in more detail, it appears that power beams are characterized by increased values of sensitivity (Figure 12), which is consistent with the fact that stronger lasers are able to penetrate through denser canopy covers. In addition, the threshold of 98% that was used by Fayad et al. [20] to remove shots with insufficient penetration roughly discriminates between power and coverage beams in our case.

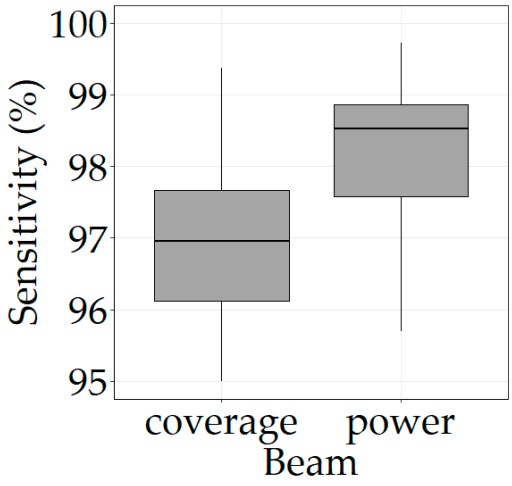


Figure 12. Relation between beam type and beam sensitivity (algorithm setting group number 5). Outliers removed for clarity.

Regarding SNR, the higher it is, the easier it is to correctly interpret the waveform and extract the ground peak from the background noise. When considering the influence of SNR according to beam type, power and coverage beams behave in different ways (Figure 6): while the former show a certain stability for the estimation of CHM, with no significant differences between low and high values of SNR, the latter exhibit a tendency to underestimate canopy heights in low ranges of SNR. The stability of power beams

advocates for the fact that, even in the context of noisy waveforms, beam sensitivity is the most impactful parameter for CHM estimation, as it is directly linked to ground detection. A weak ground mode in a given waveform is not necessarily a synonym of bad canopy height estimate if its location really corresponds to the actual ground elevation. In addition, similarly to the observation made before for sensitivity, power beams are characterized by increased values of SNR, which is also consistent with the fact that stronger lasers tend to be less sensitive to background noise.

Overall, the accuracies obtained when using *rh_95* as a direct proxy of CHM remain insufficient to consider this method as satisfying. The use of statistical approaches based on GEDI L2A metrics induces significant improvements on CHM estimates [4,23,43]. Rather than considering a single metric, multiple predictors are used as input to build and train empirical models. The approaches evaluated in this study allow for taking advantage of the richness of the information contained in the waveforms and for leveraging the interplay between all the predictors to estimate the target variable. The predictors that were selected to perform this task correspond to height metrics describing the canopy tops, the ground elevation and the vertical structure of the forest. These variables are correlated to tree heights and are often used in canopy height estimation [4,23,43]. Signal parameters such as beam type and beam sensitivity were also added as input variables since they proved to strongly impact the direct CHM estimation through *rh_95*. The implementation of a SRH model shows that waveform metrics can be used to obtain significantly better accuracies of CHM. When using all the available data to build the model, we reach notable improvements of 45% and 30% in RMSE and MAD, respectively, compared to the use of *rh_95* as a direct proxy of CHM. The model automatically selects the most relevant predictors among all the available inputs and produces a linear relation allowing for the identification of the most explanatory variables. Waveform extent, leading edge extent, *rh_95*, sensitivity and beam type emerge as the most important GEDI variables for the estimation of CHM. Waveform extent is one of the most used waveform metrics to estimate CHM through linear models. Lefsky et al. [4] firstly developed a statistical model that was based only on the waveform extent to estimate the maximum canopy height. Payn et al. [46] then incorporated the waveform leading edge extent and obtained slightly improved results on canopy height estimation. The leading edge extent characterizes the detection of canopy tops and, in our case, is selected as a major factor of CHM variability. As stated many times before, *rh_95* is one of the most correlated GEDI metrics with ALS reference heights, and it is therefore natural to prove as a significant contributor to CHM estimates. Finally, consistent with the analysis performed in this paper, sensitivity and beam type are two fundamental signal parameters that appear in Equation (1) because they are linked to GEDI's ability to detect the ground and, consequently, to derive relevant height metrics from the received waveforms.

RF is commonly used for the prediction of forest attributes when reference ground truth data are available. In the context of this study, RF shows a strong capability to leverage the available inputs to produce good CHM estimates. This method proves to be both the most accurate and the most stable regarding all the parameters of influence that were considered. The accuracies of CHM estimates remain in the same order of magnitude no matter which algorithm setting group is used (RMSE ranging from 6.7 to 8.6 m). Moreover, when considering the impact of beam type, RF is also capable of exploiting the information given by waveform metrics regardless of the configuration of the acquisition (RMSE of 5.9 and 7.7 m for power and coverage beams, respectively). The analysis of the importance informs about which variables have the most predictive power. Figure 9 shows that the variables that were previously selected by the SRH model still appear as impactful predictors for the CHM estimation task using RF. In particular, as highlighted and discussed before, beam type and sensitivity are key inputs for the models, as they are directly linked to laser penetration and ground detection. The importance of *rh_10* can be explained by the fact low percentiles of relative heights (*rh_10* and *rh_15* especially) provide information about the deep layers of the forest and thus are related to the laser penetration. Reciprocally, *rh_95* still appears as an explanatory predictor for canopy heights, as well as *rh_100*.

The CNN encoders implemented in this study did not allow for obtaining significant gains in performances, as they produced CHM estimates with accuracies similar to the ones of SRH regressors. Deep learning approaches and especially neural networks rely on the availability of big amounts of data for model training. The limited sample size of the dataset used in this analysis (less than 4000 usable GEDI footprints) may not be adequate to exploit the full capacity of these methods. Nevertheless, CNNs have the huge advantage of directly taking the L1B product as input (i.e., the waveforms), thus avoiding the extraction of descriptive metrics from the signals. As highlighted many times before, some metrics are harder to extract depending on the signal parameters, especially the ones linked to the ground detection, and therefore, it can be interesting to skip this step that may alter the waveform information and give uncertain results.

Finally, the testing of several training options advocates for the viability of model transferability across study sites that are characterized by comparable tree heights and AGB levels. Each training configuration gave similar results in terms of RMSE for both SRH and RF, except for the "Guiana > Gabon" strategy. The French Guiana ground truth data cover a smaller range of tree heights (between 21 and 58 m) compared to the Gabon dataset (between 4 and 62 m), which poses a problem for a method such as RF that is unable to extrapolate and is limited within the range of the training sample. On the contrary, the "Gabon > Guiana" strategy performed similarly as the "Full" one given the fact that the training sample covers a wide range of tree heights. All things considered, a sufficient and diverse training set is always necessary for statistical approaches and empirical models built on reference data. Such models could then be transferred and applied on other study sites that share similar structural characteristics.

5. Conclusions

In this study, several approaches to estimate CHM from GEDI data were tested while assessing the impact of GEDI acquisition and processing parameters on the accuracy of canopy height estimates. The tested methods can be classified into two main categories depending on the approaches that were considered. On one hand, we used a GEDI height metric (*rh_95*) as a direct proxy of CHM. This method is the simplest since it relies on a single value that is directly available in the GEDI L2A product, but in return, it is also the least accurate and the most dependent on GEDI signal properties. On the other hand, statistical approaches (built on the L2A product) and deep learning frameworks (based on the L1B product) were implemented and allowed for reaching improved accuracies for CHM estimates.

All things considered, GEDI acquisition and processing parameters have a strong impact on the data derived from waveforms and, consequently, on the CHM estimates produced from these data. Firstly, the algorithm setting group defines the parameters for the smoothing of received waveforms and thus directly influences the values of the metrics extracted from the signals. Depending on the forest context, a setting group is more or less adapted for the extraction of relevant metrics. In this analysis, we observed that algorithm setting group number 5 was the most suited to correctly detect the ground peak in waveforms acquired over tropical forests. This result is quite different from comparable studies that analyzed the impact of algorithm setting groups over temperate forests, for example. Secondly, the beam type, which is linked to GEDI laser power, appears as the most impactful parameter when assessing the quality of GEDI data. Power and coverage beams differ significantly in terms of laser energy, and over tropical forests and high AGB contexts, power beams are more capable of penetrating the vegetation all the way to the ground, thus providing received waveforms accurately describing the vertical structure of the forest from the top-of-canopy to the actual ground. Thirdly, in the same spirit, beam sensitivity is also linked to the laser penetration and is a key input for canopy height estimation, especially over the dense canopies that characterize tropical biomes.

Overall, the use of *rh_95* as a direct proxy of CHM exhibited relatively low accuracies. When no reference data are available (i.e., reference canopy heights), *rh_95* can still be used

in certain conditions to derive canopy heights. We recommend the use of only power beam data, as this configuration allowed for reaching better results compared to the use of all available footprints. In contrast, when reference data are available, the implementation of regression models induces significant improvements in CHM estimation accuracies. These approaches also show certain stability in relation to the signal acquisition and processing parameters that were assessed in this study. The key factor when building empirical models is the availability of sufficient amounts of data for model training. Depending on the data availability, the operational application and the desired accuracies, GEDI information can be used in diverse manners to estimate canopy heights.

Author Contributions: Conceptualization, K.L., N.B., G.l.M. and I.F.; Data curation, K.L., N.B., G.l.M. and I.F.; Formal analysis, K.L., N.B., G.l.M. and I.F.; Methodology, K.L., N.B., G.l.M. and I.F.; Software, K.L. and I.F.; Validation, K.L., N.B., G.l.M. and I.F.; Visualization, K.L.; Writing—original draft, K.L., N.B., G.l.M. and I.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the French Space Study Center (CNES, TOSCA 2022 project) and the National Research Institute for Agriculture, Food and the Environment (INRAE).

Data Availability Statement: GEDI data were obtained from the LP DAAC (https://e4ftl01.cr.usgs.gov/GEDI/, accessed on 29 August 2022).

Acknowledgments: The authors would like to thank the GEDI team and NASA's LP DAAC for providing the GEDI data.

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

References

1. 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.; et al. A Large and Persistent Carbon Sink in the World's Forests. *Science* **2011**, *333*, 988–993. [CrossRef] [PubMed]
2. Asner, G.P.; Mascaro, J. Mapping Tropical Forest Carbon: Calibrating Plot Estimates to a Simple LiDAR Metric. *Remote Sens. Environ.* **2014**, *140*, 614–624. [CrossRef]
3. Chave, J.; Andalo, C.; Brown, S.; Cairns, M.A.; Chambers, J.Q.; Eamus, D.; Fölster, H.; Fromard, F.; Higuchi, N.; Kira, T.; et al. Tree Allometry and Improved Estimation of Carbon Stocks and Balance in Tropical Forests. *Oecologia* **2005**, *145*, 87–99. [CrossRef] [PubMed]
4. Lefsky, M.A.; Harding, D.J.; Keller, M.; Cohen, W.B.; Carabajal, C.C.; Del Bom Espirito-Santo, F.; Hunter, M.O.; de Oliveira, R. Estimates of Forest Canopy Height and Aboveground Biomass Using ICESat: Icesat Estimates of Canopy Height. *Geophys. Res. Lett.* **2005**, *32*, L22S02. [CrossRef]
5. Feldpausch, T.R.; Lloyd, J.; Lewis, S.L.; Brienen, R.J.W.; Gloor, M.; Monteagudo Mendoza, A.; Lopez-Gonzalez, G.; Banin, L.; Abu Salim, K.; Affum-Baffoe, K.; et al. Tree Height Integrated into Pantropical Forest Biomass Estimates. *Biogeosciences* **2012**, *9*, 3381–3403. [CrossRef]
6. Lima, A.J.N.; Suwa, R.; de Mello Ribeiro, G.H.P.; Kajimoto, T.; dos Santos, J.; da Silva, R.P.; de Souza, C.A.S.; de Barros, P.C.; Noguchi, H.; Ishizuka, M.; et al. Allometric Models for Estimating Above- and below-Ground Biomass in Amazonian Forests at São Gabriel Da Cachoeira in the Upper Rio Negro, Brazil. *For. Ecol. Manag.* **2012**, *277*, 163–172. [CrossRef]
7. Boyd, D.S.; Danson, F.M. Satellite Remote Sensing of Forest Resources: Three Decades of Research Development. *Prog. Phys. Geogr. Earth Environ.* **2005**, *29*, 1–26. [CrossRef]
8. Véga, C.; Renaud, J.-P.; Durrieu, S.; Bouvier, M. On the Interest of Penetration Depth, Canopy Area and Volume Metrics to Improve Lidar-Based Models of Forest Parameters. *Remote Sens. Environ.* **2016**, *175*, 32–42. [CrossRef]
9. Lahssini, K.; Dayal, K.R.; Durrieu, S.; Monnet, J.-M. Joint Use of Airborne LiDAR Metrics and Topography Information to Estimate Forest Parameters via Neural Networks. In Proceedings of the 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON), Palermo, Italy, 14–16 June 2022; pp. 442–447.
10. Hancock, S.; Armston, J.; Hofton, M.; Sun, X.; Tang, H.; Duncanson, L.I.; Kellner, J.R.; Dubayah, R. The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions. *Earth Space Sci.* **2019**, *6*, 294–310. [CrossRef]
11. Karasiak, N.; Sheeren, D.; Fauvel, M.; Willm, J.; Dejoux, J.-F.; Monteil, C. Mapping Tree Species of Forests in Southwest France Using Sentinel-2 Image Time Series. In Proceedings of the 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Brugge, Belgium, 27–29 June 2017; pp. 1–4.
12. Grabska, E.; Hostert, P.; Pflugmacher, D.; Ostapowicz, K. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sens.* **2019**, *11*, 1197. [CrossRef]

13. Quegan, S.; Le Toan, T.; Chave, J.; Dall, J.; Exbrayat, J.-F.; Minh, D.H.T.; Lomas, M.; D'Alessandro, M.M.; Paillou, P.; Papathanassiou, K.; et al. The European Space Agency BIOMASS Mission: Measuring Forest above-Ground Biomass from Space. *Remote Sens. Environ.* **2019**, *227*, 44–60. [CrossRef]
14. Lahssini, K.; Teste, F.; Dayal, K.R.; Durrieu, S.; Ienco, D.; Monnet, J.-M. Combining LiDAR Metrics and Sentinel-2 Imagery to Estimate Basal Area and Wood Volume in Complex Forest Environment via Neural Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 4337–4348. [CrossRef]
15. Morin, D.; Planells, M.; Baghdadi, N.; Bouvet, A.; Fayad, I.; Le Toan, T.; Mermoz, S.; Villard, L. Improving Heterogeneous Forest Height Maps by Integrating GEDI-Based Forest Height Information in a Multi-Sensor Mapping Process. *Remote Sens.* **2022**, *14*, 2079. [CrossRef]
16. Dubayah, R.; Blair, J.B.; Goetz, S.; Fatoyinbo, L.; Hansen, M.; Healey, S.; Hofton, M.; Hurtt, G.; Kellner, J.; Luthcke, S.; et al. The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography. *Sci. Remote Sens.* **2020**, *1*, 100002. [CrossRef]
17. Baghdadi, N.N.; El Hajj, M.; Bailly, J.-S.; Fabre, F. Viability Statistics of GLAS/ICESat Data Acquired Over Tropical Forests. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1658–1664. [CrossRef]
18. Fayad, I.; Baghdadi, N.; Riedi, J. Quality Assessment of Acquired GEDI Waveforms: Case Study over France, Tunisia and French Guiana. *Remote Sens.* **2021**, *13*, 3144. [CrossRef]
19. Herzfeld, U.C.; McDonald, B.W.; Wallin, B.F.; Neumann, T.A.; Markus, T.; Brenner, A.; Field, C. Algorithm for Detection of Ground and Canopy Cover in Micropulse Photon-Counting Lidar Altimeter Data in Preparation for the ICESat-2 Mission. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2109–2125. [CrossRef]
20. Fayad, I.; Baghdadi, N.; Lahssini, K. An Assessment of the GEDI Lasers' Capabilities in Detecting Canopy Tops and Their Penetration in a Densely Vegetated, Tropical Area. *Remote Sens.* **2022**, *14*, 2969. [CrossRef]
21. Fayad, I.; Ienco, D.; Baghdadi, N.; Gaetano, R.; Alvares, C.A.; Stape, J.L.; Ferraço Scolforo, H.; Le Maire, G. A CNN-Based Approach for the Estimation of Canopy Heights and Wood Volume from GEDI Waveforms. *Remote Sens. Environ.* **2021**, *265*, 112652. [CrossRef]
22. Ho Tong Minh, D.; Le Toan, T.; Rocca, F.; Tebaldini, S.; Villard, L.; Réjou-Méchain, M.; Phillips, O.L.; Feldpausch, T.R.; Dubois-Fernandez, P.; Scipal, K.; et al. SAR Tomography for the Retrieval of Forest Biomass and Height: Cross-Validation at Two Tropical Forest Sites in French Guiana. *Remote Sens. Environ.* **2016**, *175*, 138–147. [CrossRef]
23. Fayad, I.; Baghdadi, N.; Bailly, J.-S.; Barbier, N.; Gond, V.; Hajj, M.; Fabre, F.; Bourgine, B. Canopy Height Estimation in French Guiana with LiDAR ICESat/GLAS Data Using Principal Component Analysis and Random Forest Regressions. *Remote Sens.* **2014**, *6*, 11883–11914. [CrossRef]
24. Fayad, I.; Baghdadi, N.; Bailly, J.-S.; Barbier, N.; Gond, V.; Hérault, B.; El Hajj, M.; Fabre, F.; Perrin, J. Regional Scale Rain-Forest Height Mapping Using Regression-Kriging of Spaceborne and Airborne LiDAR Data: Application on French Guiana. *Remote Sens.* **2016**, *8*, 240. [CrossRef]
25. El Moussawi, I.; Ho Tong Minh, D.; Baghdadi, N.; Abdallah, C.; Jomaah, J.; Strauss, O.; Lavalle, M. L-Band UAVSAR Tomographic Imaging in Dense Forests: Gabon Forests. *Remote Sens.* **2019**, *11*, 475. [CrossRef]
26. El Hajj, M.; Baghdadi, N.; Labrière, N.; Bailly, J.-S.; Villard, L. Mapping of Aboveground Biomass in Gabon. *Comptes Rendus Geosci.* **2019**, *351*, 321–331. [CrossRef]
27. Memiaghe, H.R.; Lutz, J.A.; Korte, L.; Alonso, A.; Kenfack, D. Ecological Importance of Small-Diameter Trees to the Structure, Diversity and Biomass of a Tropical Evergreen Forest at Rabi, Gabon. *PLoS ONE* **2016**, *11*, e0154988. [CrossRef]
28. Fayad, I.; Baghdadi, N.; Frappart, F. Comparative Analysis of GEDI's Elevation Accuracy from the First and Second Data Product Releases over Inland Waterbodies. *Remote Sens.* **2022**, *14*, 340. [CrossRef]
29. Dubayah, R.O.; Sheldon, S.L.; Clark, D.B.; Hofton, M.A.; Blair, J.B.; Hurtt, G.C.; Chazdon, R.L. Estimation of Tropical Forest Height and Biomass Dynamics Using Lidar Remote Sensing at La Selva, Costa Rica: Forest Dynamics Using Lidar. *J. Geophys. Res.* **2010**, *115*, 1–7. [CrossRef]
30. Kellner, J.R.; Clark, D.B.; Hubbell, S.P. Pervasive Canopy Dynamics Produce Short-Term Stability in a Tropical Rain Forest Landscape. *Ecol. Lett.* **2009**, *12*, 155–164. [CrossRef]
31. Slik, J.W.F.; Aiba, S.-I.; Brearley, F.Q.; Cannon, C.H.; Forshed, O.; Kitayama, K.; Nagamasu, H.; Nilus, R.; Payne, J.; Paoli, G.; et al. Environmental Correlates of Tree Biomass, Basal Area, Wood Specific Gravity and Stem Density Gradients in Borneo's Tropical Forests: Forest Carbon and Structure Gradients. *Glob. Ecol. Biogeogr.* **2010**, *19*, 50–60. [CrossRef]
32. Chave, J.; Olivier, J.; Bongers, F.; Châtelet, P.; Forget, P.-M.; van der Meer, P.; Norden, N.; Riéra, B.; Charles-Dominique, P. Above-Ground Biomass and Productivity in a Rain Forest of Eastern South America. *J. Trop. Ecol.* **2008**, *24*, 355–366. [CrossRef]
33. Réjou-Méchain, M.; Tymen, B.; Blanc, L.; Fauset, S.; Feldpausch, T.R.; Monteagudo, A.; Phillips, O.L.; Richard, H.; Chave, J. Using Repeated Small-Footprint LiDAR Acquisitions to Infer Spatial and Temporal Variations of a High-Biomass Neotropical Forest. *Remote Sens. Environ.* **2015**, *169*, 93–101. [CrossRef]
34. Vincent, G.; Sabatier, D.; Blanc, L.; Chave, J.; Weissenbacher, E.; Pélissier, R.; Fonty, E.; Molino, J.-F.; Couteron, P. Accuracy of Small Footprint Airborne LiDAR in Its Predictions of Tropical Moist Forest Stand Structure. *Remote Sens. Environ.* **2012**, *125*, 23–33. [CrossRef]

35. Adam, M.; Urbazaev, M.; Dubois, C.; Schmullius, C. Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters. *Remote Sens.* **2020**, *12*, 3948. [CrossRef]
36. Hilbert, C.; Schmullius, C. Influence of Surface Topography on ICESat/GLAS Forest Height Estimation and Waveform Shape. *Remote Sens.* **2012**, *4*, 2210–2235. [CrossRef]
37. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping Global Forest Canopy Height through Integration of GEDI and Landsat Data. *Remote Sens. Environ.* **2021**, *253*, 112165. [CrossRef]
38. Dorado-Roda, I.; Pascual, A.; Godinho, S.; Silva, C.; Botequim, B.; Rodríguez-Gonzálvez, P.; González-Ferreiro, E.; Guerra-Hernández, J. Assessing the Accuracy of GEDI Data for Canopy Height and Aboveground Biomass Estimates in Mediterranean Forests. *Remote Sens.* **2021**, *13*, 2279. [CrossRef]
39. Neuenschwander, A.; Pitts, K. The ATL08 Land and Vegetation Product for the ICESat-2 Mission. *Remote Sens. Environ.* **2019**, *221*, 247–259. [CrossRef]
40. Xi, Z.; Xu, H.; Xing, Y.; Gong, W.; Chen, G.; Yang, S. Forest Canopy Height Mapping by Synergizing ICESat-2, Sentinel-1, Sentinel-2 and Topographic Information Based on Machine Learning Methods. *Remote Sens.* **2022**, *14*, 364. [CrossRef]
41. Fayad, I.; Baghdadi, N.; Alcarde Alvares, C.; Stape, J.L.; Bailly, J.S.; Scolforo, H.F.; Cegatta, I.R.; Zribi, M.; Le Maire, G. Terrain Slope Effect on Forest Height and Wood Volume Estimation from GEDI Data. *Remote Sens.* **2021**, *13*, 2136. [CrossRef]
42. Escobar Villanueva, J.R.; Iglesias Martínez, L.; Pérez Montiel, J.I. DEM Generation from Fixed-Wing UAV Imaging and LiDAR-Derived Ground Control Points for Flood Estimations. *Sensors* **2019**, *19*, 3205. [CrossRef]
43. Fayad, I.; Baghdadi, N.N.; Alvares, C.A.; Stape, J.L.; Bailly, J.S.; Scolforo, H.F.; Zribi, M.; Maire, G.L. Assessment of GEDI's LiDAR Data for the Estimation of Canopy Heights and Wood Volume of Eucalyptus Plantations in Brazil. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7095–7110. [CrossRef]
44. Lang, N.; Kalischek, N.; Armston, J.; Schindler, K.; Dubayah, R.; Wegner, J.D. Global Canopy Height Regression and Uncertainty Estimation from GEDI LIDAR Waveforms with Deep Ensembles. *Remote Sens. Environ.* **2022**, *268*, 112760. [CrossRef]
45. Liu, A.; Cheng, X.; Chen, Z. Performance Evaluation of GEDI and ICESat-2 Laser Altimeter Data for Terrain and Canopy Height Retrievals. *Remote Sens. Environ.* **2021**, *264*, 112571. [CrossRef]
46. Payn, T.; Carnus, J.-M.; Freer-Smith, P.; Kimberley, M.; Kollert, W.; Liu, S.; Orazio, C.; Rodriguez, L.; Silva, L.N.; Wingfield, M.J. Changes in Planted Forests and Future Global Implications. *For. Ecol. Manag.* **2015**, *352*, 57–67. [CrossRef]

Chapter 3

Use of GEDI Signal and Environmental Parameters to Improve Canopy Height Estimation over Tropical Forest Ecosystems in Mayotte Island

The previous study discussed several approaches for estimating canopy height using GEDI data, focusing the impact of GEDI acquisition and processing parameters on the accuracy of the estimates. Two primary methods were explored: on one hand, direct estimation using a single GEDI height metric as a direct proxy of canopy height; on the other hand, statistical and machine learning models based on waveform metrics and signal parameters. The results indicate that laser beam type, beam sensitivity, as well as the algorithm setting group significantly influence the accuracy of canopy height estimates from GEDI data, with power beams and algorithm setting group number 5 being more effective in dense tropical forests due to laser penetration.

Following this analysis, this chapter presents the second study of this dissertation to address the second research question: how do environmental factors, such as terrain configuration and vegetation structure, affect GEDI measurements? In remote sensing applications, both the sensor and the sensed environment influence the data acquired by instruments. After gaining a better understanding of GEDI acquisition and processing parameters, the next important step is to assess how the characteristics of tropical environments impact the canopy heights measured by GEDI. This assessment is part of the broader objective of this doctoral thesis which aims at developing a methodology for canopy height mapping based on GEDI data.

The backscattered signals retrieved by GEDI and the subsequent forest variables and height metrics derived from the waveforms may contain uncertainties due to the signal distortion

caused by the topographic conditions within the footprints. Understanding and correcting slope effects is a mandatory step when performing studies over steep areas where slope strongly impacts GEDI measurements. Moreover, the vegetation structure also has a direct influence on the data collected by GEDI. In tropical biomes, LiDAR beam penetration capability is strongly dependent on forest characteristics, and the horizontal structure of the vegetation, whether it is heterogeneous, can exacerbate errors related to GEDI geolocation uncertainties.

In summary, this chapter analyzes how environmental parameters related to terrain and vegetation structure influence the canopy height estimates obtained from GEDI data over heterogeneous tropical forests in Mayotte Island (Comoro Archipelago). It further continues the improvement of canopy height estimates initiated in the previous chapter by tackling the second research question of the dissertation. By assessing the behavior of GEDI over a different tropical environment with its own challenges, this study highlights the necessity to adapt to the specific environmental context of each study site.

CANADIAN JOURNAL OF REMOTE SENSING
2024, VOL. 50, NO. 1, 2351004
https://doi.org/10.1080/07038992.2024.2351004

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Use of GEDI Signal and Environmental Parameters to Improve Canopy Height Estimation over Tropical Forest Ecosystems in Mayotte Island

Utilisation du signal GEDI et des paramètres environnementaux pour améliorer l'estimation de la hauteur de la canopée dans les écosystèmes forestiers tropicaux à Mayotte

Kamel Lahssini[a], Nicolas Baghdadi[a], Guerric Le Maire[b,c], Stéphane Dupuy[d,e], and Ibrahim Fayad[f,g]

[a]TETIS, INRAE, Montpellier 34093, France; [b]Eco&Sols, CIRAD, Montpellier 34398, France; [c]Eco&Sols, Université de Montpellier, CIRAD, INRAE, IRD, Institut Agro, Montpellier 34093, France; [d]CIRAD, UMR TETIS, Montpellier F-34398, France; [e]TETIS, Université de Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, Montpellier, France; [f]Kayrros SAS, Paris75009, France; [g]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS9 UVSQ, Université Paris-Saclay, Gif-sur-Yvette 91191, France

ABSTRACT

Canopy height is a fundamental parameter for describing forest ecosystems. GEDI is a spaceborne LiDAR system that was designed to measure vegetation's vertical structure at a global scale. This study evaluates the accuracy of GEDI-derived canopy height estimates over complex tropical forests in Mayotte Island (Overseas France) characterized by moderate height and biomass levels as well as a relatively steep terrain. The influence of GEDI signal and environmental parameters (canopy height, beam sensitivity and slope) on height estimates was assessed. Linear as well as non-linear approaches were implemented using the GEDI L2A product to estimate canopy height. Empirical models were trained on reference data derived from airborne LiDAR scanning. The results showed that using regression models built on multiple GEDI metrics yielded improved accuracies compared to a direct estimation from a single GEDI height metric. Canopy height, beam sensitivity and terrain slope were found to have a significant impact on the height metrics derived from GEDI waveforms. Conversely, both linear and non-linear regression models produced unbiased and stable estimates.

RÉSUMÉ

La hauteur de la canopée est un paramètre fondamental pour décrire les écosystèmes forestiers. GEDI est un système LiDAR spatial conçu pour mesurer la structure verticale de la végétation à l'échelle mondiale. Cette étude évalue la précision des estimations de la hauteur de la canopée à partir de GEDI sur des forêts tropicales complexes de l'île de Mayotte (France d'outre-mer) caractérisées par des hauteurs et des niveaux de biomasse modérés ainsi que par un terrain relativement escarpé. L'influence du signal GEDI et des paramètres environnementaux (hauteur de la canopée, sensibilité du faisceau laser et pente du terrain) sur les estimations de hauteur a été évaluée. Des approches linéaires et non-linéaires ont été mises en œuvre en utilisant le produit GEDI L2A pour estimer la hauteur de la canopée. Des modèles empiriques ont été entraînés sur des données de référence issues d'acquisitions par LiDAR aéroporté. Les résultats ont montré que l'utilization de modèles de régression construits à partir de plusieurs métriques GEDI permettait d'améliorer la précision par rapport à une estimation directe à partir d'une seule métrique de hauteur GEDI. La hauteur de la canopée, la sensibilité du faisceau et la pente ont eu un impact significatif sur les métriques de hauteur dérivées des formes d'onde GEDI. Par ailleurs, les modèles de régression linéaire et non-linéaire ont produit des estimations stables et sans biais.

ARTICLE HISTORY
Received 6 December 2023
Accepted 29 April 2024

Introduction

Tropical moist forests play a critical role in maintaining natural balances by serving as global carbon storage reservoirs. They are natural carbon dioxide sinks and account for more than 40% of the world's terrestrial carbon stock (Pan et al. 2011). Measuring

CONTACT Kamel Lahssini kamel.lahssini@gmail.com
© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

the standing aboveground biomass (AGB) in these forests is a fundamental step in assessing their carbon sequestration potential. AGB refers to the quantity of living vegetation above the soil, including stem, stump, branches, bark, seed and foliage. A number of studies have established allometric relationships that link the measurable structural characteristics of a forest to its AGB amounts (Asner and Mascaro 2014; Chave et al. 2005). Canopy height is a crucial parameter for quantifying biomass, as most allometric equations rely on it to derive AGB (Lefsky et al. 2005). Indeed, multiple authors have found that including canopy height in AGB prediction models can significantly reduce estimation errors and improve accuracy (Lima et al. 2012; Feldpausch et al. 2012), and that allometric relationships based on canopy height only may be used to predict AGB (Lefsky et al. 2005).

Remote sensing data has been proven to be an effective means of estimating forest characteristics at both regional and global scales over large areas that would be difficult to study otherwise (Boyd and Danson 2005). Light detection and ranging (LiDAR) technology is particularly well-suited for characterizing forest height and vertical structures. By emitting laser pulses and retrieving return signals, LiDAR can measure the three-dimensional structure of the environment with precision. In particular, full waveform (FW) systems are able to record the complete profiles of the return signals by sampling them in constant time intervals (Wehr and Lohr 1999). These systems provide geolocated one-dimensional temporal signals (referred to as return or received waveforms) describing the vertical structure of the vegetation at specific geolocations, from which a variety of forest parameters can be extracted. Airborne and spaceborne LiDARs are the two primary systems used to acquire data to describe forest vertical structures. The main advantage of airborne sensors is their high resolution (i.e., the number of returned points over a given surface), but they come at a high financial cost for data users which restricts their usage to limited areas and dates. Conversely, spaceborne LiDAR data are freely accessible and provide global coverage, but they produce low density information (spatial coverage of about 4% of the Earth's surface for the latest system) and their high operational altitudes make them more sensitive to difficult atmospheric conditions (Dubayah et al. 2020; Baghdadi et al. 2014). The latest operating spaceborne LiDAR system providing data at a global scale is the Global Ecosystem Dynamics Investigation (GEDI) embedded on the International Space Station (ISS). By emitting laser pulses that pass through the atmosphere and interact with the objects on the Earth's surface, GEDI records the return signals resulting from backscattering. These received waveforms are a direct proxy of the vegetation's vertical structure and descriptive metrics can then be extracted to characterize vegetation parameters.

In tropical biomes, characterizing forest parameters from GEDI data poses important challenges due to the dense and complex nature of these ecosystems. To address these issues, statistical methods have been developed and implemented to predict canopy heights from GEDI data (Dorado-Roda et al. 2021; Adrah et al. 2022; Gupta and Sharma 2022). Accurately describing the vertical structure of forests from top-of-canopy to the ground using GEDI depends directly on the system's ability to penetrate through the vegetation all the way to the ground. The assessment of GEDI's capability to measure canopy height has highlighted the importance of signal parameters such as beam sensitivity and intensity when dealing with densely vegetated areas (Adam et al. 2020; Ngo et al. 2022). Wang et al. (2022) concluded that the higher the waveform sensitivity, the lower the errors in GEDI canopy height estimation. More precisely, Fayad et al. (2022) found that GEDI shots with beam sensitivity greater than 98% exhibited a significantly greater ability to detect the ground and canopy tops in a tropical context. Moreover, terrain configuration has a strong impact on the waveforms acquired by GEDI and the effects of slope on the waveforms need to be accounted for and compensated, especially over steep and rugged areas. Topographic slope is indeed a critical factor affecting the precision of canopy height estimates using spaceborne LiDAR GEDI (Liu et al. 2021; Fayad et al. 2021a). For example, the waveform extent may stretch over terrain with significant elevation differences, causing an overestimation of the relevant vegetation features (Chen 2010; Kutchartt et al. 2022). Dhargay et al. (2022) investigated the impact of terrain slope on GEDI-derived canopy height in complex forest ecosystems of south-eastern Australia, revealing an increasing trend in estimation error with steeper slopes. Liu et al. (2021) noted that regions characterized by dense canopy cover or steep slopes pose significant challenge for retrieving accurate canopy height information from GEDI. Similarly, in a study over multiple Eucalyptus plantations, Fayad et al. (2021b) observed a 14% increase in RMSE for the estimation of dominant height in areas with slopes exceeding 20% compared to slopes ranging between 10 and 20%. Finally, the vegetation structure may also affect GEDI measurements and, in particular, tree heights and cover fraction have a direct impact on the metrics extracted from GEDI waveforms. Schlund et al. (2022)

built empirical models from GEDI data to estimate vegetation heights and noted that structural attributes like the canopy height itself and forest cover had a significant effect on the accuracies of height estimates, with higher and denser vegetation generally resulting in higher errors. Reciprocally, when dealing with lower tree heights, the ground return may blend with the vegetation, also causing errors in canopy height estimates. For example, Ilangakoon et al. (2018) observed that using GEDI waveforms to differentiate bare ground from low-height vegetation such as shrubs could be challenging because the waveforms share the same characteristics.

This paper aims at evaluating the impact of vegetation height, beam sensitivity and slope on the estimation of canopy heights over tropical forests in Mayotte Island (Overseas France), a particularly complex context for GEDI measurements due to its spatial heterogeneity, terrain slope (mean slope of 20°), and relatively small tree height with an average height of 15 m (Dupuy et al. 2013). The first objective is to understand how these parameters impact the data retrieved by the GEDI sensor and, consequently, the height metrics derived from the received waveforms. The context of Mayotte Island is particularly interesting as it is one of the last islands to have forest complexes in this part of the western Indian Ocean. Secondly, in the light of this understanding, the second objective is to evaluate how these parameters can be integrated as input of prediction models to improve canopy height estimates from GEDI data. Based on the results of this analysis, we aim at offering valuable recommendations to GEDI users on how to utilize their data effectively depending on their specific work configurations, goals and objectives. The information contained in GEDI products are indeed abundant and can be leveraged in various ways to derive indicators that describe the structure of forest ecosystems. In order to do so, GEDI-derived canopy height estimates are compared to reference values from a canopy height model (CHM) obtained from airborne LiDAR scanning (ALS) data.

Materials and methods

Study area

The study was conducted over two forests in the island of Mayotte. Mayotte Island (Comoro Archipelago) is an overseas department of France located in the Indian Ocean. It is one of the few remaining islands in the western Indian Ocean with forest ecosystems. These ecosystems are primarily located in five forested areas. The areas of interest considered in this study are located over the Dapani state forest and the Majimbini departmental forest (Figure 1).

Dapani forest is a vital natural ecosystem that provides numerous ecological, economic and social benefits to the region. Located in the southern part of the island, the forest covers an area of approximately 340 hectares, on a relatively hilly terrain that ranges from sea level to 500 m above sea level (asl). The forest is a unique habitat characterized by a rich

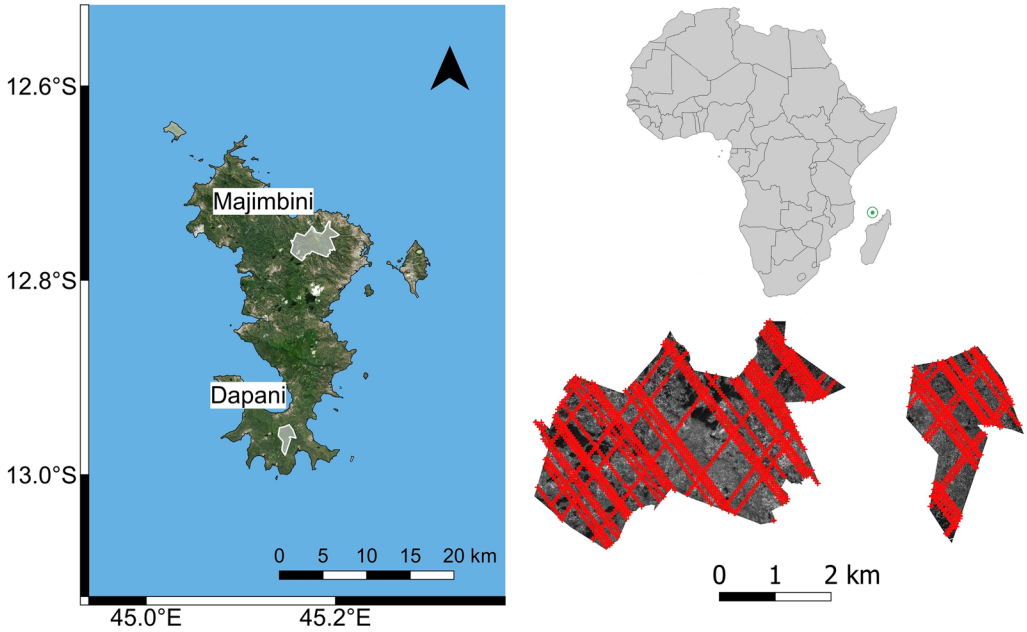


Figure 1. Location of the two study sites in Mayotte Island (ESRI Satellite®) and GEDI footprints over ALS canopy height.

diversity of endemic and endangered plant and animal species, including several species of lemurs, bats and birds (Gargominy 2003). The climate in Dapani forest is typically tropical (group Aw according to Köppen climate classification), with two distinct seasons: the dry season from June to October and the wet season from November to May. The average annual rainfall in the southern regions of Mayotte amounts to about 1,000 mm (Lachassagne et al. 2014). The forest is dominated by shallow to deep soil and the vegetation is characterized by a dense canopy of evergreen trees and shrubs that provide vital carbon sequestration and oxygen production services (Pascal and Labat 2002).

Majimbini Forest, situated on the northeast part of the island, is also a crucial ecosystem for biodiversity conservation in the region. The forest covers an area of approximately 1,200 hectares and is characterized by a variety of vegetation types, including dense forests, shrublands and grasslands (Pascal and Labat 2002). Over 500 plant species are recorded in the area, including several endemic species (Gargominy 2003). Majimbini forest is subject to a tropical climate (group Aw according to Köppen climate classification), although more humid than Dapani forest with an average annual rainfall of approximately 2,000 mm in the northern parts of Mayotte Island (Lachassagne et al. 2014). The topography of Majimbini forest is characterized by a series of rolling hills, with the highest point in the forest reaching approximately 500 m asl.

The two study sites correspond to zones with similar structural properties in terms of average heights and biomass levels, with an average canopy height of 15 m and a mean aboveground biomass (AGB) ranging between 100 and 150 Mg/ha (Santoro et al. 2021). Even though global AGB maps such as the ones we considered are produced with quite large uncertainties (error between 50 and 100 Mg/ha in average), they still allow drawing conclusions on the relatively moderate AGB levels that characterize Mayotte's forest ecosystems and that are significantly lower than the known saturation levels of LiDAR spaceborne sensors (Duncanson et al. 2022; Shendryk 2022). For the purpose of this study, all the data related to each site were merged together to produce a single database describing forests in Mayotte's tropical context.

Data and processing

GEDI data

GEDI is a FW LiDAR sensor that was specifically designed to measure the vertical structure of forest ecosystems (Dubayah et al. 2020). Between April 2019 and April 2023, GEDI has been gathering data using three 1064 nm lasers that emit 242 pulses per second. One laser is split into coverage beams and the other two remain as power beams. All the beams are then slightly dithered to create eight parallel tracks of observations over a 4.2 km swath on the ground. The beams illuminate circular footprints of 25 m in diameter on the Earth's surface and the waveforms of the return signals are recorded to measure the vertical structure of the vegetation. The profiles of the return signals are sampled at a fixed time interval of 1 ns which corresponds to a 15 cm sampling distance.

The GEDI data used in this study were processed by NASA's Land Processes Distributed Active Archive Center (LP DAAC). The datasets consist of two processing levels, namely L1B (Level 1) and L2A (Level 2). The L1B product provides geolocated and smoothed waveforms along with their ancillary parameters, with an expected geolocation accuracy of about 10 m (Roy et al. 2021). The L2A product provides elevation and height metrics at the footprint-level, such as ground elevation, canopy top height and relative height (RH) metrics. The L2A product is derived from the received waveforms in the L1B product using six possible signal processing configurations known as algorithm setting groups. These groups determine the thresholds and smoothing settings used to interpret the received waveforms, which in turn affects the height metrics computed in the L2A product. In the context of this study, we utilized the L2A metrics that were computed using algorithm setting group number 2.

GEDI shots (i.e., L1B and L2A data associated with a footprint on the ground) acquired in the period between April 2019 and March 2023 were downloaded over the two areas of interest considered in this study. To ensure the validity of the measurements obtained by the GEDI sensor and considering the potential negative impact of atmospheric conditions, various filters were applied to remove unusable and irrelevant shots before conducting the analysis:

- Shots with no mode detected ($num_detected\text{-}modes = 0$) were removed. A signal with no mode is just pure noise and does not contain any information related to the vertical structure of the forest.
- Shots with a null SNR ($SNR = 0$) were removed. These shots also correspond to pure noise.
- Shots with an incorrect detection of the ground were removed. The quality assessment of the ground detection from GEDI data are performed using the corresponding Shuttle Radar

Topography Mission (SRTM) digital elevation model (DEM). Even though a more accurate ALS DEM is available, we chose to use the SRTM product to place our study in an operational context where accurate products are not always available for users. Moreover, this filtering step does not require a very accurate DEM product and the filtering performed with the ALS DEM did not change the output. If the absolute difference between the elevation of the GEDI lowest mode (*elev_lowestmode*) and the SRTM DEM (*digital_elevation_model_srtm*) is greater than 100 m, then the shot is discarded.

- Shots with an incomplete waveform were removed. An incomplete waveform is a partial signal that does not have a sufficient number of bins to be interpretable. If the end location of the usable portion of the waveform (*search_end*) is equal to the total number of bins in the waveform (*rx_sample_count*), then the waveform is considered incomplete.
- Shots with a distance between the canopy top and the ground return (*rh_100*) lower than 3 m were removed. GEDI shows certain insensitivity to estimate the height of short objects. In fact, GEDI's lasers fire shots with a pulse width of about 15 ns, which corresponds to a distance between the first emitted photons and the last of approximately 4.5 m. In an ideal scenario (perfect reflectance and no atmospheric perturbations), the returned waveform over a bare ground would have a distance between the *toploc* (first returned photons) and *botloc* (last returned photons) of approximately 4.5 m, and the derived *rh_100* would amount to about 2.25 m (distance between *toploc* and the ground which is located in the middle between *toploc* and *botloc*). Due to noise and other sources of perturbations, this insensitivity of GEDI may increase to around 3 m.

After the filtering steps, and considering the availability of ground truth data, a total of 3384 GEDI footprints were retained out of a total of 19,272 shots (about 17%) for the combined analysis of the two study sites.

ALS data

In order to evaluate the accuracy of GEDI data in predicting canopy heights, a CHM derived from ALS data was utilized as ground truth. ALS data were acquired through surveys conducted in October 2008 by the French National Geographic Institute (IGN) in the context of the Litto3D project (Dupuy et al. 2013). All the LiDAR echoes were recorded and IGN then used automatic and interactive filtering to extract the first and last returns from the point clouds. The first returns are typically associated with the canopy top in forested areas, while the last returns usually indicate the ground, even though it is not always the case in densely vegetated areas. Therefore, IGN also conducted significant interactive processing to verify and reclassify points. A digital surface model (DSM) and a digital terrain model (DTM) at a resolution of 1 m were finally produced using TerraScan software (Terrasolid Ltd., Finland). The CHM used in this study was obtained by firstly calculating the difference between the DSM and the DTM, and then applying a rank-order operator median filter to generate the final CHM product (Dupuy et al. 2013).

Dapani and Majimbini forests are located in protected areas and consist mainly of mature old-growth and secondary woodlands. Nevertheless, given the time span between the acquisition of ground truth data and the GEDI mission, vegetation growth and changes occurred. There are no in-situ measurements of forest growth based on permanent inventory plots for these forests, which makes it impossible for a precise evaluation of changes in canopy height. Considering the rough estimate of forest AGB increase of about 1% per year in Mayotte Island (Requena Suarez et al. 2019), we expect an average height growth of 1 to 2 m in a ten-year time span considering simple allometric relationships for tropical forests (Chave et al. 2014). Some studies have investigated canopy dynamics over extended time periods in old-growth tropical rainforest landscapes. Dubayah et al. (2010) utilized two LiDAR datasets from 1998 and 2005 acquired over La Selva, a tropical biome in Costa Rica (Central America), and reported a net height loss of −0.33 m for old-growth forests and a net gain of 2.08 m for secondary forests over a 7-year time span. Similarly, Kellner et al. (2009) analyzed two LiDAR datasets acquired over La Selva (in 1997 and 2006) and found a mean height change of −0.32 m for old-growth landscapes over an 8.5-year period. Although the height did not vary significantly in average, local variations can still occur. For example, Dubayah et al. (2010) reported local variations of approximately +/−3 m for old-growth forests over a 7-year time span.

To summarize, in the present study, the discrepancies between GEDI-derived heights and ALS CHM include the possible evolutions of the forests between the acquisition dates, together with various other

factors, such as GEDI geolocation accuracy and measurement uncertainties. In the following section, we describe a filter that was applied to exclude GEDI footprints over areas where disturbances (i.e., height decrease) occurred in the period between ALS and GEDI acquisitions.

Ancillary data

Even though Dapani and Majimbini forests are located in protected areas, they are still under threat from human activities such as logging, agriculture and urbanization. Moreover, intense climatic events such as cyclones, typhoons or fires can also cause significant degradations to these ecosystems. Areas where such significant changes in height occurred need to be removed from the analysis to produce a relevant comparison between GEDI data and reference heights.

The dataset produced by the European Commission's Joint Research Center (JRC) on forest cover change in tropical moist forests (TMF dataset) is a highly structured and comprehensive record of forest cover and land use change in these ecosystems (Vancutsem et al. 2021). It is compiled from optical satellite imagery spanning several decades and covers a vast area of the tropics, including the western Indian Ocean and Mayotte Island. The dataset includes information on forest area, forest cover change, and forest loss and degradation, as well as data on human activities and infrastructure that may be contributing to these changes. For the purpose of this study, we focused on the zones of the Dapani and Majimbini forests that remained undisturbed since the year 2008. An undisturbed zone is defined as an area that did not undergo any degradation or deforestation since a given year. In this way, our study focuses on old-growth and secondary undisturbed forested areas only and, therefore, the dispersion that will be observed between GEDI-derived heights and the reference CHM cannot be linked to potential major disturbances in forest cover during the time span between acquisitions.

The TMF dataset allowed restricting the study sites to the extent of undisturbed forested areas and, out of the 3384 GEDI shots that were firstly retained after the preliminary filters, a total of 2397 footprints are finally available for the assessment of canopy height estimates performed in this analysis.

Methods

The study presented herein has two main objectives. First, using a single GEDI height metric as a direct proxy of canopy height, a comparative assessment of GEDI's capabilities in predicting canopy heights is performed. Canopy height, beam sensitivity and slope are studied independently from each other in order to evaluate their specific influence on GEDI data. Through this analysis, we aim at understanding how signal physical parameters as well as environmental configuration can impact the waveforms acquired through GEDI and thus the canopy height estimates resulting from their processing. Next, in the light of this preliminary analysis, we evaluate several regression approaches based on multiple GEDI metrics to improve the accuracy of canopy height prediction models. The aforementioned GEDI signal and environmental parameters are also included in the construction of these models to assess how their impact can be taken into account and how they can contribute to reaching better accuracies. Specifically for slope, a number of studies have been carried out to address this particular issue and to integrate slope information in estimation models. Some works retrieve terrain indices that describe the elevation configuration and directly utilize them as inputs in prediction models (Chen 2010; Xing et al. 2010). Others rely on the computation of slope-corrected waveforms to minimize the effects of slope and produce simulated metrics that are later used in prediction models (Wang et al. 2019; Fayad et al. 2021a). In the context of this study, a multilinear regression model and Random Forest (RF) estimators are implemented to take advantage of the information contained in GEDI waveforms. Specifically, to investigate the possible outcomes related to the integration of slope information, we build a RF model based on actual GEDI metrics on one hand, and a RF model trained on simulated metrics on the other hand.

GEDI and ALS canopy heights

To determine the potential of GEDI for estimating canopy heights, a single metric from the L2A product was selected as the GEDI-derived height (referred to as GEDI-CHM). From all the available values, three relative heights were identified as potential candidates to be a direct proxy of canopy height: *rh_100*, *rh_98* or *rh_95*. Relative height metrics indicate the height relative to the ground at which a specific percentile of returned energy is attained. For instance, *rh_95* denotes the height at which 95% of the waveform energy is reached. Relative heights of high percentiles are therefore considered as good indicators of canopy heights. In this study, we assessed the use of *rh_100*, *rh_98* and *rh_95* as direct proxies of canopy height in order to select the most adapted one in accordance

with our results and previous assumptions in the literature review (Potapov et al. 2021; Dorado-Roda et al. 2021).

The accuracy of GEDI estimates is determined by comparing the relative height metric value with the corresponding reference canopy height, which is obtained from the reference CHM raster. The geolocation of the received GEDI waveforms is used to overlay GEDI footprints with the CHM raster. For each usable footprint in the GEDI L2A product, zonal statistics of the CHM raster cells contained within the extent of the footprint (a 25 m circle) are extracted. We identified two statistical values as potential candidates to represent the reference canopy height (referred to as ALS-CHM): the maximum (*als100*) or the 95th percentile (*als95*). High statistics of the CHM raster cells were chosen because they are theoretically closer to the top-of-canopy signal obtained from GEDI waveforms. Given the point density of the ALS acquisition (2 points/m^2) and the CHM raster resolution (1 m), these statistics were directly computed from the raster. Adam et al. (2020) supported the use of the maximum value by the fact that received waveforms begin at the highest point of vegetation within the footprints. In another study, Hilbert and Schmullius (2012) obtained better correlations between ALS reference heights and spaceborne LiDAR-derived height estimates when using the maximum rather than the mean value. Specifically in this study, we noted a significant difference between *als100* and *als95* for some GEDI shots, especially the ones over spatially fragmented and heterogeneous canopies. Therefore, we also assessed the use of each of these two statistics to select the one most correlated with the height information contained in GEDI waveforms.

In the end, the accuracy of GEDI estimates is achieved by comparing GEDI-derived heights (GEDI-CHM) with ALS reference heights (ALS-CHM). To perform statistical analysis and evaluate the accuracies of this approach, we calculated the root mean square error (RMSE) as well as the difference between GEDI-CHM and ALS-CHM (referred to as CHM-Differences). The *rh_95* metric proved to be the most suitable indicator of canopy height and the maximum *als100* appeared to be the most suited value for reference canopy height.

Influence of signal physical parameters

The performances of GEDI's canopy height estimation depend on the laser's ability to penetrate the vegetation and detect the ground. To estimate canopy heights using a relative height metric, it is crucial to identify the ground peak in the waveform signal. Incorrect ground peak detection could be due to the laser's inability to reach the ground or to a difficulty in isolating the ground mode from the background noise that characterizes every numeric signal (Adam et al. 2020). In this prospective, beam sensitivity is an important parameter that can help understand and characterize these situations. Sensitivity is defined as the maximum canopy cover that can be penetrated considering the SNR of the waveform. A high sensitivity allows penetrating denser canopies and thus reaching the ground. This signal physical parameter is therefore of paramount importance when extracting canopy heights from GEDI data and its influence was assessed to understand how it may affect the obtained results.

Influence of environmental parameters

The forest variables and height metrics measured by the backscattered GEDI signals may contain uncertainties due to the distortion caused by the topographic conditions within the footprints. In order to quantify the influence of terrain on canopy height estimates, a slope raster was derived from the DTM raster and resampled to a 5 m resolution. This upsampling was done to minimize the impact of extreme slope values (Wang et al. 2019; Adam et al. 2020). The mean slope value within the extent of the footprint was then calculated for each GEDI waveform. To better visualize the influence of slope on GEDI measurement accuracies, three slope classes were created to later group accuracy metrics by these categories: inferior to 15°, between 15 and 25°, superior to 25°. Furthermore, vegetation and canopy characteristics can also play a role in the shape of GEDI measurements. Consequently, in order to highlight the impact of tree height, the accuracy of GEDI's canopy height estimates was assessed depending on two classes of heights: inferior to 15 m and superior to 15 m.

Regression models for canopy height estimation

After assessing the use of a single metric to derive canopy height, we built regression models between the ALS-CHM reference height data and GEDI L2A waveform metrics, as well as acquisition and environmental parameters.

In this study, three models were trained and evaluated: a Multilinear Regression (MRH), a RF regressor (RFH) based on GEDI metrics and a RF regressor (sRFH) based on simulated metrics computed from simulated ground returns (Wang et al. 2019; Fayad

et al. 2021a). Table 1 presents a list of the models along with the metrics used for their implementation. The goal is to understand how the parameters that impact canopy height estimation can be integrated to improve the obtained accuracies. To evaluate the performances of the models, a ten-fold cross validation was used, and the RMSE and CHM-Differences were calculated.

Firstly, a MRH was implemented to predict canopy heights from a set of predictive explanatory variables. In this approach, a linear relationship is built between the target variable and the predictors. Secondly, canopy heights were also estimated through non-linear non-parametric regressions using Random Forest (RF) regressors. The RF method is a machine learning algorithm that leverages ensemble of trees through a bagging strategy to predict the target variable. Compared to MRH, RFH and sRFH have the advantage of being able to model non-linear relationships between the predictors and the variable to predict. The importance of the predictors used as input variables can be quantified in order to understand the most contributing factors in the estimation task. In this analysis, RFH and sRFH differ by the way slope information is integrated to the model: the former is built on GEDI relative height metrics and mean slope of the footprints while the latter relies on metrics extracted from simulated waveforms that already take into account the slope effects. To generate new metrics from simulated waveforms, we used a method developed by Wang et al. (2019). To begin with, for each GEDI footprint of the dataset, we simulated a waveform over a bare ground with the same slope value as the actual waveform acquired over forested area. The simulated waveform is based on a Gaussian signal such as the laser pulses emitted by FW LiDAR sensors (Fayad et al. 2021a). Over steep terrain, the waveform extent increases with slope and the ground peak exhibits a broadening of its width. To account for this, the standard deviation used in the Gaussian function to simulate a bare ground return is broadened to reflect the slope of the terrain. Once the shape of the waveform over a bare ground with known slope is determined, the simulated ground return is finally overlaid on the actual waveform by aligning the positions of the signal end of both waveforms. The superposition of the simulated ground return and the original waveform allows for the computation of new waveform metrics defined as follows:

- *HT_n*: height between the signal end and the position at which *n*% of the original waveform energy is reached.
- *sim_HG_n*: height between the signal end and the position at which *n*% of the simulated ground return energy is reached.
- *sim_RHT_n*: difference between *HT_n* and *sim_HG_n*.

Table 1. List of the models used for the estimation of canopy heights and input data.

Model	Input data
MRH	*rh_95* Slope Sensitivity
RFH	*rh_n* {10% ≤ *n* ≤ 100%, step 5%} Slope Sensitivity
sRFH	*sim_RHT_n* {10% ≤ *n* ≤ 100%, step 10%} *sim_HG_n* {10% ≤ *n* ≤ 100%, step 10%} Sensitivity

Results

GEDI and ALS canopy heights

We first analyzed the simple correlation between GEDI-derived heights and ALS reference data. The results presented in Figure 2 are scatter plots of GEDI

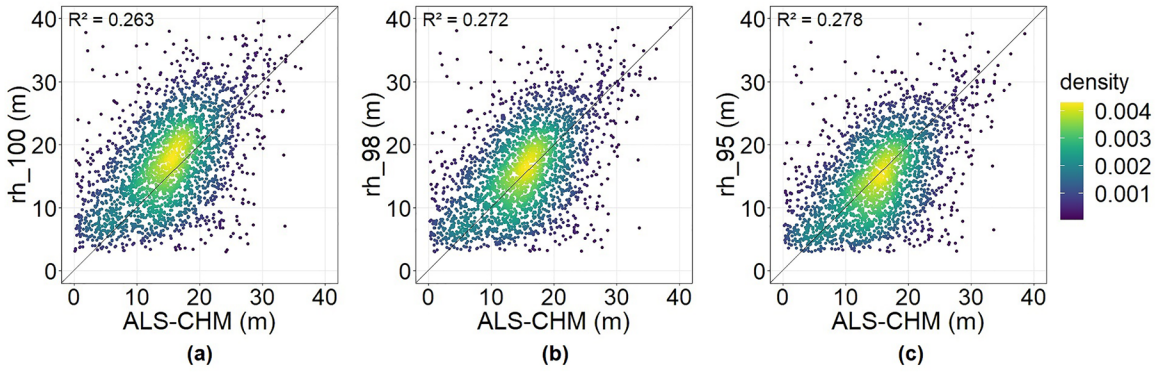


Figure 2. GEDI-CHM estimates from *rh_100* (a), *rh_98* (b) and *rh_95* (c) as a function of ALS-CHM (*als100*).

relative height metrics (*rh_100*, *rh_98* and *rh_95*) and ALS-CHM (*als100*). All RH metrics exhibit a quite limited linear correlation, with R^2 coefficients around 0.3. We note a tendency for *rh_100* to overestimate canopy heights, while *rh_98* and *rh_95* appear less biased and result in point distributions that are more centered on the bisector line. These observations are also confirmed by the accuracy metrics given in Table 2. In the light of these results, we chose for the next steps of this study to select *rh_95* as the best metric to derive GEDI-CHM.

Regarding the value used to retrieve ALS-CHM from the reference raster data, we first compared CHM-Differences with the difference between *als100* and *als95* for all GEDI footprints. Two main observations can be drawn from the results presented in Figure 3. On one hand, for GEDI shots with a relatively low difference between *als100* and *als95* (inferior to 7 m, mostly corresponding to shots over homogeneous areas), CHM-Differences remain stable, and the comparison between GEDI-CHM and ALS-CHM exhibits a null bias and an RMSE of 6.3 m. On the other hand, when considering GEDI footprints with a higher difference between *als100* and *als95* (superior to 7 m, corresponding to shots over spatially fragmented areas), CHM-Differences increase in absolute value, with a significantly stronger bias of −3.6 m and an increased RMSE of 7.7 m. Therefore, the horizontal heterogeneity of the vegetation proves to have an impact on the accuracy of canopy height estimates from GEDI data and ALS-CHM needs to be chosen accordingly.

Table 2. Accuracy of GEDI-CHM estimates (*rh_100*, *rh_98* and *rh_95*) against ALS-CHM (*als100* and *als95*).

GEDI-CHM	Bias (m)	RMSE (m)
rh_100 vs. *als100*	2.1	7.4
rh_98 vs. *als100*	0.8	6.9
rh_95 vs. *als100*	−0.3	6.6
rh_95 vs. *als95*	3.9	7.5

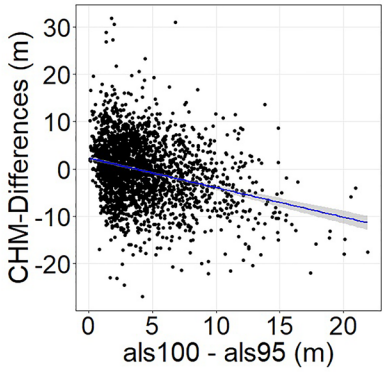


Figure 3. CHM-Differences as a function of the difference between *als100* and *als95*. The difference between *als100* and *als95* is an indicator of spatial heterogeneity.

Similarly to what was done for GEDI-CHM, Figure 4 displays scatter plots of GEDI-CHM and the ALS-CHM candidates (*als100* and *als95*). The use of *als95* for reference data induces a tendency to overestimate canopy heights, even though it shows a slightly better correlation than when *als100* is utilized. In terms of accuracy metrics, using the maximum value (i.e., *als100*) to calculate ALS-CHM allows for smaller errors and a less biased estimation, as confirmed by the results reported in Table 2.

Based on these findings, we decided to extract the maximum canopy height value *als100* within a given GEDI footprint as the ALS-CHM used for reference in this study.

Influence of canopy height, sensitivity and slope

Focusing on the influence of beam sensitivity, GEDI shots with a mean slope superior to 15° were removed in order to suppress slope effects and thus concentrate on the impact of sensitivity.

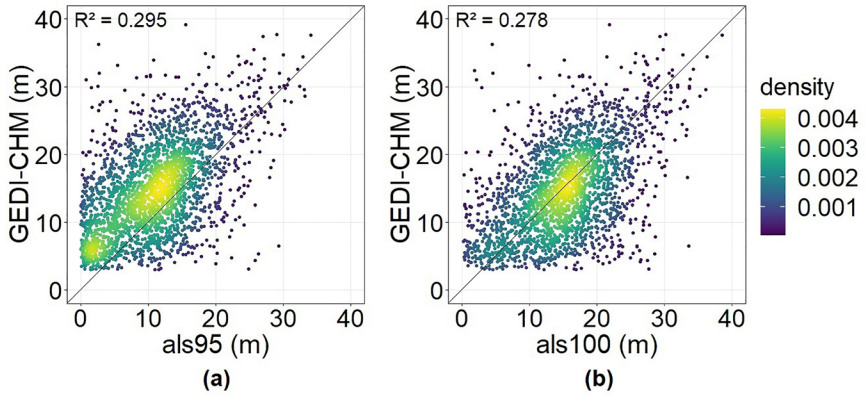


Figure 4. GEDI-CHM estimates (*rh_95*) as a function of *als95* (a) and *als100* (b).

Boxplots in Figure 5 describe the distributions of CHM-Differences (i.e., the difference between GEDI-CHM and ALS-CHM) depending on four sensitivity classes. We note that the overall CHM-Differences as well as the median value increase in absolute value when beam sensitivity decreases. Classes [0.90, 0.95] and [0.95, 0.98] display a similar behavior while GEDI shots with a sensitivity inferior to 0.90 appear to clearly underestimate canopy heights. Beam sensitivity values superior to 0.98 outperform all other classes in terms of accuracy and, contrary to them, do not show a tendency to underestimate canopy heights. These findings are confirmed by the performance metrics given in Table 3. All classes inferior to a sensitivity value of 0.98 tend to underestimate canopy heights, with bias values ranging from −3.2 to −1.8 m. Conversely, for sensitivity values superior to 0.98, *rh_95* produces relatively unbiased estimates and allow reaching the lowest error, with an RMSE value of 5.8 m. Sensitivity is linked to signal penetration and ground detection: a higher sensitivity allows for better detecting the ground and in the end produces better height estimates.

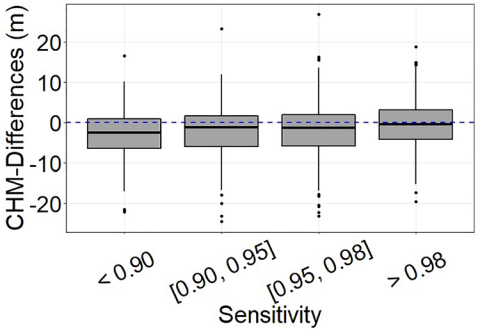


Figure 5. Boxplots of CHM-Differences depending on sensitivity class.

Table 3. Accuracy of GEDI-CHM estimates depending on beam sensitivity, tree height and mean slope.

	Bias (m)	RMSE (m)
Sensitivity		
< 0.90	−3.2	7.3
[0.90, 0.95]	−2.3	6.6
[0.95, 0.98]	−1.8	6.5
> 0.98	−0.5	5.8
Tree height		
< 15 m	0.6	4.7
> 15 m	−4.4	7.6
Slope		
< 15°	1.1	4.8
[15°, 25°]	3.3	6.5
> 25°	5.9	9.4

Similarly, to study the influence of tree height, we also kept GEDI footprints over steep areas (mean slope superior to 15°) out of the study. Figure 6 provides visual representations to understand the relationship between CHM-Differences and ALS-CHM. The main observation that can be drawn is that CHM-Differences increase significantly in absolute value with tree height. Moreover, we also observe two distinct behaviors depending on tree height class. On one hand, for heights ranging between 0 and approximately 15 m, *rh_95* generates unbiased estimates, with a moderate bias of 0.6 m and an RMSE value of 4.7 m (see Table 3). On the other hand, for ALS-CHM values superior to 15 m, estimates through *rh_95* are negatively biased and GEDI strongly underestimates canopy heights, with a bias of −4.4 m and an RMSE of 7.6 m (see Table 3). This underestimation for heights superior to 15 m is still clearly observed when considering only high-sensitivity shots (sensitivity superior to 0.98).

Regarding the impact of slope, and in the light of our previous findings, we excluded GEDI footprints with sensitivity values inferior to 0.98 as well as those with an ALS-CHM superior to 15 m. In that way, we can focus on the impact of slope only, without adding the other effects that were previously

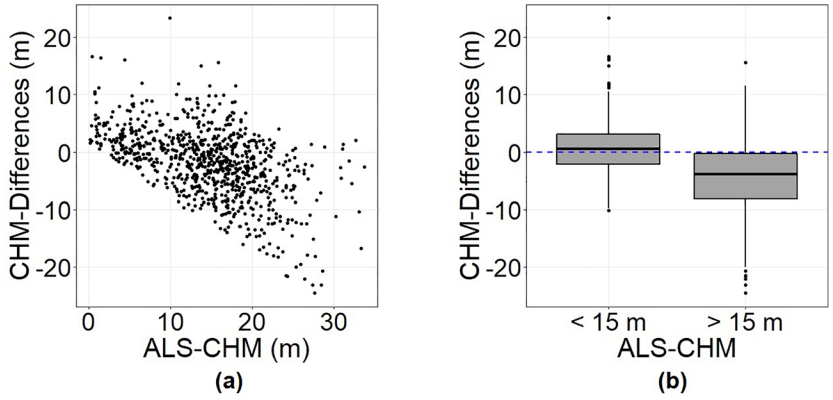


Figure 6. CHM-Differences as a function of ALS-CHM (a) and boxplots of CHM-Differences depending on ALS-CHM class (b).

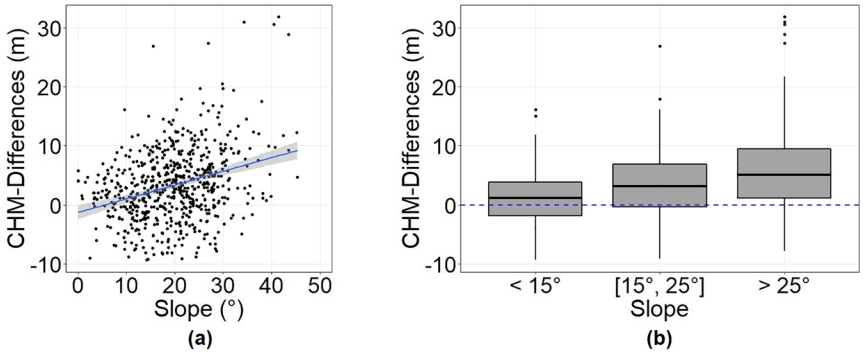


Figure 7. CHM-Differences as a function of slope (a) and boxplots of CHM-Differences depending on slope class (b).

Table 4. Accuracy of GEDI-CHM estimates for *rh_95* and the three regression models.

Estimator	Bias (m)	RMSE (m)	rRMSE
rh_95	−0.3	6.6	44%
MRH	0.0	5.7	38%
RFH	0.0	5.7	38%
sRFH	0.0	5.7	38%

studied. Figure 7a exhibits the relation between CHM-Differences and mean slope. As highlighted by the trendline, CHM-Differences increase with mean slope and steeper slopes result in an overestimation of canopy heights. Indeed, when performing the analysis by slope classes, boxplots in Figure 7b show that the overall CHM-Differences mostly correspond to positive values for slopes superior to 15° (i.e., an overestimation of canopy heights). On the contrary, GEDI footprints acquired over relatively flat terrain (mean slope inferior to 15°) allow reaching unbiased and more accurate height estimates, as confirmed by the accuracy metrics displayed in Table 3.

Regression models

In order to improve the accuracy of canopy height estimates, we implemented several regression models based on GEDI waveform metrics, signal parameters and terrain conditions. In this section, we assess the performances of these models and how the integration of additional data to GEDI height metrics can help reach better estimations.

Multilinear regression for height estimation (MRH)

The MRH model allows linking ALS-CHM with previously selected explanatory variables through a linear relation. When considering all the available input data (i.e., 2397 GEDI footprints), the relation built by the MRH model is as follows:

$$\mathrm{ALS-CHM(m)} = 3.0 + 0.5 * rh_95 - 0.1 * slope_mean + 5.7 * sensitivity \quad (1)$$

MRH produces an unbiased estimation with an RMSE of 5.7 m (see Table 4). These results highlight the fact that using a multilinear regression built on waveform metrics allows for improving the accuracy of the estimation task compared to the direct method when using only *rh_95*. More importantly, when the analysis is carried out according to the signal and terrain parameters that were previously assessed (sensitivity, slope and tree height), we note that the impact of these factors is greatly lessened and that canopy height estimates are no longer dependent on them as well as completely unbiased. When considering CHM-Differences, boxplots in Figure 8a highlight that MRH displays the same behavior for all four sensitivity classes. Similarly, Figure 8b and c show that the results given by MRH do not vary with slope and tree height.

Random Forest for height estimation (RFH and sRFH)

Both RF models also produce unbiased canopy height estimates with an RMSE of 5.7 m (see Table 4). Similarly to MRH, building RF models based on waveform metrics results in a performance gain compared to the use of *rh_95* only. All three estimators assessed in this study display equivalent accuracies in terms of bias and error (see Table 4). Additionally, RF regressors are also capable of lessening and erasing the impact of signal and terrain parameters. In the same spirit as what is observed for MRH, Figures 9a and 10a highlight the fact that canopy height estimates produced through RF models are not dependent on sensitivity. The same conclusion can be drawn

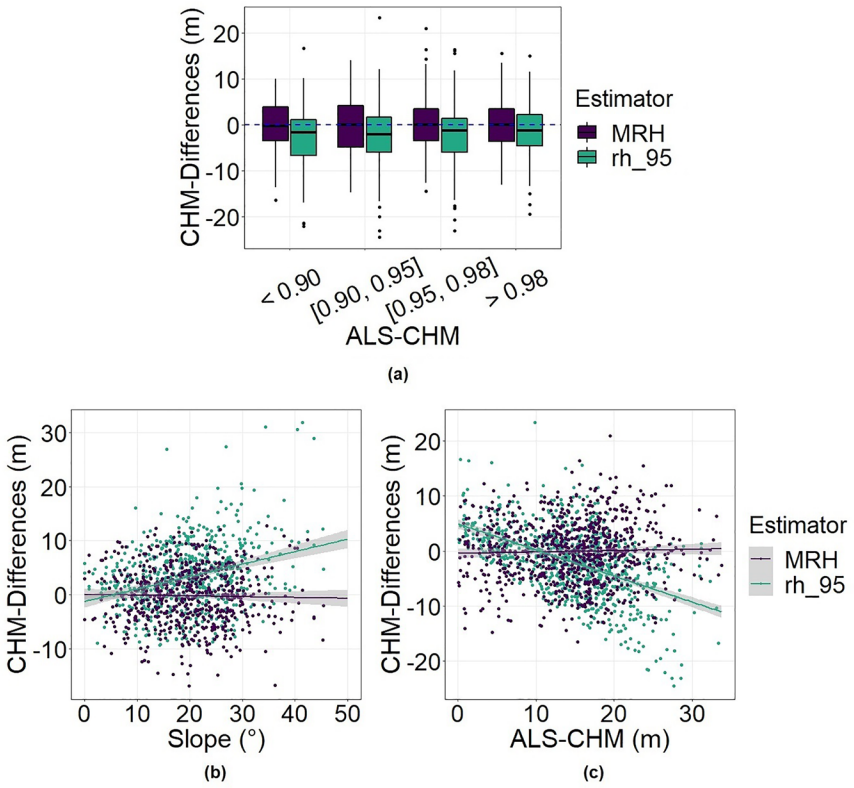


Figure 8. Boxplots of CHM-Differences depending on sensitivity class (a) and CHM-Differences as a function of slope (b) and ALS-CHM (c) for MRH and *rh_95*.

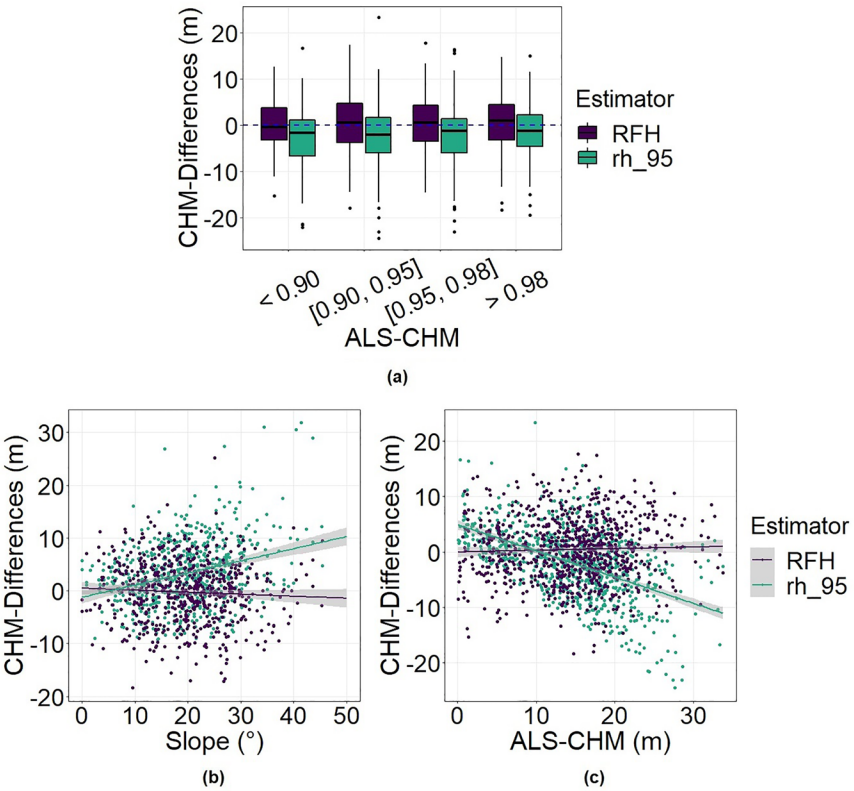


Figure 9. Boxplots of CHM-Differences depending on sensitivity class (a) and CHM-Differences as a function of slope (b) and ALS-CHM (c) for RFH and *rh_95*.

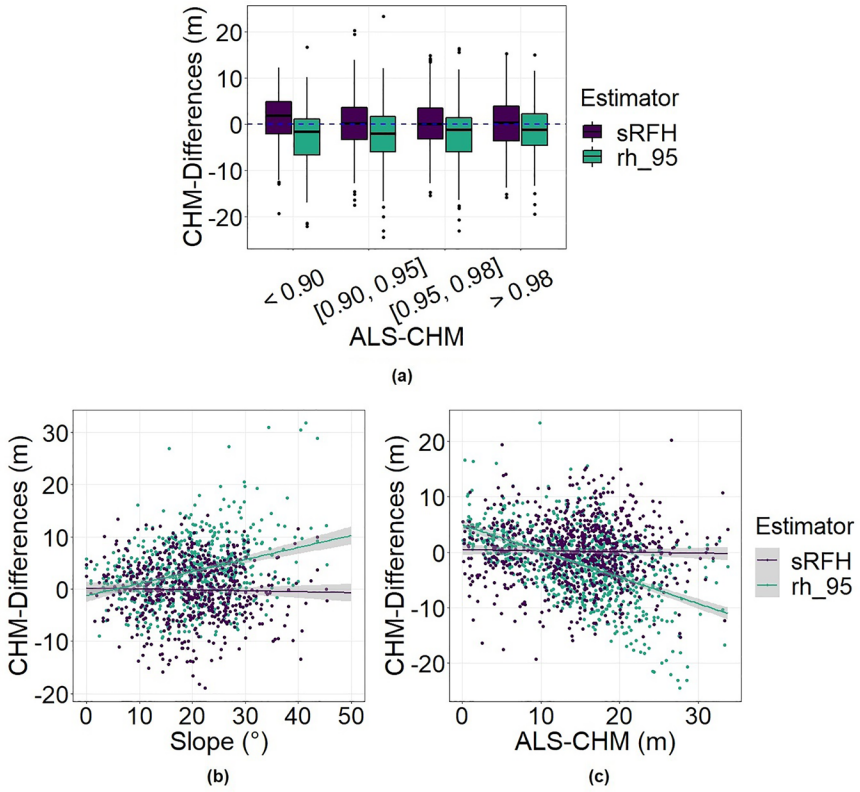


Figure 10. Boxplots of CHM-Differences depending on sensitivity class (a) and CHM-Differences as a function of slope (b) and ALS-CHM (c) for sRFH and *rh_95*.

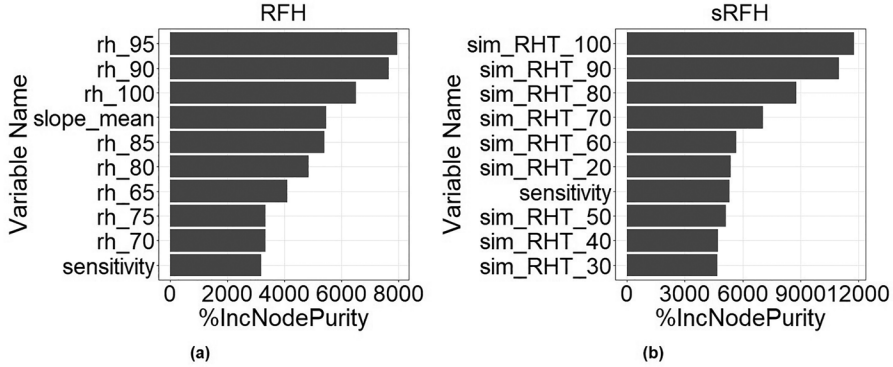


Figure 11. Importance of input variables (ten most important) using mean decrease Gini (%IncNodePurity) for RFH (a) and sRFH (b).

regarding slope and tree height, as exhibited in Figures 9b, c, 10b and c.

The analysis of importance informs about which input variables have the most predictive power (Figure 11). In the case of RFH, when looking at the ten most important predictors, relative height metrics of high percentages appear as the most important variables, with *rh_95* conveying the highest contribution to the estimation task. Slope and sensitivity also bring an interesting added value as input variables when estimating CHM through a RF model. Regarding sRFH, we note that the simulated *RHT* metrics (which contain slope information) all appear as the most contributing predictors whereas sRFH do not seem able to leverage the information given by the simulated *HG* metrics. Similarly, sensitivity emerges as a contributing factor in the estimation task through sRFH.

Overall, when comparing estimates with ALS reference data for the whole dataset, we note that both

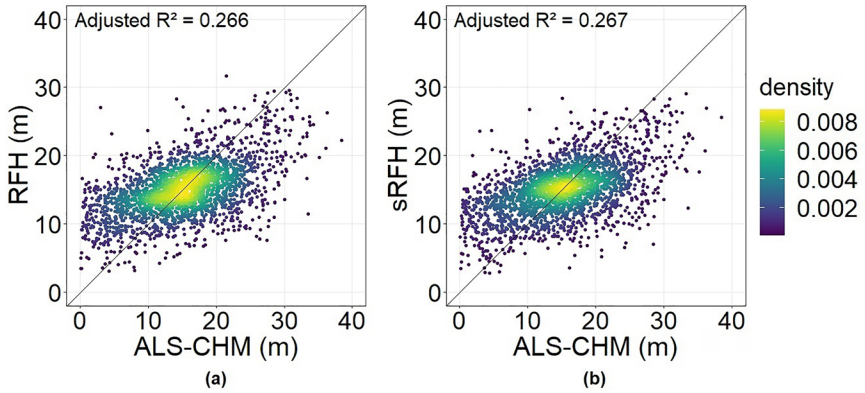


Figure 12. GEDI-CHM estimates from RFH (a) and sRFH (b) as a function of ALS-CHM.

models struggle to predict canopy heights at the edges, with an overestimation of low heights and an underestimation of high heights as shown in Figure 12.

Discussion

In the scope of our study, various methods were evaluated for estimating canopy heights from GEDI data. The results indicate that utilizing waveform metrics-based prediction models yield unbiased and relatively accurate height estimates, with a best RMSE value of around 5.7 m (38%).

The most straightforward approach for estimating canopy heights from GEDI data consists in using a specific L2A level metric as a direct indicator of CHM. The metric selected for this purpose is *rh_95*, which represents the 95th percentile of energy return height relative to the ground. Upper relative height metrics, such as the ones we considered in our analysis, are associated with the top of forest canopies. Previous studies have observed that *rh_95* showed the highest correlation with ALS-derived reference heights in tropical forest biomes, while *rh_100* tended to overestimate canopy height (Potapov et al. 2021). Nonetheless, other works in the literature have also advocated for the use of *rh_98* to derive forest heights, although in different contexts such as Mediterranean forests (Dorado-Roda et al. 2021) and African savannas (Li et al. 2023). Given the forest type of Mayotte Island and in the light of the results we obtained, *rh_95* was finally selected in this study to retrieve GEDI-CHM.

The reference heights used to assess GEDI accuracies and to build regression models were derived from a CHM obtained from ALS data. In this approach, each GEDI footprint is characterized by a distinct reference height, which is compared to the canopy height derived from *rh_95*. In this regard, the maximum CHM value within the footprint's extent proves to have a stronger correlation with GEDI height metrics, as the top location of the waveform (i.e., the location where the highest detected return in the waveform occurs) is directly linked to the first interaction of the laser signal with the highest object within the footprint (Hilbert and Schmullius 2012; Adam et al. 2020). However, considering the uncertainty of GEDI geolocation, extracting relevant reference heights from CHM rasters can be particularly challenging for canopies that exhibit fragmented spatial distributions or possess heterogeneous three-dimensional structures at scales similar to the 25 m GEDI footprint dimension. For example, Roy et al. (2021) observed that secondary forests in the western Democratic Republic of Congo are particularly likely to produce unreliable canopy height retrievals associated with geolocation uncertainty because of their spatially fragmented and heterogeneous three dimensional structure. Specifically in this study, we note that a significant difference between *als100* and *als95* corresponds to GEDI shots over heterogeneous canopies, isolated trees and forest edges, amongst other cases. These shots prove to give less reliable canopy height estimates, with a stronger bias and a larger RMSE. The error increases with heterogeneity because the value extracted as reference data are probably off and do not really correspond to the associated GEDI waveform due to the geolocation uncertainty. The impact of this uncertainty is lessened when dealing with homogeneous forests because the spatial variation of tree heights is smoother. On average, using *als100* (i.e., the maximum height value within the footprint) still remains the better option, even though it may give less reliable results locally. All things considered, as highlighted before in this study, the dispersion that is observed between GEDI-derived heights and ALS

CHM is linked to many factors, including GEDI geolocation accuracy.

Like any remote sensing instrument, GEDI is subject to geolocation errors than can impact the accuracy of its measurements. Geolocation uncertainties in the context of GEDI refer to inaccuracies in determining precisely the spatial location of the laser footprint on the Earth's surface. These errors can arise from various sources, including satellite orbital inaccuracies, instrumental calibration imperfections and atmospheric effects that impact the propagation of the laser beam through the atmosphere (Roy et al. 2021). The quantification and mitigation of geolocation errors are critical for ensuring the reliability of GEDI's data products and various approaches were developed in order to tackle this particular issue (Hancock et al. 2019; Shannon et al. 2022; Schleich et al. 2023; Tang et al. 2023). The GEDI geolocation requirement as provided in the version 2 of the data products is that each 25 m footprint center is horizontally georeferenced to within 10 m, assuming normally distributed geolocation errors with a 0 m mean and a 10 m standard deviation (Dubayah et al. 2021). To address geolocation errors in our dataset and understand how they could impact our results, an investigation was undertaken to quantify the contribution of geolocation uncertainties in the dispersion and the errors observed in our study. The process involved a controlled spatial perturbation of each GEDI footprint center, systematically shifting them within both the X and Y directions across a range from −10 m to 10 m with a 5 m step size. This shifting process resulted in the creation of a comprehensive grid encompassing 25 distinct spatial locations for each footprint. Subsequently, for each possible spatial location within the grid, the associated ALS-CHM value was extracted from the reference canopy height raster. The optimal corrected spatial location and the corresponding reference ALS-CHM value were chosen based on the closest match to *rh_95* relative height metric. As expected, this approach significantly reduced the canopy height estimation errors (Figure 13). More interestingly, we quantified an improvement of RMSE value from 6.6 m for the initial geolocations to 3.8 m for the corrected geolocations, confirming a substantial positive impact in our analysis of geolocation errors on the use of *rh_95* as a direct proxy of canopy height. Moreover, the tendencies observed regarding the influence of key parameters (sensitivity, canopy height and slope) persisted when using corrected spatial locations, albeit with reduced errors (not shown). Additionally, the implementation of regression models demonstrated a notable decrease in errors when using corrected geolocations. The RMSE value, which initially stood at 5.7 m (38%) decreased significantly to 3.2 m (21%) for both MRH and RFH. This new error on height estimation, calculated by attempting to remove the geolocation error of GEDI shots, is similar in order of magnitude to the one observed in another study over Eucalyptus plantations in Brazil (Fayad et al. 2021b), where GEDI geolocation uncertainty is less problematic given the fact these plantations present homogeneous heights on large areas. The enhancements we observed provide confirmation of the impact of geolocation uncertainties on canopy height estimates. They also offer insights into the maximum performance improvements achievable with refined geolocation. However, these enhancements may not be attainable within an operational context where ALS data are not available to correct GEDI spatial locations. Therefore, we retained the uncorrected results as the primary findings of this study.

The ability of a laser signal to penetrate through forest cover and reach the ground is directly linked to the laser's physical properties. The interpretation

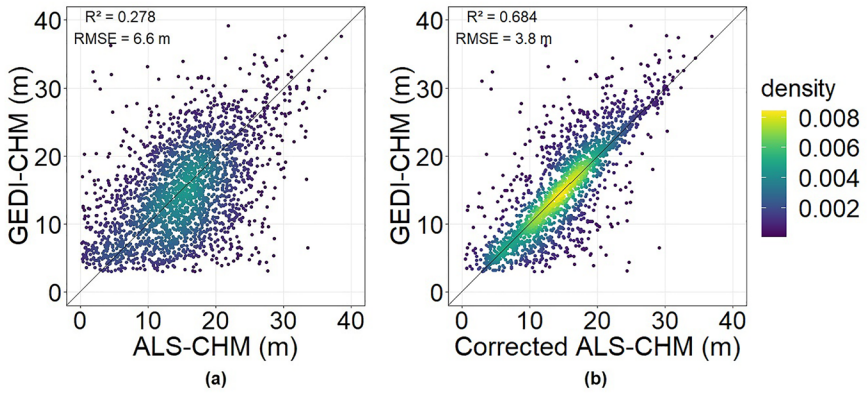


Figure 13. GEDI-CHM estimates as a function of ALS-CHM when using initial geolocations (a) and corrected geolocations (b).

of waveform data and the subsequent computation of metrics are heavily reliant on the quality and shape of the signal. This study highlights the impact of beam sensitivity and shows that it is a crucial parameter for forest height estimation, with lower sensitivities generally resulting in an underestimation of canopy heights while higher sensitivities favor a better penetration through vegetation and thus allow a better detection of the ground. In densely vegetated areas, the ground reflection that is recorded by GEDI sensors can be weak or mixed with the background noise. In that case, footprints characterized by high sensitivities are better suited for ground detection since they can penetrate denser canopies. Fayad et al. (2022) observed that footprints with sensitivities greater than 98% exhibited a deeper penetration of vegetation cover, with an average increase of 5 m in the *rh_100* values compared to footprints with a sensitivity value inferior to 98%. Although a different height metric was considered in this study, the strong correlation between *rh_95* and *rh_100* suggests that the same conclusion can be drawn regardless of the metric used to derive heights. Several other studies have also advocated for the use of only high-sensitivity data when dealing with tall, dense or complex forests. The thresholds vary depending on the forest type and the characteristics of the study site. Rajab Pourrahmati et al. (2023) indicated that removing shots with sensitivity inferior to 0.96 increased the accuracy of canopy height estimates over broadleaf and coniferous forests in Germany. A threshold of 0.95 was proposed by Dhargay et al. (2022) in the context of high-density complex forests located in the Central Highlands region of Victoria in Australia. Rishmawi et al. (2021) also used the same threshold in their study on key forest structure attributes of large-extent forests across the United States.

This study also assessed the impact of slope on the information derived from GEDI waveforms. When estimating canopy heights from a relative height metric such as *rh_95*, increasing slopes lead to an increased waveform extent, which leads to an increase of RH metrics and therefore results in an overestimation of canopy heights (Adam et al. 2020). Indeed, on steep forested terrain, LiDAR returns from the vegetation and the ground can coincide at the same height and, consequently, this effect leads to an overestimation of the vertical structure of the vegetation (Fayad et al. 2021a). The results presented in this study show that for slopes greater than 15°, *rh_95* produces positively biased canopy height estimates with relatively higher errors than when considering slope values inferior to 15°. Some studies proposed a geometric correction of slope effects by applying an offset to GEDI RH metrics (Yang et al. 2011). This offset is computed from the mean slope value as well as the footprint diameter, in order to account for the additional height seen in the increased waveform extents. Figure 14 shows the improvements obtained when applying this simple correction to the *rh_95* metric. We note that the dependency to slope is reduced and that the overall accuracies of canopy height estimates are improved. However, this approach also appears to over-correct slope effects, especially for steep slopes, and shows a tendency to underestimate canopy heights in general. The way slope information is integrated in prediction models is a key element toward reaching better accuracies for canopy height estimation.

In general, the accuracy achieved by using *rh_95* as a direct substitute for GEDI-CHM is insufficient and cannot be deemed satisfactory. Other studies obtained similar accuracies in terms of errors when using a single GEDI metric to derive canopy height. For example, Dorado-Roda et al. (2021) reported rRMSE values

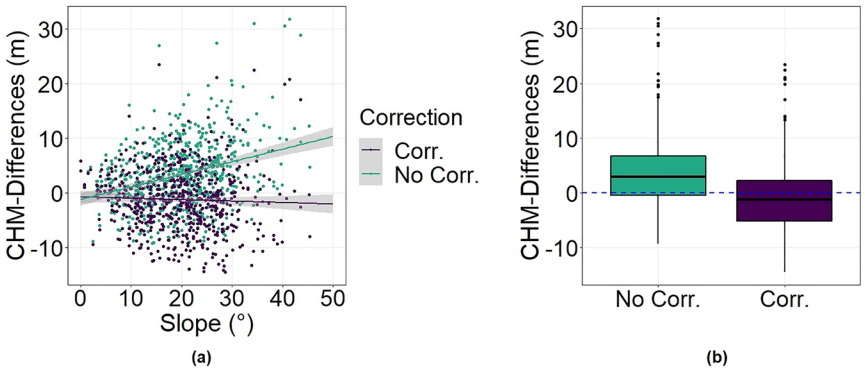


Figure 14. CHM-Differences as a function of slope (a) and boxplots of CHM-Differences (b) depending on whether a simple geometric correction is applied or not.

of up to 41% depending on tree species in Mediterranean forests. In their study over African savannas, Li et al. (2023) presented rRMSE values between 29.8% and 40.9% depending on leaf conditions. In a tropical context, V.C. Oliveira et al. (2023) obtained a percent error of 36% when using *rh_95* to derive canopy height over the Brazilian Amazon forest and, similarly, in an analysis conducted over a study site in the western part of Brazil, Urbazaev et al. (2021) presented an RMSE value of 7 m when estimating canopy height from GEDI data. Nevertheless, rRMSE was found to be much lower in the particular case of Eucalyptus plantations in Brazil (Fayad et al. 2021b), which are characterized by a very large and homogeneous cover in terms of height, an absence of gaps as well as the availability of precise field measurements. In this prospective, statistical approaches that rely on GEDI L2A metrics have been found to lead to significant improvements in CHM estimates (Lahssini et al. 2022). Instead of relying on a single metric, these approaches utilize multiple predictors as inputs to create and train empirical models. By doing so, the richness of the information contained in the waveforms can be leveraged, allowing for taking advantage of the interplay between all the predictors to estimate the target variable. The predictors selected in this study include height metrics that describe the canopy tops, the ground elevation and the vertical structure of the forest, which are known to be correlated with tree heights and are frequently used in canopy height estimation. In addition, signal physical characteristics and terrain configuration were taken into account through the inclusion of beam sensitivity and slope as input variables, since they were found to have a significant impact on direct CHM estimation via *rh_95*.

Implementing regression models demonstrates that waveform metrics combined with other external parameters have the potential to yield improved accuracies for CHM estimates. The RMSE values remain significant (5.7 m, 38%) but are nonetheless consistent with observations made in other studies over tropical ecosystems. In the global forest canopy height map for the year 2019 released by Potapov et al. (2021), RMSE values of 6.6 m and 9.1 m were documented in validations against GEDI validation data and ALS data, respectively. In another study over tropical forest ecosystems in French Guiana and Gabon, Ngo et al. (2023) reported errors of about 5 m on canopy height estimation. Conversely, on the same Eucalyptus dataset (Fayad et al. 2021b), several linear and non-linear approaches based on GEDI metrics produced much lower rRMSE values ranging between 7.8% and 12.4% depending on the model implemented. In our study, all three competitors assessed exhibit similar performances and prove to remain stable regarding all the parameters of influence that were considered. MRH uses pre-selected predictors and produces a linear relationship linking explanatory variables with canopy height. RF is commonly used for the prediction of forest stand attributes when reference in-situ data are available for model training. In the context of this study, both RFH and sRFH show a robust ability to utilize the available inputs to generate accurate CHM estimates. These models differ by the way terrain information is integrated to account for slope effects. Examining the importance reveals which variables possess the most substantial predictive power. Regarding RFH, high-percentile relative heights emerge as important indicators of canopy height (see Figure 11). These RH metrics are linked with the characterization of canopy tops and are logically selected as major factors of CHM variability. Beam sensitivity and slope, which prove to have a direct impact on the interpretability of GEDI waveforms, also appear in the most contributing input variables for RFH. When integrating terrain parameters through simulated ground returns, sRFH is capable of leveraging the slope information to produce similar estimates to RFH in terms of accuracy. Simulated RHT metrics of high percentiles are the most contributing factors in the estimation task and beam sensitivity is still an interesting variable to consider for model training and validation.

When looking at the performances of regression models depending on tree height, the scatter plots in Figure 12 highlight two specific tendencies for canopy height estimates through RF models. The same observations can be made for all the estimators assessed in this study. On one hand, for relatively high heights (i.e., superior to 25 m), GEDI tends to underestimate CHM. This is due to the fact this tree height range usually corresponds to higher AGB levels and denser vegetation covers. As discussed before, the laser penetration through vegetation is of paramount importance to measure the vertical structure of the forest, and it is more challenging for the signal to penetrate denser canopies. In these conditions, the ground peak in the waveform is extracted at a higher height than the actual ground, resulting in an underestimation of canopy height (Potapov et al. 2021; Lang et al. 2022; Lahssini et al. 2022). Conversely, for lower heights (i.e., inferior to 5 m), GEDI exhibits a tendency to overestimate CHM in a quite significant way. This overestimation of low heights is due to the waveform extent and the natural broadening of the ground return, especially over steep terrain. Indeed, if we consider a bare soil with no

vegetation, the resulting echoed waveform will present a single peak with a width of about 3 m (Fayad et al. 2021a), even though the height for this specific footprint should be 0 m. This effect is exacerbated over steep terrain, leading to an even bigger overestimation of heights. Using RH metrics to derive canopy heights for low vegetated areas is therefore more challenging because of the impact of the ground peak width on the waveform extent.

Conclusions

In this study, we explored different methods for estimating canopy height from GEDI data while examining the impact of GEDI acquisition and environmental parameters on the precision of canopy height estimates.

In a heterogeneous forest environment, accurately assessing the precision of GEDI for height estimation poses challenges, primarily due to the geolocation error of GEDI data, even though geolocation uncertainty can be considered low. Thus, in the context of Mayotte's forests, we posit that about half of the error in height estimation is attributable to the geolocation of GEDI shots. Taking all factors into consideration, the accuracy of canopy height estimates from GEDI data are strongly influenced by signal parameters and environmental features. The issue of LiDAR beam penetration is crucial for canopy height estimation, particularly for dense canopies such as the ones that characterize tropical ecosystems. In this study, we illustrate a reduction in laser beam penetration depth starting at approximately 15 m, in contrast to findings from other studies that reported full penetration of the LiDAR waveform up to heights of about 30 m. Hence, LiDAR beam penetration capability is strongly dependent on forest characteristics and penetration depth can differ between forests with the same height and biomass levels. Similar conclusions have been well documented in the literature regarding the penetration of L-band SAR data, with studies reporting different saturation levels of the radar signal for different forest types. Terrain slope also proves to significantly impact the received waveforms that are retrieved by GEDI sensors and needs to be accounted for when dealing with data that was acquired over steep areas.

In general, using *rh_95* as a direct indicator of CHM resulted in relatively low accuracies. However, *rh_95* can still be utilized to estimate canopy heights under certain conditions, particularly when reference canopy height data are unavailable. In that case, it is recommended to only use high-sensitivity data, as it yielded better results compared to using all available footprints. When reference data are available (in our case ALS data), implementing regression models can improve canopy height estimation accuracies, especially when GEDI beam sensitivity and terrain slope are accounted for in the models. The primary requirement for building empirical models is having enough data for model training. GEDI information can be utilized in various ways to estimate canopy heights, depending on data availability, operational application and the expected accuracies.

Acknowledgments

The authors would like to thank the GEDI team and NASA's LP DAAC for providing the GEDI data. This research received funding from the French Space Study Center (CNES, TOSCA 2022 project) and the National Research Institute for Agriculture, Food and the Environment (INRAE).

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Kamel Lahssini http://orcid.org/0000-0002-5359-8718
Nicolas Baghdadi http://orcid.org/0000-0002-9461-4120
Guerric Le Maire http://orcid.org/0000-0002-5227-958X
Stéphane Dupuy http://orcid.org/0000-0002-9710-5364

References

Adam, M., Urbazaev, M., Dubois, C., and Schmullius, C. 2020. "Accuracy assessment of GEDI terrain elevation and canopy height estimates in European temperate forests: Influence of environmental and acquisition parameters." *Remote Sensing*, Vol. 12(No. 23): p. 1. doi:10.3390/rs12233948.

Adrah, E., Wan Mohd Jaafar, W.S., Omar, H., Bajaj, S., Leite, R.V., Mazlan, S.M., Silva, C.A., et al. 2022. "Analyzing canopy height patterns and environmental landscape drivers in tropical forests using NASA's GEDI spaceborne LiDAR." *Remote Sensing*, Vol. 14(No. 13): pp. 3172. doi:10.3390/rs14133172.

Asner, G.P., and Mascaro, J. 2014. "Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric." *Remote Sensing of Environment*, Vol. 140 pp. 614–22. doi:10.1016/j.rse.2013.09.023.

Baghdadi, N.N., El Hajj, M., Bailly, J.-S., and Fabre, F. 2014. "Viability statistics of GLAS/ICESat data acquired over tropical forests." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 7(No. 5): pp. 1658–1664. doi:10.1109/JSTARS.2013.2273563.

Boyd, D.S., and Danson, F.M. 2005. "Satellite remote sensing of forest resources: Three decades of research development."

Progress in Physical Geography: Earth and Environment, Vol. 29(No. 1): pp. 1–26. doi:10.1191/0309133305pp432ra.

Chave, J., Andalo, C., Brown, S., Cairns, M.A., Chambers, J.Q., Eamus, D., Fölster, H., et al. 2005. "Tree allometry and improved estimation of carbon stocks and balance in tropical forests." *Oecologia*, Vol. 145(No. 1): pp. 87–99. doi:10.1007/s00442-005-0100-x.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B.C., Duque, A., et al. 2014. "Improved allometric models to estimate the aboveground biomass of tropical trees." *Global Change Biology*, Vol. 20(No. 10): pp. 3177–3190. doi:10.1111/gcb.12629.

Chen, Q. 2010. "Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry." *Remote Sensing of Environment*, Vol. 114(No. 7): pp. 1610–1627. doi:10.1016/j.rse.2010.02.016.

Dhargay, S., Lyell, C.S., Brown, T.P., Inbar, A., Sheridan, G.J., and Lane, P.N.J. 2022. "Performance of GEDI space-borne LiDAR for quantifying structural variation in the temperate forests of south-eastern Australia." *Remote Sensing*, Vol. 14(No. 15): p. 3615. doi:10.3390/rs14153615.

Dorado-Roda, I., Pascual, A., Godinho, S., Silva, C., Botequim, B., Rodríguez-Gonzálvez, P., González-Ferreiro, E., and Guerra-Hernández, J. 2021. "Assessing the accuracy of GEDI data for canopy height and aboveground biomass estimates in Mediterranean forests." *Remote Sensing*, Vol. 13(No. 12): p. 2279. doi:10.3390/rs13122279.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., et al. 2020. "The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography." *Science of Remote Sensing*, Vol. 1: p. 100002. doi:10.1016/j.srs.2020.100002.

Dubayah, R., Hofton, M., Blair, J., Armston, J., Tang, H., and Luthcke, S. 2021. *GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002*. NASA EOSDIS Land Processes Distributed Active Archive Center. doi:10.5067/GEDI/GEDI02_A.002.

Dubayah, R.O., Sheldon, S.L., Clark, D.B., Hofton, M.A., Blair, J.B., Hurtt, G.C., and Chazdon, R.L. 2010. "Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica." *Journal of Geophysical Research: Biogeosciences*, Vol. 115(No. G2): p. G00E09. doi:10.1029/2009JG000933.

Duncanson, L., Kellner, J.R., Armston, J., Dubayah, R., Minor, D.M., Hancock, S., Healey, S.P., et al. 2022. "Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation (GEDI) lidar mission." *Remote Sensing of Environment*, Vol. 270: p. 112845. doi:10.1016/j.rse.2021.112845.

Dupuy, S., Lainé, G., Tassin, J., and Sarrailh, J.-M. 2013. "Characterization of the horizontal structure of the tropical forest canopy using object-based LiDAR and multispectral image analysis." *International Journal of Applied Earth Observation and Geoinformation*, Vol. 25: pp. 76–86. doi:10.1016/j.jag.2013.04.001.

Fayad, I., Baghdadi, N., Alcarde Alvares, C., Stape, J.L., Bailly, J.S., Scolforo, H.F., Cegatta, I.R., Zribi, M., and Le Maire, G. 2021a. "Terrain slope effect on forest height and wood volume estimation from GEDI data." *Remote Sensing*, Vol. 13(No. 11): p. 2136. doi:10.3390/rs13112136.

Fayad, I., Baghdadi, N., and Lahssini, K. 2022. "An assessment of the GEDI lasers' capabilities in detecting canopy tops and their penetration in a densely vegetated, tropical area." *Remote Sensing*, Vol. 14(No. 13): p. 2969. doi:10.3390/rs14132969.

Fayad, I., Baghdadi, N.N., Alvares, C.A., Stape, J.L., Bailly, J.S., Scolforo, H.F., Zribi, M., and Maire, G.L. 2021b. "Assessment of GEDI's LiDAR data for the estimation of canopy heights and wood volume of eucalyptus plantations in Brazil." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14: pp. 7095–7110. doi:10.1109/JSTARS.2021.3092836.

Feldpausch, T.R., Lloyd, J., Lewis, S.L., Brienen, R.J.W., Gloor, M., Monteagudo Mendoza, A., Lopez-Gonzalez, G., et al. 2012. "Tree height integrated into pantropical forest biomass estimates." *Biogeosciences*, Vol. 9(No. 8): pp. 3381–3403. doi:10.5194/bg-9-3381-2012.

Gargominy, O. 2003. *Biodiversité et conservation dans les collectivités françaises d'outre-mer*. Paris: Comité français pour l'UICN.

Gupta, R., and Sharma, L.K. 2022. "Mixed tropical forests canopy height mapping from spaceborne LiDAR GEDI and multisensor imagery using machine learning models." *Remote Sensing Applications: Society and Environment*, Vol. 27: p. 100817. doi:10.1016/j.rsase.2022.100817.

Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L.I., Kellner, J.R., and Dubayah, R. 2019. "The GEDI simulator: A large-footprint waveform lidar simulator for calibration and validation of spaceborne missions." *Earth and Space Science (Hoboken, N.J.)*, Vol. 6(No. 2): pp. 294–310. doi:10.1029/2018EA000506.

Hilbert, C., and Schmullius, C. 2012. "Influence of surface topography on ICESat/GLAS forest height estimation and waveform shape." *Remote Sensing*, Vol. 4(No. 8): pp. 2210–2235. doi:10.3390/rs4082210.

Ilangakoon, N.T., Glenn, N.F., Dashti, H., Painter, T.H., Mikesell, T.D., Spaete, L.P., Mitchell, J.J., and Shannon, K. 2018. "Constraining plant functional types in a semi-arid ecosystem with waveform lidar." *Remote Sensing of Environment*, Vol. 209: pp. 497–509. doi:10.1016/j.rse.2018.02.070.

Kellner, J.R., Clark, D.B., and Hubbell, S.P. 2009. "Pervasive canopy dynamics produce short-term stability in a tropical rain forest landscape." *Ecology Letters*, Vol. 12(No. 2): pp. 155–164. doi:10.1111/j.1461-0248.2008.01274.x.

Kutchartt, E., Pedron, M., and Pirotti, F. 2022. "Assessment of canopy and ground height accuracy from GEDI LIDAR OVER steep mountain areas." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 3: pp. 431–438. doi:10.5194/isprs-annals-V-3-2022-431-2022.

Lachassagne, P., Aunay, B., Frissant, N., Guilbert, M., and Malard, A. 2014. "High-resolution conceptual hydrogeological model of complex basaltic volcanic islands: A Mayotte, Comoros, case study." *Terra Nova*, Vol. 26(No. 4): pp. 307–321. doi:10.1111/ter.12102.

Lahssini, K., Baghdadi, N., Le Maire, G., and Fayad, I. 2022. "Influence of GEDI acquisition and processing parameters on canopy height estimates over tropical forests." *Remote Sensing*, Vol. 14(No. 24): p. 6264. doi:10.3390/rs14246264.

Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., and Wegner, J.D. 2022. "Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles." *Remote Sensing of Environment*, Vol. 268: p. 112760. doi:10.1016/j.rse.2021.112760.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O., and de Oliveira, R. 2005. "Estimates of forest canopy height and aboveground biomass using ICESat." *Geophysical Research Letters*, Vol. 32(No. 22): p. L22S02. doi:10.1029/2005GL023971.

Li, X., Wessels, K., Armston, J., Hancock, S., Mathieu, R., Main, R., Naidoo, L., Erasmus, B., and Scholes, R. 2023. "First validation of GEDI canopy heights in African savannas." *Remote Sensing of Environment*, Vol. 285: p. 113402. doi:10.1016/j.rse.2022.113402.

Lima, A.J.N., Suwa, R., de Mello Ribeiro, G.H.P., Kajimoto, T., dos Santos, J., da Silva, R.P., de Souza, C.A.S., et al. 2012. "Allometric models for estimating above- and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil." *Forest Ecology and Management*, Vol. 277: pp. 163–172. doi:10.1016/j.foreco.2012.04.028.

Liu, A., Cheng, X., and Chen, Z. 2021. "Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals." *Remote Sensing of Environment*, Vol. 264: p. 112571. doi:10.1016/j.rse.2021.112571.

Ngo, Y.-N., Ho Tong Minh, D., Baghdadi, N., and Fayad, I. 2023. "Tropical forest top height by GEDI: From sparse coverage to continuous data." *Remote Sensing*, Vol. 15(No. 4): pp. 975. doi:10.3390/rs15040975.

Ngo, Y.-N., Huang, Y., Minh, D.H.T., Ferro-Famil, L., Fayad, I., and Baghdadi, N. 2022. "Tropical forest vertical structure characterization: From GEDI to P-band SAR tomography." *IEEE Geoscience and Remote Sensing Letters*, Vol. 19: pp. 1–5. doi:10.1109/LGRS.2022.3208744.

Oliveira, P.V., Zhang, X., Peterson, B., and Ometto, J.P. 2023. "Using simulated GEDI waveforms to evaluate the effects of beam sensitivity and terrain slope on GEDI L2A relative height metrics over the Brazilian Amazon Forest." *Science of Remote Sensing*, Vol. 7:p. 100083. doi:10.1016/j.srs.2023.100083.

Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., et al. 2011. "A large and persistent carbon sink in the world's forests." *Science (New York, N.Y.)*, Vol. 333(No. 6045): pp. 988–993. doi:10.1126/science.1201609.

Pascal, O., and Labat, J.-N. 2002. *Plantes et forêts de Mayotte*. Paris: Muséum national d'histoire naturelle, Institut d'écologie et de gestion de la biodiversité, Service du patrimoine naturel.

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Pickens, A., et al. 2021. "Mapping global forest canopy height through integration of GEDI and Landsat data." *Remote Sensing of Environment*, Vol. 253: p. 112165. doi:10.1016/j.rse.2020.112165.

Rajab Pourrahmati, M., Baghdadi, N., and Fayad, I. 2023. "Comparison of GEDI LiDAR Data Capability for Forest Canopy Height Estimation over Broadleaf and Needleleaf Forests." *Remote Sensing*, Vol. 15(No. 6): p. 1522. doi:10.3390/rs15061522.

Requena Suarez, D., Rozendaal, D.M.A., De Sy, V., Phillips, O.L., Alvarez-Dávila, E., Anderson-Teixeira, K., Araujo-Murakami, A., et al. 2019. "Estimating aboveground net biomass change for tropical and subtropical forests: Refinement of IPCC default rates using forest plot data." *Global Change Biology*, Vol. 25(No. 11): pp. 3609–3624. doi:10.1111/gcb.14767.

Rishmawi, K., Huang, C., and Zhan, X. 2021. "Monitoring Key Forest Structure Attributes across the Conterminous United States by Integrating GEDI LiDAR Measurements and VIIRS Data." *Remote Sensing*, Vol. 13(No. 3): p. 442. doi:10.3390/rs13030442.

Roy, D.P., Kashongwe, H.B., and Armston, J. 2021. "The impact of geolocation uncertainty on GEDI tropical forest canopy height estimation and change monitoring." *Science of Remote Sensing*, Vol. 4: p. 100024. doi:10.1016/j.srs.2021.100024.

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D.M.A., Avitabile, V., Araza, A., De Bruin, S., et al. 2021. "The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations." *Earth System Science Data*, Vol. 13(No. 8): pp. 3927–3950. doi:10.5194/essd-13-3927-2021.

Schleich, A., Durrieu, S., Soma, M., and Vega, C. 2023. "Improving GEDI footprint geolocation using a high-resolution digital elevation model." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16: pp. 7718–7732. doi:10.1109/JSTARS.2023.3298991.

Schlund, M., Wenzel, A., Camarretta, N., Stiegler, C., and Erasmi, S. 2022. "Vegetation canopy height estimation in dynamic tropical landscapes with TanDEM-X supported by GEDI data." *Methods in Ecology and Evolution*, Vol. 14 (No. 7): pp. 1639–1656. doi:10.1111/2041-210X.13933.

Shannon, E.S., Finley, A.O., Hayes, D.J., Noralez, S.N., Weiskittel, A.R., Cook, B.D., and Babcock, C. 2024. "Quantifying and correcting geolocation error in spaceborne LiDAR forest canopy observations using high spatial accuracy ALS: A Bayesian model approach." *Environmetrics:* p. e2840. https://doi.org/10.1002/env.2840.

Shendryk, Y. 2022. "Fusing GEDI with earth observation data for large area aboveground biomass mapping." *International Journal of Applied Earth Observation and Geoinformation*, Vol. 115: p. 103108. doi:10.1016/j.jag.2022.103108.

Tang, H., Stoker, J., Luthcke, S., Armston, J., Lee, K., Blair, B., and Hofton, M. 2023. "Evaluating and mitigating the impact of systematic geolocation error on canopy height measurement performance of GEDI." *Remote Sensing of Environment*, Vol. 291: p. 113571. doi:10.1016/j.rse.2023.113571.

Urbazaev, M., Hess, L., Sato, L., Ometto, J., Thiel, C., Dubois, C., Adam, M., and Schmullius, C. 2021. "Accuracy assessment of terrain and canopy height estimates from ICESat-2 and GEDI LiDAR missions in temperate and tropical forests: First results." In *Proceedings of the* Silvilaser. TU Wien. doi:10.34726/WIM.1984.

Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego, J., Aragão, L.E.O.C., and Nasi, R. 2021. "Long-term (1990–2019) monitoring of forest cover changes in the humid tropics." *Science*

Advances, Vol. 7(No. 10): p. eabe1603. doi:10.1126/sciadv.abe1603.

Wang, C., Elmore, A.J., Numata, I., Cochrane, M.A., Shaogang, L., Huang, J., Zhao, Y., and Li, Y. 2022. "Factors affecting relative height and ground elevation estimations of GEDI among forest types across the conterminous USA." *GIScience & Remote Sensing*, Vol. 59(No. 1): pp. 975–999. doi:10.1080/15481603.2022.2085354.

Wang, Y., Ni, W., Sun, G., Chi, H., Zhang, Z., and Guo, Z. 2019. "Slope-adaptive waveform metrics of large footprint lidar for estimation of forest aboveground biomass." *Remote Sensing of Environment*, Vol. 224: pp. 386–400. doi:10.1016/j.rse.2019.02.017.

Wehr, A., and Lohr, U. 1999. "Airborne laser scanning—An introduction and overview." *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54(No. 2–3): pp. 68–82. doi:10.1016/S0924-2716(99)00011-8.

Xing, Y., De Gier, A., Zhang, J., and Wang, L. 2010. "An improved method for estimating forest canopy height using ICESat-GLAS full waveform data over sloping terrain: A case study in Changbai mountains, China." *International Journal of Applied Earth Observation and Geoinformation*, Vol. 12(No. 5): pp. 385–392. doi:10.1016/j.jag.2010.04.010.

Yang, W., Ni-Meister, W., and Lee, S. 2011. "Assessment of the impacts of surface topography, off-nadir pointing and vegetation structure on vegetation lidar waveforms using an extended geometric optical and radiative transfer model." *Remote Sensing of Environment*, Vol. 115(No. 11): pp. 2810–2822. doi:10.1016/j.rse.2010.02.021.

Chapter 4

Canopy Height Mapping in French Guiana Using Multi-Source Satellite Data and Environmental Information in a U-Net Architecture

Building upon the conclusions of the previous analyses that focused on the performance assessment of GEDI in tropical environments and the various influences on canopy height measurements, this chapter presents the third study of the dissertation to address the third research question: how can effective data fusion models be developed to leverage multiple data sources for canopy height mapping? A high-resolution canopy height map at the regional scale of French Guiana is presented and discussed. This map was generated through the optimal integration of GEDI data with other remote sensing sources and environmental descriptors.

After exhaustive evaluation of GEDI's capabilities and limitations in tropical forest ecosystems, the operational goal of this doctoral thesis shifts towards producing a continuous canopy height map of French Guiana at a spatial resolution of 10 m. This study presents a deep learning approach that leverages the complementarity between GEDI's sparse measurements and other freely accessible remote sensing data (optical and radar), as well as environmental variables related to the canopy structure and the ecological context of French Guiana (height above nearest drainage and forest landscape types). The integration of these diverse data sources is achieved through a U-Net architecture, which is particularly well-suited for handling complex and heterogeneous spatial information. The model, named *CHNET*, is trained using a GEDI height metric as reference canopy height and the insights gained from the previous analyses enable an optimal utilization of these data and guide an iterative refinement process in the form of successive training scenarios to enhance the accuracy of the final canopy height predictions.

In summary, this chapter demonstrates how GEDI inaccuracies can be mitigated in an operational application and highlights the contributions of relevant environmental descriptors to characterize canopy height. The canopy height map produced can serve as a valuable input for estimating AGB in French Guiana and thus assessing the region's carbon stocks. Additionally, it can be useful for validating the upcoming products from ESA's BIOMASS mission. In the end, this study provides an operational and reproductible framework for canopy height mapping based on freely available data that can be adapted and used in different tropical regions.

frontiers | Frontiers in Remote Sensing

TYPE Original Research
PUBLISHED 25 November 2024
DOI 10.3389/frsen.2024.1484900

Check for updates

OPEN ACCESS

EDITED BY
Celso H. L. Silva-Junior,
Instituto de Pesquisa Ambiental da Amazonia (IPAM), Brazil

REVIEWED BY
Patrick Jantz,
Northern Arizona University, United States
Sugandha Arora,
University of Münster, Germany

*CORRESPONDENCE
Kamel Lahssini,
kamel.lahssini@inrae.fr

RECEIVED 22 August 2024
ACCEPTED 31 October 2024
PUBLISHED 25 November 2024

CITATION
Lahssini K, Baghdadi N, le Maire G, Fayad I and Villard L (2024) Canopy height mapping in French Guiana using multi-source satellite data and environmental information in a U-Net architecture.
Front. Remote Sens. 5:1484900.
doi: 10.3389/frsen.2024.1484900

COPYRIGHT
© 2024 Lahssini, Baghdadi, le Maire, Fayad and Villard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Canopy height mapping in French Guiana using multi-source satellite data and environmental information in a U-Net architecture

Kamel Lahssini[1]*, Nicolas Baghdadi[1], Guerric le Maire[2,3], Ibrahim Fayad[4,5] and Ludovic Villard[6]

[1]TETIS, INRAE, Montpellier, France, [2]Eco&Sols, CIRAD, Montpellier, France, [3]Eco&Sols, CIRAD, INRAE, IRD, Institut Agro, Université de Montpellier, Montpellier, France, [4]Kayrros SAS, Paris, France, [5]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS9 UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France, [6]CESBIO, Université Toulouse III - Paul Sabatier, Toulouse, France

Canopy height is a key indicator of tropical forest structure. In this study, we present a deep learning application to map canopy height in French Guiana using freely available multi-source satellite data (optical and radar) and complementary environmental information. The potential of a U-Net architecture trained on sparse and unevenly distributed GEDI data to generate a continuous canopy height map at a regional scale was assessed. The developed model, named *CHNET*, successfully produced a canopy height map of French Guiana at a 10-m spatial resolution, achieving relatively good accuracy compared to a validation airborne LiDAR scanning (ALS) dataset. The study demonstrates that relevant environmental descriptors, namely, height above nearest drainage (HAND) and forest landscape types (FLT), significantly contribute to the model's accuracy, highlighting that these descriptors bring important information on canopy structural properties and that the *CHNET* framework can efficiently use this information to improve canopy height prediction. Another critical aspect highlighted is the necessity of addressing GEDI data inaccuracies and geolocation uncertainties, which is essential for any GEDI-based application. However, challenges remain, particularly in characterizing tall canopies, as our *CHNET* model exhibits a tendency to underestimate canopy heights greater than 35 m. A large part of this error arises from the use of GEDI measurements as reference, given the fact these data exhibit certain saturation in tropical biomes. Future improvements in the analysis of GEDI signal as well as the implementation of robust models are essential for better characterization of dense and tall tropical forest ecosystems.

KEYWORDS

canopy height, data fusion, deep learning, French Guiana, GEDI, lidar, tropical forests

1 Introduction

Tropical forests play a major role in maintaining global ecological balance. Beyond the richness of their biodiversity, they are essential for carbon sequestration, absorbing significant amounts of carbon dioxide and thereby mitigating the effects of climate change (Pan et al., 2011). Mapping forest resources and carbon is crucial for enhancing

forest management practices and achieving goals related to environmental preservation. The accurate characterization and monitoring of the standing aboveground biomass (AGB) have become imperative to understand the carbon sequestration potential of tropical ecosystems. In this regard, several studies have highlighted relationships between the structural characteristics of a forest and its AGB levels (Chave et al., 2005; Asner and Mascaro, 2014). In tropical regions like French Guiana, the dense and complex forest structure makes the precise estimation of AGB particularly challenging (Saatchi et al., 2011). Accurately describing forest structure, including canopy height, is essential for reliable biomass estimates at a regional scale (Joetzjer et al., 2017). Canopy height is indeed one of the most important inputs to estimate biomass, and most allometric equations incorporate this structural parameter to calculate AGB (Lefsky et al., 2005; Lima et al., 2012; Feldpausch et al., 2012).

Remote sensing is a powerful tool for forest monitoring because it offers the ability to collect extensive data across large and often inaccessible areas (Boyd and Danson, 2005). Light detection and ranging (LiDAR), in particular, is highly effective for characterizing vegetation profiles and structural parameters. Among LiDAR technologies, airborne LiDAR scanning (ALS) stands out as the most accurate and offers the highest resolution data (in terms of returned points over a given surface). However, it is also more costly for data users and its spatial coverage is limited to specific areas. In French Guiana, where accessibility is limited, ALS data have been widely used for localized studies (Ho Tong Ming et al., 2016; Joetzjer et al., 2017; Lahssini et al., 2022), but the high cost of this method limits its applicability for large-scale applications. In contrast, spaceborne sensors, while having a lower spatial resolution, offer better revisit times and global coverage, making them invaluable for large-scale and continuous monitoring. These sensors have demonstrated their capability to produce regional or global forest products (Tyukavina et al., 2015) and effectively map canopy height (Fayad et al., 2014; Pourrahmati et al., 2015). The Global Ecosystem Dynamics Investigation (GEDI) mission is the most recent advancement in spaceborne LiDAR technology, and it was specifically designed to measure the structure of the vegetation at a global scale (Dubayah et al., 2020). GEDI has proven its ability to map canopy height accurately and can be complemented with other remote sensing data sources to provide comprehensive information on forest structure and dynamics (Potapov et al., 2021; Lang et al., 2023). Continuous data sources are indeed useful to transform the sparse coverage of GEDI measurements into continuous canopy height maps.

One complementary data source often used with GEDI information is optical data, and studies have already used this complementarity in the context of French Guiana (Ngo et al., 2023). Optical sensors are extensively used to examine the state parameters of forest stands and surface vegetation. Using radiometric information obtained from optical systems, it is possible to characterize stand compositions and vegetation biophysical parameters by analyzing the spectral information contained in the reflectance images (Lemaire et al., 2008; Grabska et al., 2019). One of the primary advantages of most recent satellite optical sensors lies in their capacity to provide complete coverage and high-resolution (both spatial and temporal) multispectral imagery that offers accurate information on forest structure and composition. This information can be used in forest type classification, species identification, and monitoring of canopy cover (Hansen et al., 2013; Vancutsem et al., 2021). Many optical data products can be acquired freely by data users and are available from several satellite platforms, enabling long-term and large-scale assessments of tropical forests. Nonetheless, there are also limitations to consider. First and foremost, cloud cover can impede data acquisition, especially in tropical regions with unique environmental conditions and frequent cloud cover. Furthermore, optical sensors are passive systems that are sensitive to solar illumination conditions, which can limit data acquisition to specific times of the day. They are also sensitive to sun-sensor angle view effects, which are difficult to correct. Additionally, optical sensors are limited in their ability to penetrate canopy cover, thus impeding the study of understory vegetation and ground-level features. A combination of optical data with other remote sensing techniques can therefore offer a more comprehensive view of forest ecosystems.

In this perspective, synthetic aperture radar (SAR) technologies bring specific information on canopy structure and have been effectively employed alongside GEDI data in French Guiana (Ngo et al., 2023; Qi et al., 2023). SAR sensors use lower frequency electromagnetic waves (microwaves) than LiDAR but still offer advantages for studying forest ecosystems. One of their primary advantages is their ability to penetrate cloud cover and to operate in darkness, allowing for data acquisition in all weather conditions and at any moment of the day. Due to their longer wavelengths, radar signals can also penetrate the forest canopy to a certain extent (depending on the wavelength, i.e., deeper penetration with longer wavelengths) and provide information on forest structure and the lower layers of the canopy. Moreover, SAR data are highly sensitive to changes in canopy structure, which makes it suitable for detecting changes such as disturbances, deforestation, and degradation (Rahman and Sumantyo, 2010; Bouvet et al., 2018; Watanabe et al., 2018). Using different frequency bands can be advantageous for studying vegetation because they complement each other (Berninger et al., 2018; Sothe et al., 2022). On one hand, the L-band, with its longer wavelength can penetrate deeper into the vegetation and provide information on the lower layers of the canopy (Baghdadi et al., 2015). This information is particularly relevant for estimating forest structure variables (Mermoz et al., 2014). On the other hand, the C-band, with its shorter wavelength, is limited in its penetration capabilities but can still provide detailed information about the leaf density of upper canopies (Fagua et al., 2019), which is indirectly related to canopy height (McDowell et al., 2002). While LiDAR remains the most appropriate technology for deriving accurate structural descriptors of forest ecosystems, optical and radar data can provide complementary information and multimodal approaches allow combining the advantages of each individual sensor type. Indeed, spaceborne LiDAR sparse measurements do not exhibit a sufficient spatial density for accurate and continuous mapping of canopy height. Therefore, the complementarity between optical, radar, and LiDAR data is leveraged to produce canopy height maps.

Such multimodal approaches have been employed successfully over French Guiana to overcome the limitations of individual sensors in canopy height mapping. However, most existing studies have either focused on very localized areas, such as

research plots (Ngo et al., 2023), or have been conducted at global scales (Potapov et al., 2021; Lang et al., 2023; Qi et al., 2023). In contrast, our study addresses the gap in regional-scale applications, which allows us to explore how data fusion approaches behave and can be optimized at an intermediate yet countrywide scale. Employing an appropriate data fusion framework is essential for the optimal processing of multiple data sources, with the goal of producing a final canopy height map of French Guiana. Standard approaches, such as linear models, regression analysis, and Random Forest provide widely accepted methods for combining data (Morin et al., 2019). As the richness and complexity of multimodal information from various sensors increase, deep learning (LeCun et al., 2015) approaches prove to be particularly useful and effective. Deep learning methods have become prevalent in remote sensing applications due to their powerful ability to analyze and process data from different sources, times, and scales (Hong et al., 2021). By stacking nonlinear functions, neural networks enable the modeling of complex relationships between input and output variables. In multimodal applications, two main strategies are employed to combine heterogeneous and complementary information: early fusion and late fusion (Hong et al., 2021). Early fusion combines data from different sources at the initial stage and then processes them through a neural network. In contrast, late fusion involves analyzing each data source separately using dedicated branches and then combining the intermediate results through additional neural network layers for final processing. Since the introduction of Convolutional Neural Networks (CNNs), there has been a substantial improvement in the accuracy of image analysis tasks. The U-Net architecture (Ronneberger et al., 2015), a specialized type of CNN widely used in image segmentation and remote sensing studies (Yuan et al., 2021), demonstrates an effective neural network model for these applications. U-Net's design, which includes a contracting path to capture local context and a symmetric expanding path for precise localization, is particularly well-suited for handling spatial data. With the goal of predicting and mapping canopy height at a regional scale, this architecture is quite adequate because of its ability to integrate detailed spatial information from multiple remote sensing data. Consequently, U-Net has been extensively used in numerous studies for canopy height mapping using diverse remote sensing data sources (Gazzea et al., 2023; Wagner et al., 2024; Schwartz et al., 2024; Ge et al., 2022). However, most existing U-Net frameworks have exclusively utilized sensor data, and there is still a need to explore how ancillary information related to canopy structure and ecological context can be leveraged by this type of architecture in a regional-scale application.

In general, canopy height mapping approaches often rely exclusively on remote sensing data. However, considering environmental parameters alongside these data sources can provide significant added value. For instance, in some studies, integrating topographical data into the modeling process has led to improvements in the accuracy and reliability of forest biophysical variables estimation (Lahssini et al., 2022). Environmental parameters have indeed a direct impact on canopy dynamics. Identifying the most relevant descriptors to integrate into a deep learning architecture may bring an additional contribution, enabling the architecture to extract complex relationships between these descriptors and canopy height. In this study, we propose an end-to-end early fusion deep learning model that effectively leverages the complementarity between optical (Sentinel-2), radar (Sentinel-1 and ALOS), and environmental data to produce a 10-m resolution canopy height map of French Guiana (South America). The height above nearest drainage (HAND) and the forest landscape types (FLT) were retained as environmental descriptors that are directly related to canopy height. The proposed framework, named *CHNET* (Canopy Height estimation NETwork), implements an early fusion strategy through a U-Net architecture. GEDI measurements are used as reference canopy heights for the *CHNET* model calibration. The main objective of our study is to assess how tropical canopy height can be accurately estimated and mapped from GEDI information and multiple data sources. Our study operates at an intermediate regional scale by accounting for the specificities of the French Guiana context while still addressing the larger scale of the entire country. We aim at understanding how the enrichment and optimization of model inputs, such as integrating additional environmental variables and refining the GEDI data used for model calibration, can potentially lead to improved and more accurate results.

2 Materials and methods

2.1 Study area

French Guiana, a French Overseas Territory, is located in the Amazon forest biome on the northern coast of South America. It spans a total area of 83,534 km^2, of which over 80,000 km^2 are covered by forests (Fayad et al., 2016). The predominant forest type is mature old-growth tropical rainforest, with some regions consisting of secondary forests (Guitet et al., 2013). Coastal areas also feature savannas and mangroves, while rainforest accounts for more than 90% of the territory. Timber harvesting and agricultural activities are mainly confined to the sub-coastal areas near large towns and along the main roads (Guitet et al., 2015). The region's topography is mostly flat, with elevations rarely exceeding 200 m, though small hills and mountains occasionally rise across the landscape (Guitet et al., 2013). Notably, about 70% of the terrain slopes are less than 5° (Fayad et al., 2016). French Guiana experiences a tropical and hot climate, with a Köppen climate classification of tropical rainforest (Af) across most of the territory (Beck et al., 2018). The mean annual temperature is approximately 26°C and rainfall varies significantly across the territory, with the northeast receiving up to 4,000 mm per year, while the southern and western areas receive around 2,000 mm annually (Guitet et al., 2015). The combination of climatic, geological, and geomorphological factors provides the conditions for the development of a wide range of forest structures at a regional scale. Canopy height in French Guiana's tropical forests generally ranges between 20 and 40 m, with emergent trees occasionally reaching up to 60 m (Fayad et al., 2014). AGB levels vary significantly and generally range from around 150 Mg/ha to over 600 Mg/ha, with higher levels typically found in mature and undisturbed forest areas (Fayad et al., 2016).

2.2 Datasets and processing

2.2.1 GEDI data

GEDI, operated by NASA aboard the International Space Station (ISS) since December 2018, is a full-waveform LiDAR

sensor designed to capture high-resolution vertical structure data of Earth's forests. The system employs three 1,064 nm lasers emitting 242 pulses per second, generating energy return waveforms (L1B product) and derived height metrics (L2A product) within 25 m diameter circular footprints. One laser is split into two half-power beams (coverage beams), while the other two lasers remain at full power (power beams). These beams are then slightly dithered to generate eight parallel tracks of observations: four tracks for coverage beams and four tracks for power beams. These tracks cover a 4.2-km swath on the ground, with a 60-m spacing along the track and a 600-m spacing across tracks.

GEDI data are processed by NASA's Land Processes Distributed Active Archive Center (LP DAAC). The L1B product includes geolocated and smoothed waveforms, along with ancillary parameters, and offers a horizontal location precision improved to 10 m (Dubayah et al., 2021). The L2A product delivers elevation and height metrics for each footprint, including measurements of ground elevation, canopy top height and relative height (RH) metrics. The L2A product is generated from the L1B waveforms using six distinct signal processing configurations, referred to as algorithm setting groups. These configurations define the thresholds and smoothing parameters used to analyze the received waveforms, thereby influencing the height metrics extracted in the L2A product. Comparing all available configurations indicates that algorithm setting group number five generally yields better accuracy on average compared to other groups in the tropical context of French Guiana (Lahssini et al., 2022). GEDI's ability to estimate canopy height is contingent on its ability to detect the ground (Sun et al., 2022). Due to the dense vegetation in tropical areas, the ground peak in GEDI waveforms is often weaker in intensity and difficult to distinguish from the background noise. Algorithm setting group number five is characterized by the lowest waveform signal end threshold. In a dense environment, a low signal end threshold enhances the differentiation of weak ground returns from noise in the signal, while a higher threshold tends to lead to errors in the detection of the ground peak in the waveform. Conversely, in less dense forest ecosystems, a low threshold can result in the interpretation of noise as the ground peak, thus leading to an overestimation of canopy heights (Adam et al., 2020). GEDI waveforms obtained over tropical forests often present weak intensity ground peaks, suggesting that a lower signal end threshold is more suitable for these ecosystems. Considering the dense and homogeneous context of French Guiana, as well as the results of related works in the literature (Lahssini et al., 2022), we therefore utilized the L2A metrics that were computed using algorithm setting group number 5.

GEDI RH metrics are considered as a reliable representation of canopy heights and have been used as reference variables to calibrate height estimation models (Lang et al., 2023). The L2A product contains several RH metrics representing the height relative to the ground corresponding to a given percentile of total return energy. For instance, *rh_95* represents the height relative to the ground which contains 95% of the waveform energy. High-percentile RH metrics are therefore good proxies of canopy height as they are related to the top-of-canopy detection. Theoretically, *rh_100* would be expected to be the best representation of canopy height and should correspond to the distance between the top-of-canopy and the ground. However, this metric can be impacted by atmospheric

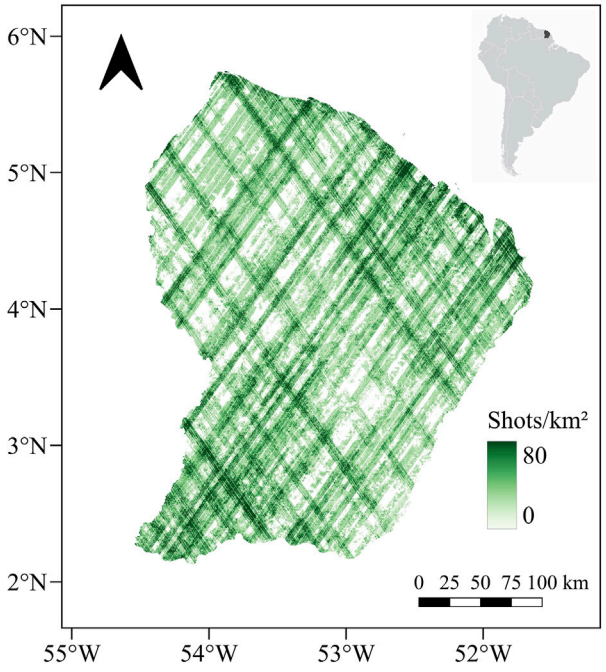


FIGURE 1
Density of valid GEDI data (number of shots per km²) over French Guiana (grid of 500 m).

conditions, incorrect ground peak detection in the waveform as well as variability in both vegetation and ground conditions (Schwartz et al., 2024). In this study, we chose to use *rh_95* as a direct proxy of canopy height since it proved to yield the best performances in the tropical context of French Guiana (Lahssini et al., 2022). Several previous studies have also advocated for the use of *rh_95* to model canopy height in tropical biomes (Potapov et al., 2021; Fayad et al., 2021; Lahssini et al., 2024).

Overall, 11,798,179 GEDI shots acquired between April 2019 and May 2022 over French Guiana were collected from the GEDI Level 2A Geolocated Elevation and Height Metrics product (GEDI02_A v002) made available by NASA (Dubayah et al., 2021). Atmospheric disturbances can strongly impact GEDI measurements, which is why the data meeting the following criteria were considered invalid and were removed: (1) shots with no mode detected (*num_detectedmodes* = 0), as they correspond to pure noise; (2) shots with a null signal-to-noise ratio (SNR), as they are also pure noise; (3) shots displaying an erroneous ground elevation, i.e., when the absolute difference between Shuttle Radar Topography Mission (SRTM) ground elevation (*digital_elevation_model_srtm*) and GEDI ground detection (*elev_lowestmode*) is greater than 100 m; (4) shots exhibiting an incomplete waveform, i.e., partial signals where the useable part of the waveform (*search_end*) is equal to the total number of bins in the waveform (*rex_sample_count*); (5) shots with *rh_100* values lower than 3 m, as GEDI is well suited for vegetation but is unable to measure the height of small objects because of laser pulse width (Dubayah et al., 2020). Following these procedures, a total of 3,891,348 valid footprints (33%) were extracted. Furthermore, in order to provide our framework with useable and interpretable metrics, only GEDI shots with a SNR greater than 10 dB

(2,127,076 shots, 18%) were retained as reference data for building our canopy height estimation models (Figure 1).

The corresponding GEDI *rh_95* values were rasterized on a 10 m grid aligned with the other remote sensing data sources. For each GEDI shot, the associated *rh_95* value was assigned to the pixel corresponding to the center of the footprint. This approach has been applied in several other studies to produce continuous canopy height maps. Lang et al., 2023 used a similar rasterizetion method in their global canopy height map to match GEDI data with the Sentinel-2 grid. They argued that this allows the model to optimize its loss function with respect to valid reference pixels during training. Potapov et al. (2021) employed the same approach in a study using Landsat data at a different spatial resolution (30 m), demonstrating that matching GEDI footprints' centers to the grid of optical sensors is a widely accepted method for canopy height mapping. Despite the difference in resolutions between GEDI (25 m) and Sentinel-2 (10 m at best), rasterizing GEDI data to a finer grid can provide additional detail, especially in areas with heterogeneous canopy structures. Schwartz et al. (2024) explored this scale mismatch in their study on canopy height mapping in the Landes forest (France). They tested a modified prediction model at a 20-m resolution (i.e., rasterizing GEDI data to a 20-m grid and producing a canopy height map at 20 m), specifically designed to reduce the scale mismatch between GEDI and Sentinel data. They found that their original model with a 10-m resolution performed slightly better in terms of accuracy and that it was also better at capturing small-scale canopy variations such as gaps or holes. This suggests that rasterizing GEDI data to a higher-resolution grid can provide more detailed information, particularly in capturing heterogeneous canopy structures. All things considered, while the scale mismatch could contribute to some variability in canopy height estimates, this approach still retains meaningful spatial variations and has been widely used in similar studies with good results in practice.

2.2.2 Optical data

The Sentinel-2 (S2) satellite constellation, operated by the European Space Agency (ESA), consists of two satellites, Sentinel-2A and Sentinel-2B, which operate in tandem to ensure a high revisit frequency (5 days) and complete coverage. It provides high-resolution multispectral imagery in 13 spectral bands ranging from visible and near-infrared to shortwave infrared wavelengths. This radiometric diversity enables detailed characterization of stand compositions and vegetation biophysical parameters by analyzing the spectral information contained in the reflectance images (Karasiak et al., 2017; Grabska et al., 2019). The data acquired from S2 are first processed to Level-1C, providing top-of-atmosphere reflectance values that are geometrically corrected. In this study, we utilized Level-2A products, which include atmospherically corrected bottom-of-atmosphere reflectance values using ESA's Sen2Cor processor (Main-Knorn et al., 2017).

Using the Google Earth Engine (GEE) data catalog and developing platform, we selected S2 images with a maximum cloud cover percentage of 80% (i.e., images with more than 80% cloud coverage were excluded) over the area of interest in the period between November 2019 and February 2021, with the goal of creating a cloud-free composite image for the year 2020. A cloud mask was applied to each image using sophisticated filters and thresholds from the PINO algorithm version 26 (Simonetti et al., 2021). The cloud and shadow mask computation was based on the Level-1C products and the results were eventually applied to the correspondent Level-2A images. After cloud masking, we generated a composite image for the year 2020 by computing the per-pixel median value across the collection of Level-2A S2 images. We opted for the median composite to mitigate the influence of outlier pixels, such as those affected by clouds or cloud shadows that may persist despite the cloud mask (Corbane et al., 2020; Pimple et al., 2022). Ten spectral bands were retained and, if necessary, resampled to 10 m: B2 (Blue), B3 (Green), B4 (Red), B5-B6-B7 (Vegetation Red Edge), B8 (Near Infrared, NIR), B8A (Narrow NIR), and B11-B12 (Short Wave Infrared, SWIR). The pixel values represent surface reflectance scaled by 10,000. The Sentinel-2 data used in this study were acquired from two distinct orbits. Despite S2 atmospheric corrections, which are more challenging in tropical contexts, differences in reflectance values between the two orbits remain. To mitigate this effect, and given that the two orbits overlap, we conducted a relative reflectance normalization for each spectral band separately. This process involved minimizing a cost function based on the mean and standard deviation within the overlapping area (Cresson and Saint-Geours, 2015).

2.2.3 Radar data

Radar data provide penetration capabilities through cloud cover and offer valuable information on forest structure (Baghdadi et al., 2015). Depending on their wavelength, radar signals can also penetrate the forest canopy to different depths, providing information on forest structure and on the lower layers of the canopy. Sentinel-1 (S1) is a C-band Synthetic Aperture Radar (SAR) system (wavelength of about 6 cm) while the Advanced Land Observing Satellite (ALOS) operates in the L-band (wavelength of about 25 cm), which allows us to benefit from the complementarity between these two bands.

S1 is comprised of two satellites with a 12-day repeat cycle, which results in a combined revisit time of 6 days for the same location on Earth. In this study, we utilized Ground Range Detected (GRD) scenes featuring dual-band cross-polarization: Vertical-Vertical (VV) and Vertical-Horizontal (VH), with a pixel size of 10 m. These scenes underwent preprocessing with S1 Toolbox, including thermal noise removal, radiometric calibration, and terrain correction using SRTM. The S1 GRD products provide Sigma-0 backscattering coefficients (in dB), which quantify the microwave radiation scattered back to the radar system. Using GEE interface, we extracted S1 images encompassing the area of interest in the period between November 2019 and February 2021 and the Sigma-0 values were calibrated using Gamma-0 normalization in order to correct the effects of radar signal attenuation due to varying incidence angles. Subsequently, we separated the data into ascending and descending orbits to generate mean composite images for the year 2020 for each orbit category and polarization. While median compositing is commonly used for optical imagery to create cloud-free composites, SAR data are not affected by clouds, enabling the use of mean compositing for constructing a data stack (Verhegghen et al., 2016; Barenblitt et al., 2024). This process resulted in a composite stack composed of four layers: VV ascending, VH ascending, VV descending, and VH descending.

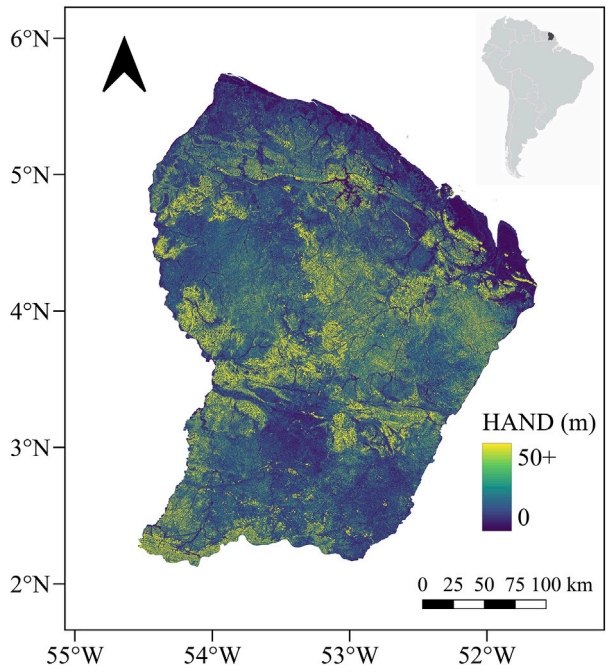


FIGURE 2
Height above nearest drainage (HAND) over French Guiana. 93% of pixels present a HAND value between 0 and 50 m. The maximum HAND value is 639 m.

ALOS Phased Array-type L-band SAR (PALSAR) is a system operated by the Japan Aerospace Exploration Agency (JAXA), featuring a sun-synchronous orbit and an observation frequency of 46 days. In this study, we utilized Normalized Radar Backscatter (NRB) scenes from the ALOS-2 PALSAR-2 ScanSAR Level 2.2 data product, which consists of dual-polarization observations: Horizontal-Horizontal (HH) and Horizontal-Vertical (HV), initially at a spatial resolution of 25 m that was resampled to 10 m. These scenes were subject to ortho-rectification as well as radiometric terrain correction and are provided in the Gamma-0 backscatter convention (in dB). Through the GEE platform, we selected ALOS-2 PALSAR-2 images covering the study site between November 2019 and February 2021 to produce a mean composite image for each dual-polarization, resulting in a composite stack comprised of two bands: HH and HV.

2.2.4 Environmental data

Topography has an impact on both remote sensing acquisitions and forest structure (Liu et al., 2021; Kutchartt et al., 2022). French Guiana is fairly flat, which is why topographic factors were not taken into account to correct GEDI waveforms, as their impact on canopy height estimates is minimal in these conditions (Fayad et al., 2021). Nonetheless, topography remains a crucial hydrologic driver because it determines the direction and velocity of water flows. Flow directions establish hydrological connections between various points within a basin, which in turn directly impact forest structure and dynamics (Ferry et al., 2010; Jucker et al., 2018; Muscarella et al., 2020). Consequently, canopy features are closely linked with soil and hydrological conditions, relationships that optical or radar imagery alone cannot fully reveal. For instance, in the Amazon, it is challenging to differentiate non-floodable swampy forests, which are characterized by waterlogged soils but are not subject to regular river flooding, from non-floodable terra-firme forests, characterized by well-drained soils, using only two-dimensional spectral data (Rennó et al., 2008). To address this, the height above nearest drainage (HAND) algorithm was developed to introduce the concept of height difference along flow channels, or drainage capacity, as a distinctive terrain descriptor (Rennó et al., 2008). This algorithm generates a normalized digital elevation model, which normalizes topography to the relative heights along the drainage network. Each pixel value corresponds to the vertical distance in meters to the nearest drainage. As a result, HAND provides essential context for interpreting canopy height variations. In this study, we utilized the Global HAND product available in the GEE catalog. In French Guiana, we extracted the global 30 m HAND product using a threshold of 100 river head cells, and we resampled it to a spatial resolution of 10 m to match the expected resolution of our canopy height map (Figure 2). The threshold of river head cells specifies the minimum number of grid cells required to contribute to water flow and establish the starting point of a stream or a river. This threshold determines where the flow accumulation is sufficient to initiate a river head in the digital elevation model and sets the scale for drainage network extraction. It controls the density and extent of the river network identified in the HAND model.

Forest landscape types (FLT) provide information on the broader ecological context of French Guiana. This territory may appear simply as a vast and uniform area, but it is in fact composed of interconnected and complex ecosystems. In this study, a map delineating 20 forest classes, produced by the French National Forests Office (ONF), was utilized (Catalogue des habitats forestiers de Guyane, 2015). This FLT map was computed from several spatial analyses: (1) a geomorphological landscape map generated by ONF using a spatial analysis of a digital surface model derived from SRTM data (Guitet et al., 2013); (2) a predictive map of low-lying areas also developed by ONF using the same SRTM data and adapting the HAND algorithm; (3) a forest vegetation map based on SPOT-VEGETATION images, which allows representing forest structure variation (Gond et al., 2011); (4) a land cover layer for the coastal strip produced by ONF from aerial photography to provide precise delineation of specific coastal habitats; (5) a human footprint map highlighting disturbances related to forestry and mining activities (De Thoisy et al., 2010), supplemented with mapping of areas exploited since 1945. These datasets allowed building a comprehensive basis for describing the diverse forest ecosystems in French Guiana. The incorporation of environmental data, such as forest landscapes, into canopy height or biomass characterization approaches holds significant interest and has been explored in various studies to produce improved results (Fayad et al., 2016; Morin et al., 2022; Morin et al., 2023; Fassnacht et al., 2021). The FLT map used in this study was integrated in our prediction models in the form of a raster at a spatial resolution of 10 m, with each pixel value corresponding to a given class of forest landscape (Figure 3).

2.2.5 ALS data

ALS data were employed as reference ground truth to validate our canopy height map. These data were acquired by ONF through several surveys over various study sites in French Guiana between

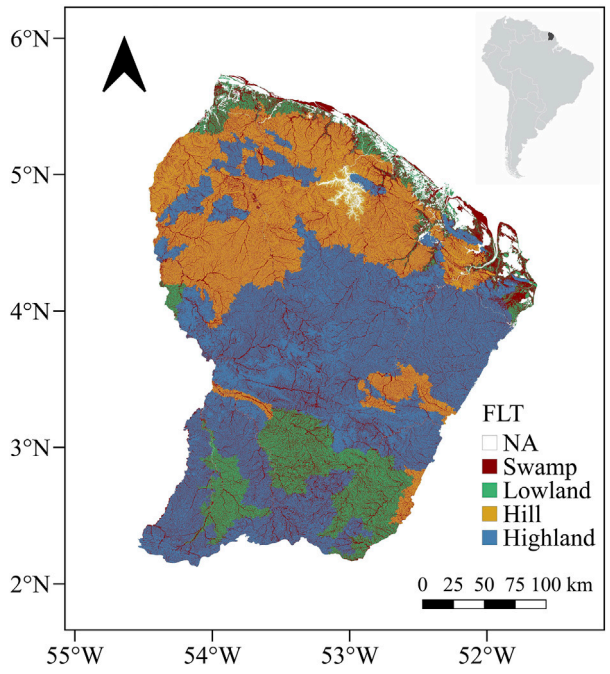


FIGURE 3
Main forest landscape types (FLT) of French Guiana. Each main class is divided into subclasses, resulting in a total of 20 FLT. NA represents non-forested areas, which include regions affected by past human disturbances such as logging and mining activities prior to 2015, as well as naturally non-forested areas. However, this map does not capture more recent disturbances and natural forest degradations.

2017 and 2020. The ALS acquisitions were characterized by technical specifications that included an average of 10–12 points per m^2 in average, with each pulse having a diameter of approximately 20 cm, and a wide scan angle of ±30°. A canopy height model (CHM) was derived by ONF at a resolution of 1 m (maximum height per 1-m grid cell), covering an area of 1,562 km^2. The CHM was resampled to a resolution of 10 m using the maximum value to match the resolution of our canopy height map. The choice of the maximum value in the resampling process is theoretically justified by the fact it better represents the top-of-canopy signal captured in GEDI waveforms, which are used as training data to build our prediction models. Adam et al. (2020) supported this approach, noting that the signal in return waveforms starts at the highest point of vegetation within the GEDI footprint. In a similar comparative study, Hilbert and Schmullius (2012) demonstrated that using the maximum rather than the mean value of ALS reference heights yielded a better correlation with GEDI metrics. Furthermore, Lahssini et al. (2024) found that GEDI-derived canopy heights exhibited stronger correlations with the maximum ALS heights within GEDI footprints.

2.3 Methods

2.3.1 Multimodal U-Net architecture

To produce a continuous canopy height map, we implemented a pixel-wise regression, which is a specialized task in machine learning to predict continuous values for each pixel in an image. This approach is particularly useful and relevant in our application, where each pixel represents a unique geospatial location and our goal is to estimate the canopy height for each of these locations based on their spectral and spatial features. In particular, fully convolutional networks (FCNs) are a type of deep learning model designed for spatially dense prediction tasks. They have been transformed from classical convolutional networks to handle inputs of any size and generate outputs of corresponding dimensions with efficient processing during both training and inference phases (Long et al., 2014). For the purpose of our study, a U-Net neural network architecture (Milesi, 2022) was used due to its proven efficacy in regression tasks for forest characterization involving multiple remote sensing data sources (Gazzea et al., 2023; Wagner et al., 2024; Schwartz et al., 2024; Ge et al., 2022). To manage and exploit the available multimodal information, each geospatial location (i.e., a pixel of 10 m) is described by means of multiple information sources as detailed in Subsection 2.2. U-Net is particularly good at capturing complex spatial dependencies in images and interpreting the local spatial context. Also, one of the key strengths of deep learning frameworks such as U-Net lies in their ability to integrate and leverage the complementarity between different datasets. The U-Net architecture is shaped like a 'U' because it sequentially downsamples then upsamples images. This design allows the network to learn robust feature representations at multiple scales. In the encoder part (contracting path), the input image undergoes two 3 × 3 convolutions, which are then followed by a Rectified Linear Unit (ReLU) activation function. This sequence is subsequently downsampled using a 2 × 2 max pooling operation with a stride of 2. This process is repeated four times, with the number of feature channels being doubled at each step to effectively capture hierarchical features at different scales. In total, there are 10 convolutional layers and 4 max pooling layers in the encoder part of the architecture. The decoder part (expansive path) mirrors the encoder but in reverse. It starts with bilinear upsampling, followed by concatenation with the corresponding feature map from the contracting path. This output is then processed through two 3 × 3 convolutions followed by a ReLU activation function. This process is also repeated four times similarly to the contracting path. This design helps in preserving spatial information throughout the network and improving the network's ability to localize features accurately. The final layer of the network consists of a 1 × 1 convolution, which produces a single-channel output image. This output image is the result of the pixel-wise regression performed by the U-Net model, i.e., a canopy height map derived from the multi-band input image. The decoder part of the network thus consists of 8 convolutional layers and 4 bilinear upsampling layers, maintaining symmetry with the encoder. The model, referred to as *CHNET* (Canopy Height estimation NETwork), and its components are described in detail in Figure 4. This figure provides a complete overview of the model's architecture as well as the flow of data through the network's layers, highlighting the encoder-decoder structure and the pixel-wise regression approach.

2.3.2 Experimental settings

To train the *CHNET* model, we employed a structured training process involving model calibration through a training/validation

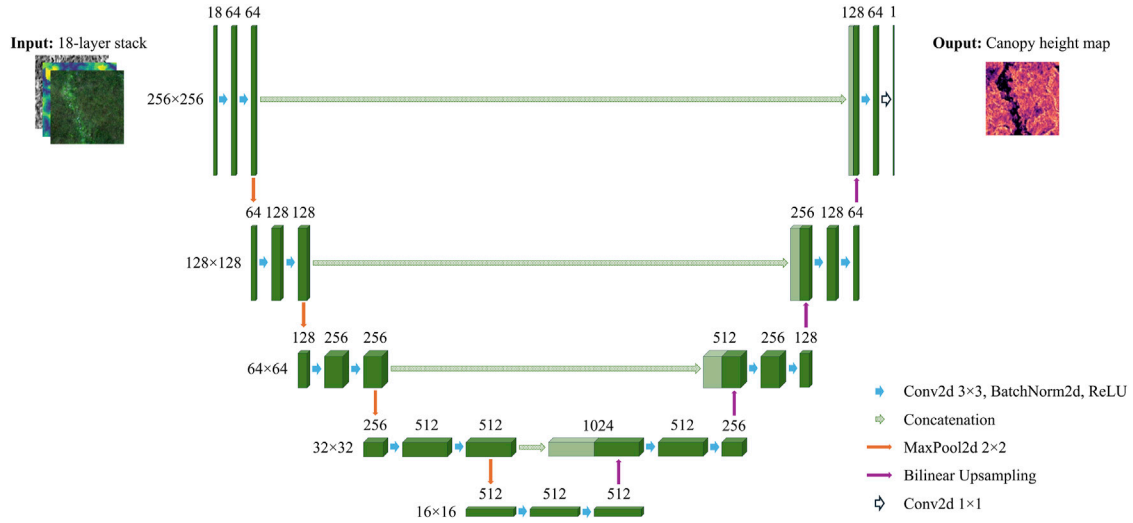


FIGURE 4
Overview of the *CHNET* framework. The input data consist of up to 18 layers depending on the training scenario. The output is a canopy height map. Image width and height as well as the number of channels are presented at each step of the model.

procedure (Ienco et al., 2017). The values of the different input sources were scaled between 0 and 1 (using minimum-maximum normalization) to help the model converge more quickly and efficiently by ensuring all features contribute proportionately to the learning process. The reference canopy height values (i.e., the GEDI *rh_95* values) were spatially separated into training, validation, and test datasets. To do so, French Guiana was divided into tiles of 768 × 768 pixels, with each tile covering an area of approximately 60 km^2. For model calibration, we only considered tiles with a minimum of 500 GEDI footprints, resulting in a total of 1,281 tiles. This dataset was allocated into training, validation, and test sets in an 80%-10%-10% split, corresponding to 1,024, 128, and 129 tiles, respectively. To avoid overfitting and improve the model's generalization to unseen data, we employed random spatial sampling during the training process to increase the diversity of the training samples. Specifically, from each training tile, we extracted random subsets of 256 × 256 pixels containing at least one GEDI shot. This sampling process introduced variability by ensuring that each training batch included different spatial regions of the tiles. While the model still learns meaningful patterns, this approach prevents it from simply memorizing specific configurations present in the training data. Instead, this randomized extraction of image patches encourages the model to learn more general and robust features while reducing the likelihood of overfitting to specific data configurations, as it exposes the network to diverse spatial contexts and pixel-level variations. Each 256 × 256 input image comprised normalized layers from the various data sources described in Subsection 2.2, while the corresponding reference image contained rasterized GEDI *rh_95* values.

The *CHNET* model's output was compared to these reference heights using a robust adaptive loss function (Barron, 2017), which computed the loss only for geospatial locations with a valid *rh_95* value. This loss function is a versatile definition that generalizes the Welsch (Dennis and Welsch, 1978), Geman-McClure (Geman and McClure, 1985), Cauchy (Black and Anandan, 1996), pseudo-Huber (Charbonnier et al., 1994), generalized Charbonnier (Sun et al., 2010), and L2 loss (least squared) functions. It is advantageous due to its capacity to adapt to varying noise levels and outliers in the data, providing a more stable and robust training process. As it unifies multiple common loss functions into a single formulation, it can adapt its shape based on the data it is optimizing and therefore it can handle a wide range of data distributions and noise levels. The loss function is defined as:

$$f(x, \alpha, c) = \frac{|\alpha - 2|}{\alpha}\left(\left(\frac{(x/c)^2}{|\alpha - 2|} + 1\right)^{\alpha/2} - 1\right) \quad (1)$$

x is the residual (i.e., the difference between predicted and true values), $\alpha \in \mathbb{R}$ is a parameter that controls the shape of the loss function, and $c > 0$ is a scale parameter that sets the size of the quadratic bowl near $x = 0$.

The parameters α and c in Equation 1 are learned during training using a gradient-based optimization method. The model automatically adapts the loss function to the data's characteristics. In our *CHNET* model training phase, the gradient of this loss was computed with respect to each model weight, and the weights were subsequently adjusted through loss backpropagation. To do so, we utilized a stochastic gradient descent (SGD) optimizer (Robbins and Monro, 1951) with a momentum of 0.9 and a cyclic learning rate scheduler with a "triangular2" policy (Smith, 2015). The scheduler's base learning rate value was set to 10^{-5} and the maximum to 0.1, with 488 steps for a half cycle (8 times the batch size). After each epoch of 1,024 images, organized as 64 batches of 16 images, we evaluated the loss on the validation data. The learning stage was conducted over 100 epochs, which allowed stabilizing the validation loss and completing the training process. We used early stopping based on validation loss to avoid overfitting during training. By monitoring performance on the validation set, we stopped training when the validation loss stopped improving for several

TABLE 1 Training scenarios and associated data.

Scenario	Input layers	GEDI reference
1	S2 + S1 + ALOS	• SNR >10 dB
2	S2 + S1 + ALOS	• SNR >10 dB • Refiltered
3	S2 + S1 + ALOS HAND + FLT	• SNR >10 dB • Refiltered
4	S2 + S1 + ALOS HAND + FLT	• SNR >10 dB • Refiltered • Geo-correction

consecutive epochs, indicating that further training would likely lead to overfitting to the training data.

2.3.3 Training scenarios

To understand the importance of each input data source, we designed four training scenarios based on combinations of different input layers and reference data. The first *CHNET* model (Scenario 1) was trained on optical and radar data as well as all the GEDI shots retained in the process described in Subsection 2.2. A first canopy height map of French Guiana at 10 m resolution was produced using this model. Following this first step, GEDI data were filtered again using the Scenario 1 canopy height map: for *rh_95* values lower than 20 m, GEDI shots that presented a difference between Scenario 1 canopy height and *rh_95* greater than 10 m were discarded. The goal of this filter was to remove low-height GEDI footprints that the model significantly overestimated, as these likely indicate unreliable GEDI data due to issues with vegetation signal penetration. Similarly, for *rh_95* values greater than 50 m, GEDI shots that exhibited a difference between *rh_95* and Scenario 1 canopy height greater than 10 m were removed. The goal of this filter was to eliminate GEDI shots that strongly overestimate height because of noise in the waveform that is interpreted as ground return by algorithm setting group number 5. The resulting refiltered GEDI database (1,875,608 shots, 16% of the initial dataset) was used as reference in Scenario 2. For the third model (Scenario 3), we enriched the Scenario 2 configuration with the environmental descriptors: HAND and FLT were added to the model input data. Finally, in Scenario 4 we accounted for GEDI geolocation uncertainty. Geolocation uncertainties in the context of GEDI refer to inaccuracies in determining precisely the spatial locations of the footprints on the Earth surface. The GEDI geolocation requirement as provided in the version 2 of the data products is that each footprint center is horizontally georeferenced to within 10 m, assuming normally distributed geolocation errors with a 0 m mean and a 10 m standard deviation (Dubayah et al., 2021). To address geolocation errors in our dataset, we implemented a controlled spatial perturbation (geo-correction) of each GEDI footprint center, systematically shifting them within both the *X* and *Y* directions across a radius of 10 m. Subsequently, for each possible spatial location, the optimal corrected footprint center was chosen based on the closest match to Scenario 3 canopy height map. The resulting corrected GEDI database was used as reference data for new model training in Scenario 4. Table 1 summarizes the characteristics of each scenario. The assessment of the models' performances was done considering the independent GEDI test set, the ALS ground truth data, and the global canopy height map produced by Lang et al. (2023). The bias, the root mean square error (RMSE), and the relative RMSE (rRMSE) were chosen as performance metrics to evaluate the accuracies of our models.

TABLE 2 Accuracy metrics of *CHNET* canopy height estimates vs. GEDI test dataset.

Scenario	Bias (m)	RMSE (m)	rRMSE
1	−2.8	8.7	13.2%
2	−1.0	7.6	11.5%
3	0.9	6.8	10.3%
4	0.7	5.7	8.6%

2.3.4 Canopy height map production

To produce the final canopy height map of French Guiana at a spatial resolution of 10 m, the study site was divided into tiles of 256 × 256 pixels for the purpose of applying the trained *CHNET* model to predict canopy height. Each tile was processed individually, and the results were subsequently merged to produce the final canopy height map. Due to the inherent nature of convolutional layers in the U-Net architecture, *CHNET* tends to struggle with accurately predicting edge values within each tile. This is because convolutions, while effective at capturing local spatial patterns, can introduce boundary artifacts because of padding and reduce prediction accuracy at the edges. To solve this issue, we divided the study area into overlapping tiles to ensure that the edges of one tile overlap with the central region of adjacent tiles. This overlap allows for the edge values in the final canopy height map to be replaced by the corresponding values from the overlapping regions of the adjacent tiles, which are predicted with higher accuracy.

3 Results

3.1 Evaluation with GEDI data

The accuracies of each training scenario are assessed against the independent GEDI test dataset established during the data preparation for *CHNET* calibration (*cf.* Subsection 2.3). This dataset was not used in the model training and validation process and was kept specifically for assessing performances. Table 2 shows, for each training scenario, the accuracy metrics between canopy height predictions and *rh_95* values from the GEDI test dataset.

The evaluation of canopy height estimates produced by *CHNET* against the GEDI test dataset reveals notable trends across the different scenarios. The model consistently demonstrates improvement across Scenarios 1 to 4, reflected in decreasing bias and RMSE values. The optimization of GEDI data used as reference height for model training (Scenario 2) allows reaching lower errors and relatively less biased estimates compared to the original GEDI dataset (Scenario 1). The integration of terrain and landscape descriptors (Scenario 3) further improves the results and exhibits a performance gain of about 10% in terms of RMSE compared to the model trained on optical and radar data only (Scenario 2). Notably,

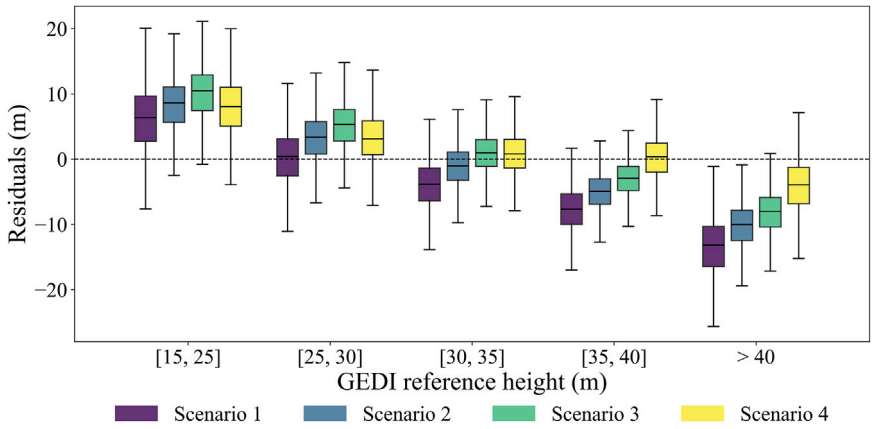


FIGURE 5
Residual analysis depending on canopy height class for all scenarios. Positive residuals indicate that predictions are greater than GEDI reference values. The boxplot shows the median, the quartiles, as well as the 10th and 90th percentiles.

Scenario 3 also shows a transition from a negative to positive bias value, indicating a tendency to mitigate the underestimation observed in preceding scenarios. Accounting for GEDI geolocation uncertainty through the iterative geo-correction process (Scenario 4) brings a significant improvement and produces the best model in terms of accuracy, with an RMSE value of 5.7 m and a minimal bias of 0.7 m. All things considered, both the enhancement and the enrichment of input data induce notable improvements in the accuracies of *CHNET*, with a global RMSE gain of about 35% between Scenarios 1 and 4.

The GEDI test set is important for evaluating the behavior of our models on an independent dataset that was not involved in the learning process. It allows us to assess the generalizability and the robustness of the models. Additionally, characterizing the canopy height map using the entire GEDI reference data, including the training and validation datasets, is essential to evaluate the quality and accuracy of each scenario's final map. By using all the available data, we can assess how the models perform across diverse conditions and regions. To achieve this, we computed the residuals of the models, defined as the differences between *CHNET* predicted canopy heights and GEDI reference values.

In particular, one key parameter to investigate is how the models perform across different canopy height classes, which is interesting given the height ranges found in French Guiana. Since most canopy heights in French Guiana forests range between 25 and 40 m (60% of the GEDI reference database), we analyzed the residuals according to the following height classes (in m): [15, 25], [25, 30], [30, 35], [35, 40], and >40. Heights below 15 m are quite infrequent in the study site and thus were excluded from the analysis. In the most relevant height ranges, i.e., between 25 and 40 m, we used a finer step of 5 m for each class. The box plot in Figure 5 illustrates the residuals (measured in m) of *CHNET* estimates across the GEDI reference height intervals for all training scenarios. The results show that for all scenarios, the residuals generally decrease as the GEDI reference height increases, indicating that the models tend to overestimate lower *rh_95* values on one hand, and underestimate higher *rh_95* values on the other hand. Regarding the overestimation, when reaching a reference height of up to 30 m, all scenarios still show positive residuals. The enhancement of GEDI data implemented in Scenario 2 and the inclusion of additional environmental descriptors in Scenario 3 do not yield better residuals for heights below 30 m compared to Scenario 1. Conversely, the geo-correction process developed in Scenario 4 results in lower residuals for heights between 15 and 30 m compared to prior scenarios, even though Scenario 1 still exhibits a slightly better performance. Regarding reference heights between 30 and 40 m, Scenario 4 produces unbiased estimates while other scenarios tend to underestimate canopy height. Overall, Scenario 1 exhibits the largest variability, with a tendency to consistently underestimate, especially in the [35, 40] m interval. Scenarios 2 and 3 bring notable improvements but are still outperformed by Scenario 4. Lastly, for heights greater than 40 m, all scenarios produce underestimated predictions, with Scenario 4 demonstrating the least underestimation and Scenario 1 the most pronounced.

Landscape type is another key parameter to investigate in the heterogeneous and diverse context of French Guiana's forest ecosystems. In this perspective, we analyzed the residuals according to four main forest landscape classes: swamp, lowland, hill, and highland. The box plot in Figure 6 displays the residuals (measured in m) of *CHNET* estimates across these four forest landscape categories for all training scenarios. The results show that for all FLT, the residuals gradually decrease in absolute value as we progress from Scenario 1 to Scenario 4. Overall, Scenario 1 shows the biggest variability and consistently underestimates canopy height across all four landscape classes. Scenario 2 also presents a slight tendency to underestimate. A notable observation that can be drawn from these results is that the inclusion of HAND and FLT in model training (Scenario 3) results in unbiased estimates per forest class. Furthermore, the optimal approach is once again Scenario 4, which exhibits significantly lower variability in residuals. Regarding FLT in general, variability is typically larger in highland areas, which are known to present greater challenges for land remote sensing applications compared to other landscapes.

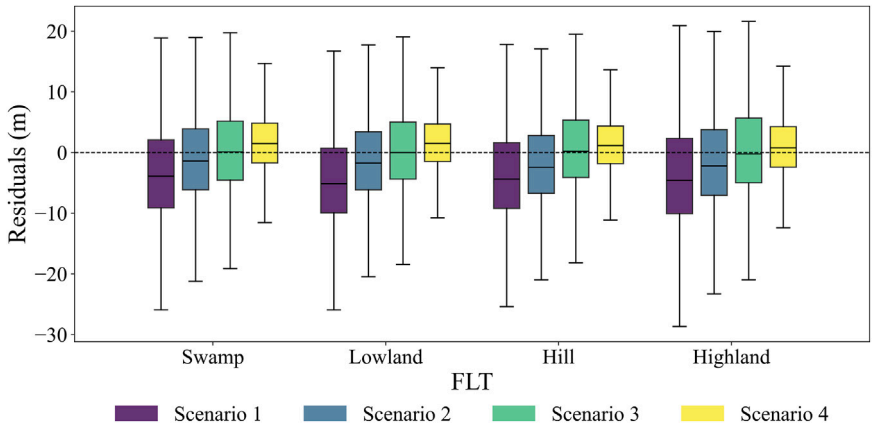


FIGURE 6
Residuals analysis depending on main forest landscape types (FLT) for all scenarios. Positive residuals indicate that predictions are greater than GEDI reference values. The boxplot shows the median, the quartiles, as well as the 10th and 90th percentiles.

TABLE 3 Accuracy metrics of *CHNET* canopy height estimates vs. ALS data.

Scenario	Bias (m)	RMSE (m)	rRMSE
1	−6.6	8.7	13.2%
2	−4.8	7.4	11.2%
3	−1.9	5.8	8.8%
4	−0.2	5.8	8.8%

3.2 Evaluation with independent ALS validation data

Canopy height models derived from ALS surveys were used to evaluate the performances of *CHNET* as well as to confirm the relevance of using GEDI height metrics as reference for canopy height. Table 3 presents, for each training scenario, the accuracy metrics between *CHNET* canopy height predictions and canopy heights retrieved from ALS data.

The evaluation of *CHNET* canopy height estimates against ALS data reveals a clear progression in accuracies across the four training scenarios. Overall, prediction models trained on GEDI data tend to underestimate canopy height, as highlighted by the negative bias values. Scenario 1, which is the one based on the least refined GEDI database in our comparative study, is characterized by a significant underestimation of canopy heights, with the strongest negative bias value of −6.6 m. The optimization of GEDI reference data implemented in Scenario 2 induces improvements, demonstrating the benefit of enhanced GEDI data integration. The most substantial improvements are observed in Scenarios 3 and 4, where the inclusion of additional environmental descriptors and the geo-correction process significantly improve the accuracy of canopy height estimates. This results in a global RMSE improvement of about 32% between Scenarios 1 and 4, along with a significant reduction in bias. The progressive rRMSE reduction from 13.2% in Scenario 1%–8.8% in Scenario 4 also underscores the effectiveness of these enhancements in providing more reliable canopy height estimates. Indeed, the geolocation correction (Scenario 4) nearly eliminates bias between *CHNET* estimates and ALS ground truth data. However, there is no notable enhancement observed in RMSE when performing geo-correction.

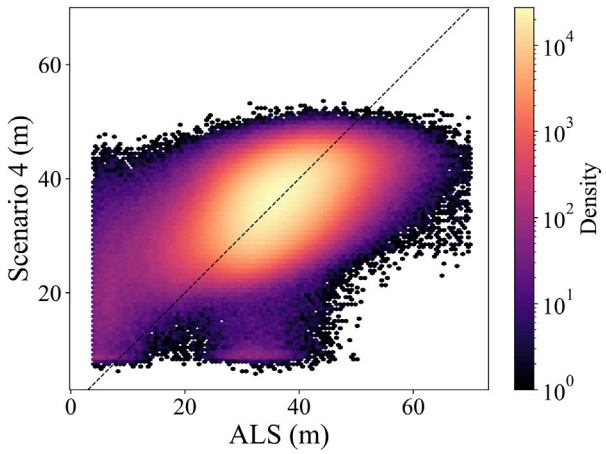


FIGURE 7
Scenario 4 *CHNET* canopy height estimates as a function of ALS ground truth data.

Figure 7 presents the scatter plot of Scenario 4 *CHNET* predictions against ALS ground truth data. Even though this configuration is the best performing model, it still exhibits noticeable difficulties at the edges, demonstrating a tendency to overestimate lower heights and to underestimate higher ones. However, it also shows robust performance within the height range of 30–40 m, which corresponds to the predominant height range observed in the forests of French Guiana. These findings are consistent with our earlier analysis using GEDI data (Figure 5), where Scenario 4 consistently provided accurate and unbiased estimates within the [30, 35] and [35, 40] m height intervals, while exhibiting tendencies towards overestimation at lower heights (lower than 30 m) and underestimation at higher heights (greater than 40 m).

TABLE 4 Accuracy metrics of *CHNET* canopy height estimates vs. Lang et al. global canopy height model.

Scenario	Bias (m)	RMSE (m)	rRMSE
1	−7.4	8.4	17.1%
2	−2.9	4.3	8.7%
3	−4.9	5.9	12.1%
4	−1.9	5.1	10.4%

TABLE 5 Accuracy metrics of Lang et al. global canopy height model vs. GEDI test dataset and ALS data.

Test data	Bias (m)	RMSE (m)	rRMSE
GEDI	3.3	8.0	12.1%
ALS	1.0	5.6	8.5%

3.3 Accuracies against global canopy height map

To assess our estimates relative to established models, we compared the *CHNET* predictions to Lang et al. high-resolution canopy height model of the Earth (Lang et al., 2023). Table 4 summarizes the comparison between the two models across the four different training scenarios. Overall, *CHNET* tends to underestimate canopy height when compared to Lang et al. estimates, as highlighted by the negative bias values. Enhancing the GEDI data used as reference to train *CHNET* (Scenario 2) induces a notable improvement with regards to Lang et al. model, achieving a minimal RMSE value of 4.3 m. However, incorporating environmental descriptors (Scenario 3) does not bring additional benefits compared to Scenario 2. The inclusion of HAND and FLT in our model does not lead to enhanced performance when assessed against Lang et al. map.

These observations are further refined when using ALS reference data to characterize both models together (Figure 8). In general, the residuals of Lang et al. estimates are consistently above those of *CHNET*, which explains the negative bias values reported in Table 4. Notably, for both Lang et al. and *CHNET*, the residuals increase as the ALS reference height decreases. Lang et al. overestimation tendency is particularly significant in the [15, 25] m height range. The key difference between the two models lies in the height at which the transition from overestimation to underestimation happens: Lang et al. tend to overestimate heights up to 35 m, whereas *CHNET* produces relatively unbiased estimates at 30 m. Conversely, for canopy heights greater than 35 m, Lang et al. outperforms *CHNET*, providing unbiased estimates in the [35, 40] m range and exhibiting less underestimation for canopy heights above 40 m.

In assessing the performances of Lang et al. canopy height map, which has its own uncertainties like any model-based product, we compared it against the same GEDI test dataset and ALS data to understand its accuracy relative to our own map (Table 5). The analysis reveals that Lang et al. map generally tends to overestimate canopy heights in French Guiana, a pattern consistent with our previous findings where our *CHNET* model typically underestimates canopy height compared to Lang et al. estimates. When evaluated against ALS data, Lang et al. product exhibits a slightly better RMSE value than our *CHNET* framework. However, Lang et al. map also exhibits a notably higher positive bias, confirming a tendency to overestimate canopy height. In contrast, results from the GEDI test dataset are more favorable to our model. Specifically, Lang et al. map demonstrates significant error metrics, which are in line with the geographical error analysis presented in their study on held-out GEDI validation data.

3.4 Canopy height map

We chose to implement the Scenario 4 *CHNET* model to the whole area of interest, as it yielded the best accuracies relative to GEDI and ALS reference data. Moreover, this approach is also the most developed as it relies on all available data and uses an

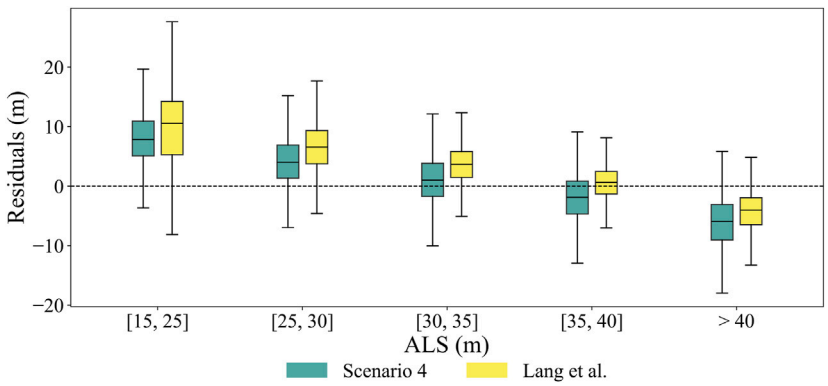


FIGURE 8
Residual analysis depending on ALS height class for Scenario 4 *CHNET* and Lang et al. map. Positive residuals indicate that predictions are greater than ALS reference values. The boxplot shows the median, the quartiles, as well as the 10th and 90th percentiles.

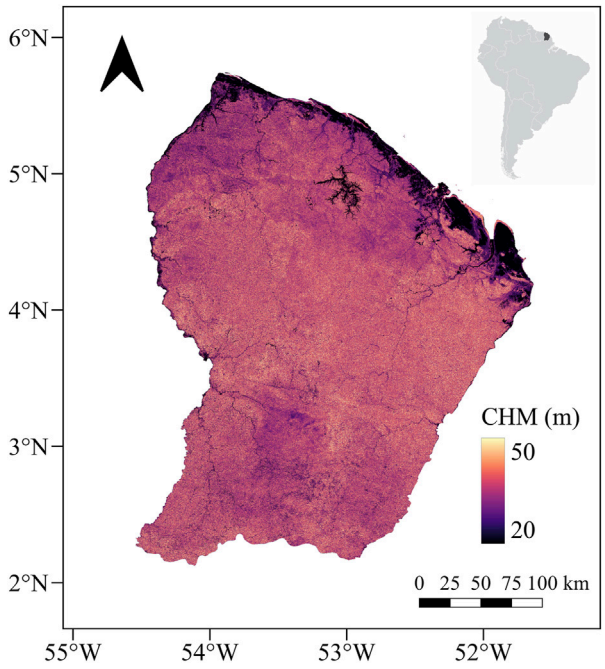


FIGURE 9
Canopy Height Model (CHM) of French Guiana at 10-m resolution predicted by *CHNET* built under the configuration of Scenario 4.

optimized GEDI database for training. Figure 9 presents the results obtained and the final canopy height map.

Despite *CHNET* being trained and validated on sparse reference data (rasterized GEDI *rh_95* metric), it successfully generates a continuous canopy height map with good precision. Indeed, this map offers a comprehensive depiction of forest structures and other landscape features observable in the remote sensing data. It effectively distinguishes height differences across the study area while describing the spatial distribution and variability of canopy heights in French Guiana.

4 Discussion

In this study, a U-Net model was assessed under different training conditions to generate a canopy height map of French Guiana. Each consecutive training scenario involved optimizing and enriching the data used in the learning process. Ultimately, the model that incorporated diverse data sources, including optical, radar, and environmental parameters, along with a refiltered and geolocation-corrected GEDI database, yielded the best results. This optimized *CHNET* model was applied to the entire study area to produce a high-resolution 10-m map of canopy height across French Guiana (Figure 9).

4.1 On the challenges of GEDI data

To train our *CHNET* model, a GEDI height metric from the L2A product was used as direct indicator of canopy height. Using GEDI metrics as ground truth data and considering them as a reliable representation of canopy height is the most straightforward and the most used approach in deep learning applications (Lang et al., 2023; Schwartz et al., 2024; Fayad et al., 2024). Theoretically, *rh_100* represents the top of the canopy cover, but other high-percentile RH metrics are used in practical applications because they are less sensitive to noise and contain less outliers (Li et al., 2024). We specifically chose the *rh_95* metric in accordance with a previous study where this metric showed the best performance in the dense tropical context of French Guiana (Lahssini et al., 2022). Other studies have also advocated for the use of this metric (Potapov et al., 2021; Lahssini et al., 2024; Dorado-Roda et al., 2021), although metrics like *rh_98* have also been used (Lang et al., 2023). Despite GEDI's unprecedented capabilities for forest structure characterization, there are still limitations to using these data as reference ground truth for empirical model calibration. Especially in our study area, the ability of GEDI to accurately retrieve canopy height in complex, tall, and dense forest structures is still uncertain. In tropical biomes, the GEDI signal may sometimes not penetrate enough to reach the ground, leading to incorrect height estimations (Lahssini et al., 2022). Moreover, the uncertainties in the GEDI footprint locations can lead to significant errors (Roy et al., 2021). Our training scenarios in this study were specifically designed to address and mitigate these existing challenges.

The GEDI L2A metrics are based on the L2A processing algorithm, which utilizes six different sets of parameters for waveform processing. Consequently, there are six possible values for canopy height for any given footprint. We opted for the data derived from algorithm setting group number 5, as previous research indicated it delivered significantly better accuracy than other groups, including the GEDI automatic algorithm selection, and particularly in the tropical context of French Guiana (Lahssini et al., 2022; Liu et al., 2024). Algorithm setting group number five is defined by the lowest waveform signal end threshold compared to other groups, which allows for better extraction of low-intensity ground returns. However, this advantage comes with a trade-off, as algorithm setting group number five tends to overestimate lower heights. Low-height areas are generally not concerned by signal penetration challenges. Due to its greater sensitivity to noise, algorithm setting group number 5 may interpret noise below the actual ground as information on ground elevation, leading to overestimations of low canopy heights. This behavior is observed in the results of our *CHNET* model and is consistent with similar models in existing literature (Schneider et al., 2020; Schlund et al., 2023). In our case, using GEDI metrics computed with the parameters of algorithm setting group number five to calibrate *CHNET* tends to produce overestimated predictions for heights lower than 30 m. While algorithm setting group number five offers a better detection of weak ground returns, the parameters of this signal processing algorithm naturally lead to an overestimation of lower heights. Alternative approaches could be explored to mitigate this overestimation of lower heights. For instance, Potapov et al. (2021) adopted an approach in their global canopy height map where they averaged the middle four values of the RH metrics across all six algorithm configurations. Given the global scale of their study, relying on a single configuration would not be suitable to capture the diversity of forest structures and ecosystems accurately. Their method, used

in their global dataset, aims to harmonize GEDI metrics across diverse biomes by avoiding extreme values (minimum and maximum of the six setting groups) and minimizing the biases introduced by a single algorithm configuration. Similarly, in their supervised machine learning approach to interpret GEDI waveforms and regress canopy top height globally, Lang et al. (2022) emphasized that, while algorithm setting group number 2 performed best over all continental areas, the performance of individual algorithm setting groups changes regionally. These findings, although obtained at global scales, suggest that local adaptations could possibly improve the reference heights used to train our model. Given the variety of forest landscapes across French Guiana, future studies could consider the use of different GEDI algorithm setting groups for different forest types. By exploiting several algorithm settings each adapted to the specificities of distinct areas, it might be possible to reduce the overestimation of lower canopy heights, a known limitation of algorithm setting group number 5, while preserving the accuracy for taller canopies.

While GEDI tends to overestimate lower canopy heights, it conversely underestimates higher canopy heights. Even though we are using the algorithm setting group with the lowest signal end threshold, a phenomenon of underestimation is present for high heights. Existing standard methods for canopy height estimation often face challenges in accurately measuring tall canopies, with height estimates typically saturating at approximately 25–30 m (Potapov et al., 2021; Healey et al., 2020). In our case, the transition appears at around 35 m. For heights greater than 35 m, our *CHNET* model, which is built on GEDI data, starts to underestimate canopy height. Tall canopies are generally associated with densely vegetated areas, which pose a significant challenge for GEDI laser signal penetration. This issue is particularly present with coverage beams, which are half-power beams compared to the full-power configuration. Pulses emitted with coverage lasers generally have more difficulties reaching the ground because of the density of the vegetation, causing the extracted ground peak in the waveforms to appear at a higher elevation than the actual ground level. This discrepancy leads to an overestimation of the ground height and, consequently, and underestimation of canopy heights. Several studies have highlighted the link between laser energy and vegetation penetration. For example, Fayad et al. (2022) observed that coverage lasers exhibited significantly lower performance for canopy height estimation compared to full-power signals such as GEDI power beams and NASA's Land Vegetation and Ice Sensor (LVIS). In this context, many studies recommend using only high-power beams for better canopy height estimates (Lahssini et al., 2022; Liu et al., 2021; Wang et al., 2022). However, deep learning approaches heavily rely on a sufficient amount of reference data for training, which is why most models built on GEDI data do not filter by beam type and use all the available footprints (Lang et al., 2023; Schwartz et al., 2024; Potapov et al., 2021). Furthermore, full-power beams also face challenges when penetrating dense canopies, as laser energy is not the only parameter affecting signal penetration. According to specifications, the GEDI instrument is capable of measuring vertical canopy profiles in environments with up to 95% canopy cover for coverage beams and up to 98% for power beams (Dubayah et al., 2020). All things considered, in a region like French Guiana, GEDI inherently exhibits a tendency to underestimate tall canopy heights in densely vegetated areas, which in turn impacts the *CHNET* model estimates in the same way. The saturation of optical and radar data also contributes to this underestimation.

In our study, we explored various training scenarios to optimize canopy height estimation from GEDI data, and we found that the inclusion of geolocation correction for GEDI footprints produced the most accurate results. GEDI geolocation uncertainty is a major challenge that limits the benefits of the data. It is a common issue across all remote sensing instruments due to orbital dynamics, instrument calibration, and atmospheric disturbances (Roy et al., 2021). Addressing these uncertainties is crucial for any GEDI-based application and various studies proposed mitigation strategies (Hancock et al., 2019; Shannon et al., 2024; Schleich et al., 2023; Tang et al., 2023). The expected accuracy of the GEDI version 2 data product is 10 m. In a previous study, we had estimated that geolocation uncertainty might account for approximately half of the error observed in canopy height estimates (Lahssini et al., 2024). To address geolocation errors, we implemented an iterative spatial perturbation of each GEDI footprint center, leveraging previously developed models to refine footprint locations. This approach significantly improved the accuracy of canopy height estimates through *CHNET*, demonstrating the potential benefits of using an enhanced GEDI database. Importantly, these enhancements can be achieved in an operational context without the need for additional validation datasets, relying solely on an iterative process that consists in consecutive model development under diverse training scenarios. Nonetheless, no reduction in error was observed when comparing model performance against ALS data before and after geo-correction (Table 3). This could be explained by the nature of the correction applied. The geo-correction improves the spatial accuracy of the GEDI footprints' locations, which is reflected in the improved performance of the model when compared against the GEDI test set (Table 2). This is expected because we are directly correcting GEDI information. However, ALS data is independent and has no direct connection to the GEDI footprints' locations. The discrepancy between GEDI and ALS measurements could limit the impact of the geo-correction. Since the model was trained on GEDI data, it may still carry the biases inherent to GEDI, and these would persist even after geo-correction when compared to ALS reference canopy heights. In previous training scenarios (i.e., Scenarios 1–3), we introduced new data or removed irrelevant information, which gave the model a different perspective on canopy height, and improvements were seen in comparison to ALS. In Scenario 4, no new data is added or removed (only a spatial correction is performed), and it could explain why no improvement in error is observed when assessing the model against ALS data.

Overall, using GEDI RH metrics as a direct and reliable proxy for canopy height has limitations due to the characteristics of the GEDI sensor. However, when mapping large areas using an empirical approach like *CHNET*, it is essential to have enough reference data for training. The optimal reference data would come from *in-situ* measurements or high-resolution sensors such as ALS. In our case, while we do have ALS data, it is insufficient for training our model, because it covers a relatively small area compared to the entire study site (less than 2%) and it is limited to the north of French Guiana, thus lacking diversity for a robust training set. To understand the accuracy of GEDI data and its

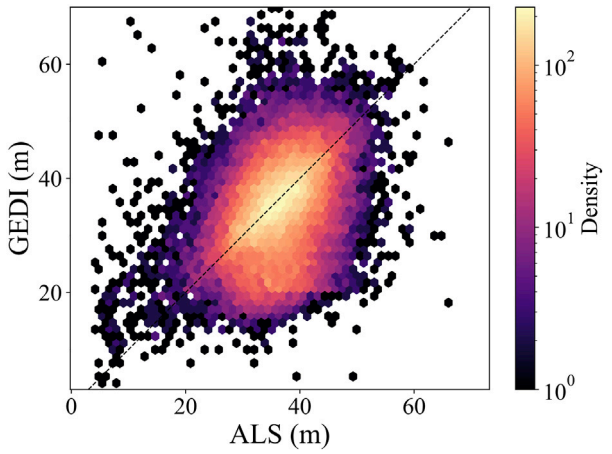


FIGURE 10
GEDI *rh_95* (Scenario 3) as a function of ALS ground truth data. We used the Scenario 3 GEDI reference data to highlight the actual relationship between GEDI data as they are provided and ALS.

relevance as reference data, we performed a comparison between GEDI *rh_95* values and the available ALS data. The results, as shown in Figure 10, indicate that GEDI is far from being a perfect ground truth, with an RMSE of 8.2 m and a tendency to underestimate canopy height, with a bias of −1.6 m. A large portion of the *CHNET* model's error can therefore be attributed to GEDI's inaccuracies. In a previous study, Lahssini et al., 2022 demonstrated that reference canopy height values could be improved by building models based on RH metrics and calibrated on actual ground truth data. While such an approach is effective at local scales, it becomes challenging at a regional scale, such as across the entirety of French Guiana, due to the limited availability of ALS data over the entire study site. When studying larger study areas using GEDI data, an iterative process of refining and enhancing the input reference data at each step proved effective. Indeed, it allowed bringing notable improvements and partly overcoming the challenges of working with GEDI data.

4.2 On the value of input data sources

The complementarity between optical and radar is particularly valuable to derive canopy height, as each sensor provides specific information. They are commonly used to characterize forest ecosystems, and their complementarity has been leveraged in several studies across a range of biomes, including temperate and tropical regions (Schwartz et al., 2024; Morin et al., 2023; Fagua et al., 2019). Optical data, such as that from S2, is generally known for its ability to provide information related to stand composition (Karasiak et al., 2017; Grabska et al., 2019). However, it also offers valuable features for characterizing structural parameters of the vegetation, especially through certain spectral bands (Lahssini et al., 2022). In fact, the reflectance of a pixel is primarily influenced by the characteristics of the foliage, such as its spectral properties, quantity, and orientation, rather than trunk biomass or tree height. However, the spatial arrangement of pixels, often referred to as texture, gives valuable information that can be linked to structural parameters of the forest. Deep learning techniques consider both spectral and textural aspects. In our case, the *CHNET* model's architecture allows extracting and linking local spatial features to canopy height. Other studies using machine learning algorithms have demonstrated that incorporating multispectral data enhances the accuracy of forest canopy height estimation and decreases the uncertainty of these estimates, even for tall forests (García et al., 2018). However, optical data can be limited by the saturation of reflectance in tropical forests, and this saturation makes it hard to distinguish vegetation heights above a certain level (Simard et al., 2011). Sentinel-2 data alone are not the optimal source for predicting forest variables that are mainly influenced by tree and stand structure, especially canopy height.

Therefore, to build our *CHNET* model, we complemented these data with other sources, particularly radar information. In this study, we utilized two radar satellite platforms operating in different bands: C-band for S1 and L-band for ALOS. Different polarizations of radar data, as implemented in our *CHNET* model, further enhance forest characterization. HH and VV polarizations are generally associated with surface scattering and can be useful for assessing forest density and structure (Wijaya et al., 2015). Cross-polarizations like HV and VH are more sensitive to volume scattering, which can help describe canopy structure (Ulaby et al., 1990). In the future, the European Space Agency (ESA) planned BIOMASS mission, which is specifically designed for forest structure and biomass estimation, will collect P-band SAR data and provide global estimates of forest biomass and height (Quegan et al., 2019). P-band SAR has a significantly longer wavelength compared to C-band or L-band systems, which allows penetrating dense canopies more effectively. Indeed, longer wavelengths are less affected by canopy scattering and absorption, allowing for greater penetration into the canopy (Khati et al., 2018). Moreover, even L-band radar signals are known to encounter difficulties penetrating canopies when biomass levels are beyond 150 Mg/ha (Mermoz et al., 2015), a common scenario in French Guiana. Specifically for the BIOMASS mission, dual-polarization and interferometric capabilities will also bring a valuable opportunity for better discrimination of different types of vegetation and terrain, making it an ideal complement to GEDI data.

In our study, we observed that enriching the remote sensing input data with environmental descriptors HAND and FLT induced a significant enhancement in the canopy height estimates produced by *CHNET*. Environmental parameters generally impact both the studied environment and the data collected by the sensors acquiring measurements over that environment. Specifically, hydrological and geomorphological descriptors like HAND and FLT are directly related to forest state, structure, and dynamics. They provide valuable contextual information that complements optical and radar data. HAND gives contextual information about a pixel's vertical proximity to the river within its watershed. Such information can relate to water table depth, access to water sources through aboveground or belowground drainage flows, and variations in vegetation species along drainage transects. For example, Schietti et al. (2014) identified a link between HAND and floristic composition changes in the Amazon. HAND captures the elevation relative to drainage networks, which influences water availability and soil moisture, critical factors affecting in turn vegetation growth, canopy height, and AGB. Indeed, some studies have established a link between HAND and forest structural parameters. For example, the link between HAND and

biomass in Eucalyptus clonal plantations in Brazil indicates that the functioning and the dynamics of a same species can be very different over a gradient of vertical distance from the river (Stape and Alvares, 2023). In a region like French Guiana, characterized by complex and highly developed water networks as well as a rich variety of species, hydrological configuration therefore strongly impacts forest structure. Incorporating HAND into our *CHNET* model provides unique and specific information on canopy height that is not available from the other data sources we used, which explains the observed improvements in model performance.

Similarly, FLT categorizes forest types based on their landscape characteristics These characteristics include soil type, geological and geomorphological factors, terrain elevation, and vegetation types. This categorization provides key insights into the structural diversity of French Guiana's forest ecosystems, and those factors ultimately impact the spatial arrangements of canopy height. Integrating FLT data into the inputs of *CHNET* helps the model gain a more detailed understanding of how different forest landscape types exhibit distinct canopy height profiles. In their study on canopy height estimation from multi-source remote sensing data using a Random Forest model, Jin et al. (2018) noted that models trained on given locations or vegetation types were not transferable to other settings. However, they also found that training their algorithm with data from diverse sites and vegetation types enabled the development of a universal model capable of accurately predicting canopy height across various locations and vegetation types. Our results also underline the importance of including environmental descriptors when implementing and applying a model to extensive areas characterized by diverse ecological conditions. HAND and FLT contain environmental information that is directly linked to ecological processes influencing canopy dynamics, and the *CHNET* model effectively incorporates this information alongside optical and radar inputs to identify complex relationships between environmental descriptors and canopy height. It is important to note that the FLT map utilized in this study was produced in 2015 by ONF and incorporates data on human activities to map non-forested areas, as reflected in the NA class in Figure 3. This includes areas affected by human disturbances up to 2015, such as urbanization and historical logging or mining. However, more recent disturbances, such as illegal logging and mining activities after 2015, or natural forest degradation, are not represented. While this limitation could affect canopy height predictions by potentially underestimating the impact of human activities, the conclusions of our study remain robust. Nonetheless, accounting for these disturbances could further refine the input data used by the *CHNET* model. The Joint Research Center's (JRC) Tropical Moist Forest (TMF) dataset (Vancutsem et al., 2021), which provides detailed records of forest cover and land use changes in tropical regions, could be a useful resource in this regard. It has been successfully used in other studies to focus on undisturbed forested areas and facilitate consistent comparisons across datasets from different periods (Lahssini et al., 2024). Even if all data sources used in this study were obtained from relatively contemporary acquisitions, the inclusion of ancillary information about forest cover changes could still allow for an interesting refinement of the input data, by filtering out even more disturbed or non-forested areas beyond what the FLT's NA class shows (Figure 3). Using this additional information, the *CHNET* model would gain a more accurate understanding of land use and forest type during its training process, which would potentially lead to better accuracy in canopy height estimations.

While the *CHNET* model does not allow for traditional feature importance analysis (as is possible in machine learning models like Random Forest), we assessed the contribution of environmental predictors by comparing models trained with and without these inputs. The observed improvement in canopy height predictions suggests that HAND and FLT are indeed informative because of their strong correlation with forest structural properties, which has been noted in previous studies (Rennó et al., 2008; Fayad et al., 2016). Although HAND and FLT were effective in this study, other candidate variables could also be explored to further improve the *CHNET* model's canopy height estimates. Yang et al., 2016 found that climate and soil factors, such as water stress, precipitation, and soil fertility, can be significant in explaining canopy height variations at large scales. They suggested that soil and climate jointly explain a quite significant portion (about 30%) of height variability in tropical forests. Fayad et al. (2016) also highlighted the importance of environmental variables like topography, geological features, and rainfall data in mapping forest biomass in French Guiana, which correlates with forest structure and canopy height. Their study emphasized that variables capturing terrain and hydrological conditions, similar to the information given by the HAND variable, are key to explaining AGB distribution. However, it is also important to consider that not all abiotic variables will always improve model performance. For example, while Yang et al., 2016 noted that soil and climate variables can explain large-scale height variations, they might not be as directly useful for small-scale canopy height predictions, because local terrain and forest characteristics often dominate at such scales. Therefore, the inclusion of these variables was not prioritized in our study due to the specific ecological and geographical context of French Guiana, where HAND and FLT were expected to play a dominant role in forest structure variability. Future work could nonetheless explore incorporating other candidate variables, such as soil properties or climatic factors, to test whether they could add further predictive power, particularly in regions with distinct ecological characteristics. While these additional variables might be informative, they should be chosen carefully based on their relevance to canopy structure and the specific scale of the study.

4.3 On the relevance of the U-Net architecture

When dealing with heterogeneous data sources and sparse reference data, the U-Net architecture represents a particularly robust approach. It is well-suited for tasks involving complex data integration and spatial feature extraction due to its structure that combines a contracting path to capture context and an expanding path to enable precise localization (Ronneberger et al., 2015). Previous research has indeed shown that performance decreases significantly, especially for tall canopies, when models are unable to learn spatial features (Lang et al., 2019). The *CHNET* model's capacity to derive and analyze features from diverse inputs, i.e., optical and radar remote sensing data alongside environmental descriptors, is a key advantage that we leveraged in this study to

produce our final canopy height map. Remote sensing data inherently contains both spectral and spatial features, and the U-Net architecture is designed to efficiently extract and interpret these multi-dimensional features. The convolutional layers in U-Net neural networks are great at capturing both spectral variations and spatial patterns directly from the raw data, thus removing the need for preliminary feature engineering like computing vegetation indices or other spectral indicators. In their canopy height model of the Earth produced with an ensemble of FCNs, Lang et al., 2023 noted that while there is a good correlation between the Normalized Difference Vegetation Index (NDVI) and the estimated canopy height, the relationship learned by the model between the S2 image features and canopy height is much more complex and cannot be fully described by a straightforward index like NDVI. Therefore, the direct utilization of raw spectral bands ensures that the model can exploit the full richness of the input data.

Furthermore, the U-Net architecture is advantageous when reference data are sparse, which is the case of GEDI measurements. This capability is of paramount importance when mapping canopy height over a large region such as French Guiana, where the only reference data available in sufficient quantity for model training consists of GEDI measurements that are sparsely and unevenly distributed (Figure 1). The model's ability to generalize from sparse reference points ensures that predictions are spatially consistent across the entire study area. This characteristic was particularly beneficial when training our *CHNET* model on sparse reference GEDI data and when applying it to produce our final canopy height map. Moreover, given the complexity of the factors impacting canopy height, we chose to build our *CHNET* model using multiple complementary data sources through a data fusion approach. Indeed, hydrological and geomorphological descriptors related to forest growth have demonstrated their ability to provide supplementary information not found in classical remote sensing data. The strategy for integrating and combining all this information together is important for maximizing its value. Some approaches process each data source separately using dedicated encoders before performing data fusion (Chen et al., 2024). In our study, the convolutional approach in the *CHNET* framework preserves the spatial hierarchies and relationships within the data, which is crucial for accurately mapping a variable like canopy height that varies continuously across landscapes (Lang et al., 2019). Convolutional filters allow linking multi-dimensional features with the target variable. All things considered, the U-Net model's ability to integrate multi-source data, to identify relevant spectral and spatial features without the need for pre-computed indicators, and to perform robustly with sparse reference data makes it a relevant choice for producing a high-resolution canopy height map of French Guiana in the conditions described in this study. While more recent segmentation architectures like Transformer-based models (Fayad et al., 2024; Tolan et al., 2024) or attention-enhanced networks (Zhao et al., 2024) have shown success in forest canopy height retrieval, the U-Net's well-established performance and efficiency in handling multimodal data, combined with its relatively lightweight architecture, made it the more practical choice for the scale and resolution of this study. Newer architectures might offer enhanced feature extraction, but they also often come with increased computational and memory requirements. Hence, the choice of U-Net allowed us to reach a balance between computational efficiency and model performance. Moreover, the study's significant conclusions are about the data themselves, as we achieved clear and consistent improvements in our canopy height predictions with U-Net. In the end, our findings highlight that, regardless of the architecture employed, the quality and integration of the data are important factors in producing accurate canopy height estimates.

5 Conclusion

This study presents an operational deep learning application that leverages multi-source satellite data and ancillary environmental information, all of which are freely available to users on a global scale. We demonstrate the potential of a U-Net architecture trained on sparse and unevenly distributed data to generate a continuous map of canopy height at a regional scale. Our assessment reveals that GEDI can be effectively used to generate canopy height products, particularly when integrated with other data sources into a deep learning estimation model.

Our *CHNET* framework successfully produced a canopy height map of French Guiana at a 10-m spatial resolution with relatively good accuracy. Notably, we observe that it is essential to account for GEDI inaccuracies as well as geolocation uncertainties, and this consideration should be integrated into any GEDI-based application. The iterative geolocation correction approach presented in this study is novel in that, unlike many existing correction methods that rely on external datasets (for example, airborne LiDAR point clouds) and calibration procedures, it is highly operational and does not require additional data to be performed. Given its efficiency and simplicity, this approach could be explored for use in other regions, both tropical and non-tropical, to enhance the accuracy of GEDI-based applications. Moreover, considering ancillary data that provides complementary insights into the structural dynamics of the canopy, which classical remote sensing data cannot fully reveal, has proven to be of significant interest. Incorporating relevant environmental descriptors closely linked to forest growth contributes significantly to the model's accuracy, as our deep learning framework *CHNET* was able to leverage these complex relationships to improve estimates. The integration method of multimodal data is crucial for fully exploiting the diverse and multi-level information contained in the different input layers, especially given the specificities of GEDI reference data.

Despite these advancements, challenges remain, particularly in accurately assessing tall canopies with GEDI data. Addressing these challenges is critical in tropical forest studies. For instance, improved GEDI-derived reference canopy height values could be obtained by applying linear or nonlinear models based on GEDI RH metrics and calibrated on actual ground data, such as ALS, provided that there is a sufficient amount of data available to build these models. Future improvements in the characterization of dense canopy structures are essential for better understanding tropical forest ecosystems and the

Earth's carbon cycle, with a need for both enhanced data quality and improved methodologies.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

KL: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft. NB: Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing–review and editing. GL: Conceptualization, Investigation, Methodology, Supervision, Validation, Writing–review and editing. IF: Conceptualization, Investigation, Methodology, Software, Validation, Writing–review and editing. LV: Data curation, Writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research received funding from the National Center for Space Studies (CNES, TOSCA 2024 project) and the National Research Institute for Agriculture, Food and the Environment (INRAE).

Acknowledgments

The authors would like to thank the GEDI team and NASA's LP DAAC for providing the GEDI data. The authors also wish to thank the National Forests Office (ONF) and Caroline Bedeau for processing and providing the ALS data.

Conflict of interest

Author IF was employed by Kayrros SAS.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adam, M., Urbazaev, M., Dubois, C., and Schmullius, C. (2020). Accuracy assessment of GEDI terrain elevation and canopy height estimates in European temperate forests: influence of environmental and acquisition parameters. *Remote Sens.* 12 (23), 3948. doi:10.3390/rs12233948

Asner, G. P., and Mascaro, J. (2014). Mapping tropical forest carbon: calibrating plot estimates to a simple LiDAR metric. *Remote Sens. Environ.* 140 (January), 614–624. doi:10.1016/j.rse.2013.09.023

Baghdadi, N., Le Maire, G., Bailly, J.-S., Ose, K., Nouvellon, Y., Zribi, M., et al. (2015). Evaluation of ALOS/PALSAR L-band data for the estimation of *Eucalyptus* plantations aboveground biomass in Brazil. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 8 (8), 3802–3811. doi:10.1109/JSTARS.2014.2353661

Barenblitt, A., Fatoyinbo, L., Thomas, N., Stovall, A., De Sousa, C., Nwobi, C., et al. (2024). Invasion in the Niger delta: remote sensing of mangrove conversion to invasive *Nypa fruticans* from 2015 to 2020. *Remote Sens. Ecol. Conservation* 10 (1), 5–23. doi:10.1002/rse2.353

Barron, J. T. (2017). A general and adaptive robust loss function. *arXiv*. doi:10.48550/ARXIV.1701.03077

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Sci. Data* 5 (1), 180214. doi:10.1038/sdata.2018.214

Berninger, A., Lohberger, S., Stängel, M., and Siegert, F. (2018). SAR-based estimation of above-ground biomass and its changes in tropical forests of kalimantan using L- and C-band. *Remote Sens.* 10 (6), 831. doi:10.3390/rs10060831

Black, M. J., and Anandan, P. (1996). The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Comput. Vis. Image Underst.* 63 (1), 75–104. doi:10.1006/cviu.1996.0006

Bouvet, A., Mermoz, S., Ballère, M., Koleck, T., and Toan, T.Le (2018). Use of the SAR shadowing effect for deforestation detection with sentinel-1 time series. *Remote Sens.* 10 (8), 1250. doi:10.3390/rs10081250

Boyd, D. S., and Danson, F. M. (2005). Satellite remote sensing of forest resources: three decades of research development. *Prog. Phys. Geogr. Earth Environ.* 29 (1), 1–26. doi:10.1191/0309133305pp432ra

Catalogue des habitats forestiers de Guyane. (2015). Cayenne (Guyane): direction régionale ONF Guyane.

Charbonnier, P., Blanc-Feraud, L., Aubert, G., and Barlaud, M. (1994). "Two deterministic half-quadratic regularization algorithms for computed imaging," in Proceedings of 1st international Conference on image processing, *2* (Austin, TX, USA: IEEE Comput. Soc. Press), 168–172. doi:10.1109/ICIP.1994.413553

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., et al. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* 145 (1), 87–99. doi:10.1007/s00442-005-0100-x

Chen, M., Dong, W., Hao, Yu, Woodhouse, I. H., Ryan, C. M., Liu, H., et al. (2024). Multimodal deep learning enables forest height mapping from patchy spaceborne lidar using sar and passive optical satellite data. doi:10.2139/ssrn.4898106

Corbane, C., Politis, P., Kempeneers, P., Simonetti, D., Soille, P., Burger, A., et al. (2020). A global cloud free pixel- based image composite from sentinel-2 data. *Data Brief* 31 (August), 105737. doi:10.1016/j.dib.2020.105737

Cresson, R., and Saint-Geours, N. (2015). Natural color satellite image mosaicking using quadratic programming in decorrelated color Space. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 8 (8), 4151–4162. doi:10.1109/JSTARS.2015.2449233

De Thoisy, B., Richard-Hansen, B. C., Bertrand, G., Joubert, P., Obstancias, J., Winterton, P., et al. (2010). Rapid evaluation of threats to biodiversity: human footprint score and large vertebrate species responses in French Guiana. *Biodivers. Conservation* 19 (6), 1567–1584. doi:10.1007/s10531-010-9787-z

Dennis, J. E., and Welsch, R. E. (1978). Techniques for nonlinear least squares and robust regression. *Commun. Statistics - Simul. Comput.* 7 (4), 345–359. doi:10.1080/03610917808812083

Dorado-Roda, I., Pascual, A., Godinho, S., Silva, C., Botequim, B., Rodríguez-Gonzálvez, P., et al. (2021). Assessing the accuracy of GEDI data for canopy height and aboveground biomass estimates in mediterranean forests. *Remote Sens.* 13 (12), 2279. doi:10.3390/rs13122279

Dubayah, R., Bryan Blair, J., Scott, G., Fatoyinbo, L., Hansen, M., Healey, S., et al. (2020). The global ecosystem dynamics investigation: high-resolution laser ranging of the Earth's forests and topography. *Sci. Remote Sens.* 1 (June), 100002. doi:10.1016/j.srs.2020.100002

Dubayah, R., Hofton, M., Blair, J., Armston, J., Tang, H., and Luthcke, S. (2021). "GEDI L2A elevation and height metrics data global footprint level V002," in *NASA EOSDIS land processes distributed active archive center*. doi:10.5067/GEDI/GEDI02_A.002

Fagua, J. C., Jantz, P., Rodriguez-Buritica, S., Duncanson, L., and Goetz, S. J. (2019). Integrating LiDAR, multispectral and SAR data to estimate and map canopy height in tropical forests. *Remote Sens.* 11 (22), 2697. doi:10.3390/rs11222697

Fassnacht, F. E., Poblete-Olivares, J., Rivero, L., Lopatin, J., Ceballos-Comisso, A., and Galleguillos, M. (2021). Using sentinel-2 and canopy height models to derive a landscape-level biomass map covering multiple vegetation types. *Int. J. Appl. Earth Observation Geoinformation* 94 (February), 102236. doi:10.1016/j.jag.2020.102236

Fayad, I., Baghdadi, N., Alvares, C. A., Stape, J. L., Bailly, J. S., Scolforo, H. F., et al. (2021). Terrain slope effect on forest height and wood volume estimation from GEDI data. *Remote Sens.* 13 (11), 2136. doi:10.3390/rs13112136

Fayad, I., Baghdadi, N., Bailly, J.-S., Barbier, N., Gond, V., Hajj, M., et al. (2014). Canopy height estimation in French Guiana with LiDAR ICESat/GLAS data using principal component analysis and random forest regressions. *Remote Sens.* 6 (12), 11883–11914. doi:10.3390/rs61211883

Fayad, I., Baghdadi, N., Guitet, S., Bailly, J.-S., Hérault, B., Gond, V., et al. (2016). Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data. *Int. J. Appl. Earth Observation Geoinformation* 52 (October), 502–514. doi:10.1016/j.jag.2016.07.015

Fayad, I., Baghdadi, N., and Lahssini, K. (2022). An assessment of the GEDI lasers' capabilities in detecting canopy tops and their penetration in a densely vegetated, tropical area. *Remote Sens.* 14 (13), 2969. doi:10.3390/rs14132969

Fayad, I., Ciais, P., Schwartz, M., Wigneron, J.-P., Baghdadi, N., De Truchis, A., et al. (2024). Hy-TeC: a hybrid vision transformer model for high-resolution and large-scale mapping of canopy height. *Remote Sens. Environ.* 302 (March), 113945. doi:10.1016/j.rse.2023.113945

Fayad, I., Ienco, D., Baghdadi, N., Gaetano, R., Alvares, C. A., Stape, J. L., et al. (2021). A CNN-based approach for the estimation of canopy heights and wood volume from GEDI waveforms. *Remote Sens. Environ.* 265 (November), 112652. doi:10.1016/j.rse.2021.112652

Feldpausch, T. R., Lloyd, J., Lewis, S. L., Brienen, R. J. W., Gloor, M., Monteagudo Mendoza, A., et al. (2012). Tree height integrated into pantropical forest biomass estimates. *Biogeosciences* 9 (8), 3381–3403. doi:10.5194/bg-9-3381-2012

Ferry, B., Morneau, F., Bontemps, J.-D., Blanc, L., and Vincent, F. (2010). Higher treefall rates on slopes and waterlogged soils result in lower stand biomass and productivity in a tropical rain forest. *J. Ecol.* 98 (1), 106–116. doi:10.1111/j.1365-2745.2009.01604.x

García, M., Saatchi, S., Ustin, S., and Balzter, H. (2018). Modelling forest canopy height by integrating airborne LiDAR samples with satellite radar and multispectral imagery. *Int. J. Appl. Earth Observation Geoinformation* 66 (April), 159–173. doi:10.1016/j.jag.2017.11.017

Gazzea, M., Solheim, A., and Arghandeh, R. (2023). High-resolution mapping of forest structure from integrated SAR and optical images using an enhanced U-net method. *Sci. Remote Sens.* 8 (December), 100093. doi:10.1016/j.srs.2023.100093

Ge, S., Gu, H., Su, W., Praks, J., and Antropov, O. (2022). Improved semisupervised UNet deep learning model for forest height mapping with satellite SAR and optical data. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 15, 5776–5787. doi:10.1109/JSTARS.2022.3188201

Geman, S., and McClure, D. E. (1985). "Bayesian image analysis: an application to single photon emission tomography," in *Proceedings of the American statistical association*, 12–18.

Gond, V., Vincent, F., Molino, J.-F., Brunaux, O., Ingrassia, F., Joubert, P., et al. (2011). Broad-scale spatial pattern of forest landscape types in the Guiana shield. *Int. J. Appl. Earth Observation Geoinformation* 13 (3), 357–367. doi:10.1016/j.jag.2011.01.004

Grabska, E., Hostert, P., Pflugmacher, D., and Ostapowicz, K. (2019). Forest stand species mapping using the sentinel-2 time series. *Remote Sens.* 11 (10), 1197. doi:10.3390/rs11101197

Guitet, S., Cornu, J.-F., Brunaux, O., Betbeder, J., Carozza, J.-M., and Richard-Hansen, C. (2013). Landform and landscape mapping, French Guiana (South America). *J. Maps* 9 (3), 325–335. doi:10.1080/17445647.2013.785371

Guitet, S., Pélissier, R., Brunaux, O., Jaouen, G., and Sabatier, D. (2015). Geomorphological landscape features explain floristic patterns in French Guiana rainforest. *Biodivers. Conservation* 24 (5), 1215–1237. doi:10.1007/s10531-014-0854-8

Hancock, S., Armston, J., Hofton, M., Sun, X., Tang, H., Duncanson, L. I., et al. (2019). The GEDI simulator: a large-footprint waveform lidar simulator for calibration and validation of spaceborne missions. *Earth Space Sci.* 6 (2), 294–310. doi:10.1029/2018EA000506

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science* 342 (6160), 850–853. doi:10.1126/science.1244693

Healey, S. P., Yang, Z., Gorelick, N., and Simon, I. (2020). Highly local model calibration with a new GEDI LiDAR asset on Google Earth engine reduces Landsat forest height signal saturation. *Remote Sens.* 12 (17), 2840. doi:10.3390/rs12172840

Hilbert, C., and Schmullius, C. (2012). Influence of surface topography on ICESat/GLAS forest height estimation and waveform shape. *Remote Sens.* 4 (8), 2210–2235. doi:10.3390/rs4082210

Ho Tong Ming, D., Dinh, T.Le T., Rocca, F., Tebaldini, S., Villard, L., Réjou-Méchain, M., et al. (2016). SAR tomography for the retrieval of forest biomass and height: cross-validation at two tropical forest sites in French Guiana. *Remote Sens. Environ.* 175 (March), 138–147. doi:10.1016/j.rse.2015.12.037

Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., et al. (2021). More diverse means better: multimodal deep learning meets remote-sensing imagery classification. *IEEE Trans. Geoscience Remote Sens.* 59 (5), 4340–4354. doi:10.1109/TGRS.2020.3016820

Ienco, D., Gaetano, R., Dupaquier, C., and Maurel, P. (2017). Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geoscience Remote Sens. Lett.* 14 (10), 1685–1689. doi:10.1109/LGRS.2017.2728698

Jin, S., Su, Y., Gao, S., Hu, T., Liu, J., and Guo, Q. (2018). The transferability of random forest in canopy height estimation from multi-source remote sensing data. *Remote Sens.* 10 (8), 1183. doi:10.3390/rs10081183

Joetzjer, E., Pillet, M., Ciais, P., Barbier, N., Chave, J., Schlund, M., et al. (2017). Assimilating satellite-based canopy height within an ecosystem model to estimate aboveground forest biomass. *Geophys. Res. Lett.* 44 (13), 6823–6832. doi:10.1002/2017GL074150

Jucker, T., Bongalov, B., Burslem, D. F. R. P., Nilus, R., Dalponte, M., Lewis, S. L., et al. (2018). "Topography shapes the structure, composition and function of tropical forest landscapes,". Editor M. Uriarte, 21, 989–1000. doi:10.1111/ele.12964*Ecol. Lett.*7

Karasiak, N., Sheeren, D., Fauvel, M., Willm, J., Dejoux, J.-F., and Monteil, C. (2017). "Mapping tree species of forests in southwest France using sentinel-2 image time series," in *2017 9th international workshop on the analysis of multitemporal remote sensing images (MultiTemp)* (Brugge, Belgium: IEEE), 1–4. doi:10.1109/Multi-Temp.2017.8035215

Khati, U., Singh, G., and Kumar, S. (2018). Potential of space-borne PolInSAR for forest canopy height estimation over India—a case study using fully polarimetric *L* -*C* - and *X* -band SAR data. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 11 (7), 2406–2416. doi:10.1109/JSTARS.2018.2835388

Kutchartt, E., Pedron, M., and Pirotti, F. (2022). Assessment of canopy and ground height accuracy from gedi lidar over steep mountain areas. *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.* V-3–2022, 431–438. doi:10.5194/isprs-annals-V-3-2022-431-2022

Lahssini, K., Baghdadi, N., Le Maire, G., Dupuy, S., and Fayad, I. (2024). Use of GEDI signal and environmental parameters to improve canopy height estimation over tropical forest ecosystems in Mayotte island. *Can. J. Remote Sens.* 50 (1), 2351004. doi:10.1080/07038992.2024.2351004

Lahssini, K., Baghdadi, N., Le Maire, G., and Fayad, I. (2022). Influence of GEDI acquisition and processing parameters on canopy height estimates over tropical forests. *Remote Sens.* 14 (24), 6264. doi:10.3390/rs14246264

Lahssini, K., Teste, F., Dayal, K. R., Durrieu, S., Ienco, D., and Monnet, J.-M. (2022). Combining LiDAR metrics and sentinel-2 imagery to estimate basal area and wood volume in complex forest environment via neural networks. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 15, 4337–4348. doi:10.1109/JSTARS.2022.3175609

Lang, N., Jetz, W., Schindler, K., and Jan, D. W. (2023). A high-resolution canopy height model of the Earth. *Nat. Ecol. and Evol.* 7 (11), 1778–1789. doi:10.1038/s41559-023-02206-6

Lang, N., Kalischek, N., Armston, J., Schindler, K., Dubayah, R., and Wegner, J. D. (2022). Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. *Remote Sens. Environ.* 268 (January), 112760. doi:10.1016/j.rse.2021.112760

Lang, N., Schindler, K., and Jan, D. W. (2019). Country-wide high-resolution vegetation height mapping with sentinel-2. *Remote Sens. Environ.* 233 (November), 111347. doi:10.1016/j.rse.2019.111347

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521 (7553), 436–444. doi:10.1038/nature14539

Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Espirito-Santo, F. D. B., et al. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophys. Res. Lett.* 32 (22), 2005GL023971. doi:10.1029/2005GL023971

Lemaire, G., Francois, C., Soudani, K., Berveiller, D., Pontailler, J., Breda, N., et al. (2008). Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sens. Environ.* 112 (10), 3846–3864. doi:10.1016/j.rse.2008.06.005

Li, H., Li, X., Kato, T., Hayashi, M., Fu, J., and Hiroshima, T. (2024). Accuracy assessment of GEDI terrain elevation, canopy height, and aboveground biomass density estimates in Japanese artificial forests. *Sci. Remote Sens.* 10 (December), 100144. doi:10.1016/j.srs.2024.100144

Lima, A. J. N., Suwa, R., Gabriel Henrique Pires, De M. R., Kajimoto, T., Dos Santos, J., Pereira Da Silva, R., et al. (2012). Allometric models for estimating above- and below-ground biomass in amazonian forests at são gabriel da cachoeira in the upper rio negro, Brazil. *For. Ecol. Manag.* 277 (August), 163–172. doi:10.1016/j.foreco.2012.04.028

Liu, A., Cheng, X., and Chen, Z. (2021). Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sens. Environ.* 264 (October), 112571. doi:10.1016/j.rse.2021.112571

Liu, X., Neigh, C. S. R., Pardini, M., and Forkel, M. (2024). Estimating forest height and above-ground biomass in tropical forests using P-band TomoSAR and GEDI observations. *Int. J. Remote Sens.* 45 (9), 3129–3148. doi:10.1080/01431161.2024.2343134

Long, J., Shelhamer, E., and Darrell, T. (2014). Fully convolutional networks for semantic segmentation. *arXiv*. doi:10.48550/ARXIV.1411.4038

Main-Knorn, M., Pflug, B., Louis, J., Vincent, D., Müller-Wilm, U., and Gascon, F. (2017). "Sen2Cor for sentinel-2,". *Image and signal processing for remote sensing XXIII*. Editors L. Bruzzone, F. Bovolo, and J. A. Benediktsson (Warsaw, Poland: SPIE), 3. doi:10.1117/12.2278218

McDowell, N., Barnard, H., Bond, B., Hinckley, T., Hubbard, R., Ishii, H., et al. (2002). The relationship between tree height and leaf area: sapwood area ratio. *Oecologia* 132 (1), 12–20. doi:10.1007/s00442-002-0904-x

Mermoz, S., Le Toan, T., Villard, L., Réjou-Méchain, M., and Seifert-Granzin, J. (2014). Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sens. Environ.* 155 (December), 109–119. doi:10.1016/j.rse.2014.01.029

Mermoz, S., Réjou-Méchain, M., Villard, L., Le Toan, T., Rossi, V., and Gourlet-Fleury, S. (2015). Decrease of L-band SAR backscatter with biomass of dense forests. *Remote Sens. Environ.* 159 (March), 307–317. doi:10.1016/j.rse.2014.12.019

Milesi, A. (2022). "U-net: semantic segmentation with PyTorch."

Morin, D., Planells, M., Baghdadi, N., Bouvet, A., Fayad, I., Le Toan, T., et al. (2022). Improving heterogeneous forest height maps by integrating GEDI-based forest height information in a multi-sensor mapping process. *Remote Sens.* 14 (9), 2079. doi:10.3390/rs14092079

Morin, D., Planells, M., Guyon, D., Villard, L., Mermoz, S., Bouvet, A., et al. (2019). Estimation and mapping of forest structure parameters from open access satellite images: development of a generic method with a study case on coniferous plantation. *Remote Sens.* 11 (11), 1275. doi:10.3390/rs11111275

Morin, D., Planells, M., Mermoz, S., and Mouret, F. (2023). Estimation of forest height and biomass from open-access multi-sensor satellite imagery and GEDI lidar data: high-resolution maps of metropolitan France. *arXiv*. doi:10.48550/ARXIV.2310.14662

Muscarella, R., Kolyaie, S., Morton, D. C., Zimmerman, J. K., and Uriarte, M. (2020). "Effects of topography on tropical forest structure depend on climate context,". *Journal of ecology*. Editor T. Jucker, 108, 145–159. doi:10.1111/1365-2745.132611

Ngo, Y.-N., Minh, D.Ho T., Baghdadi, N., and Fayad, I. (2023). Tropical forest top height by GEDI: from sparse coverage to continuous data. *Remote Sens.* 15 (4), 975. doi:10.3390/rs15040975

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., et al. (2011). A large and persistent carbon sink in the world's forests. *Science* 333 (6045), 988–993. doi:10.1126/science.1201609

Pimple, U., Leadprathom, K., Simonetti, D., Sitthi, A., Peters, R., Pungkul, S., et al. (2022). Assessing mangrove species diversity, zonation and functional indicators in response to natural, regenerated, and rehabilitated succession. *J. Environ. Manag.* 318 (September), 115507. doi:10.1016/j.jenvman.2022.115507

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., et al. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.* 253 (February), 112165. doi:10.1016/j.rse.2020.112165

Pourrahmati, M. R., Baghdadi, N. N., Darvishsefat, A. A., Namiranian, M., Fayad, I., Bailly, J.-S., et al. (2015). Capability of GLAS/ICESat data to estimate forest canopy height and volume in mountainous forests of Iran. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 8 (11), 5246–5261. doi:10.1109/JSTARS.2015.2478478

Qi, W., Armston, J., Choi, C., Stovall, A., Saarela, S., Pardini, M., et al. (2023). "Mapping large-scale pantropical forest canopy height by integrating GEDI lidar and TanDEM-X InSAR data. doi:10.21203/rs.3.rs-3306982/v1

Quegan, S., Le Toan, T., Chave, J., Dall, J., Exbrayat, J.-F., Minh, D.Ho T., et al. (2019). The European Space agency BIOMASS mission: measuring forest above-ground biomass from Space. *Remote Sens. Environ.* 227 (June), 44–60. doi:10.1016/j.rse.2019.03.032

Rahman, M. M., and Sumantyo, J. T. S. (2010). Mapping tropical forest cover and deforestation using synthetic aperture radar (SAR) images. *Appl. Geomatics* 2 (3), 113–121. doi:10.1007/s12518-010-0026-9

Rennó, C. D., Donato Nobre, A., Adriana Cuartas, L., Soares, J. V., Hodnett, M. G., Tomasella, J., et al. (2008). HAND, a new terrain descriptor using SRTM-DEM: mapping terra-firme rainforest environments in Amazonia. *Remote Sens. Environ.* 112 (9), 3469–3481. doi:10.1016/j.rse.2008.03.018

Robbins, H., and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statistics* 22 (3), 400–407. doi:10.1214/aoms/1177729586

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. *arXiv*, 234–241. doi:10.1007/978-3-319-24574-4_28

Roy, D. P., Kashongwe, H. B., and Armston, J. (2021). The impact of geolocation uncertainty on GEDI tropical forest canopy height estimation and change monitoring. *Sci. Remote Sens.* 4 (December), 100024. doi:10.1016/j.srs.2021.100024

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl. Acad. Sci.* 108 (24), 9899–9904. doi:10.1073/pnas.1019576108

Schietti, J., Emilio, T., Rennó, C. D., Drucker, D. P., Costa, F. R. C., Nogueira, A., et al. (2014). Vertical distance from drainage drives floristic composition changes in an amazonian rainforest. *Plant Ecol. and Divers.* 7 (1–2), 241–253. doi:10.1080/17550874.2013.783642

Schleich, A., Durrieu, S., Soma, M., and Vega, C. (2023). Improving GEDI footprint geolocation using a high-resolution digital elevation model. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 16, 7718–7732. doi:10.1109/JSTARS.2023.3298991

Schlund, M., Wenzel, A., Camarretta, N., Stiegler, C., and Erasmi, S. (2023). Vegetation canopy height estimation in dynamic tropical landscapes with TanDEM-X supported by GEDI data. *Methods Ecol. Evol.* 14 (7), 1639–1656. doi:10.1111/2041-210X.13933

Schneider, F. D., António, F., Hancock, S., Duncanson, L. I., Dubayah, R. O., Pavlick, R. P., et al. (2020). Towards mapping the diversity of canopy structure from Space with GEDI. *Environ. Res. Lett.* 15 (11), 115006. doi:10.1088/1748-9326/ab9e99

Schwartz, M., Ciais, P., Ottlé, C., De Truchis, A., Vega, C., Fayad, I., et al. (2024). High-resolution canopy height map in the Landes forest (France) based on GEDI, sentinel-1, and sentinel-2 data with a deep learning approach. *Int. J. Appl. Earth Observation Geoinformation* 128 (April), 103711. doi:10.1016/j.jag.2024.103711

Shannon, E. S., Finley, A. O., Hayes, D. J., Noralez, S. N., Weiskittel, A. R., Cook, B. D., et al. (2024). Quantifying and correcting geolocation error in spaceborne LiDAR forest canopy observations using high spatial accuracy data: a bayesian model approach. *Environmetrics* 35 (4), e2840. doi:10.1002/env.2840

Simard, M., Pinto, N., Fisher, J. B., and Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res.* 116 (G4), G04021. doi:10.1029/2011JG001708

Simonetti, D., Pimple, U., Langner, A., and Marelli, A. (2021). Pan-tropical sentinel-2 cloud-free annual composite datasets. *Data Brief* 39 (December), 107488. doi:10.1016/j.dib.2021.107488

Smith, L. N. (2015). Cyclical learning rates for training neural networks. *arXiv*. doi:10.48550/ARXIV.1506.01186

Sothe, C., Gonsamo, A., Lourenço, R. B., Kurz, W. A., and Snider, J. (2022). Spatially continuous mapping of forest canopy height in Canada by combining GEDI and ICESat-2 with PALSAR and Sentinel. *Remote Sens.* 14 (20), 5158. doi:10.3390/rs14205158

Stape, J. L., and Alvares, C. A. (2023). Physiographic position drives Eucalyptus productivity in mato grosso do sul, Brazil. *Série Técnica IPEF* 26 (48), 289–293. doi:10.18671/sertec.v26n48.057

Sun, D., Roth, S., and Black, M. J. (2010). "Secrets of optical flow estimation and their principles," in *2010 IEEE computer society conference on computer vision and pattern recognition* (San Francisco, CA, USA: IEEE), 2432–2439. doi:10.1109/CVPR.2010.5539939

Sun, M., Cui, L., Park, J., García, M., Zhou, Y., Silva, C. A., et al. (2022). Evaluation of NASA's GEDI lidar observations for estimating biomass in temperate and tropical forests. *Forests* 13 (10), 1686. doi:10.3390/f13101686

Tang, H., Stoker, J., Luthcke, S., Armston, J., Lee, K., Bryan, B., et al. (2023). Evaluating and mitigating the impact of systematic geolocation error on canopy height measurement performance of GEDI. *Remote Sens. Environ.* 291 (June), 113571. doi:10.1016/j.rse.2023.113571

Tolan, J., Yang, H.-I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., et al. (2024). Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sens. Environ.* 300 (January), 113888. doi:10.1016/j.rse.2023.113888

Tyukavina, A., Baccini, A., Hansen, M. C., Potapov, P. V., Stehman, S. V., Houghton, R. A., et al. (2015). Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012. *Environ. Res. Lett.* 10 (7), 074002. doi:10.1088/1748-9326/10/7/074002

Ulaby, F. T., Sarabandi, K., McDONALD, K., Whitt, M., and Craig Dobson, M. (1990). Michigan microwave canopy scattering model. *Int. J. Remote Sens.* 11 (7), 1223–1253. doi:10.1080/01431169008955090

Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., et al. (2021). Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. *Sci. Adv.* 7 (10), eabe1603. doi:10.1126/sciadv.abe1603

Verhegghen, A., Eva, H., Ceccherini, G., Achard, F., Gond, V., Gourlet-Fleury, S., et al. (2016). The potential of Sentinel satellites for burnt area mapping and monitoring in the Congo basin forests. *Remote Sens.* 8 (12), 986. doi:10.3390/rs8120986

Wagner, F. H., Roberts, S., Ritz, A. L., Carter, G., Dalagnol, R., Favrichon, S., et al. (2024). Sub-meter tree height mapping of California using aerial images and LiDAR-

informed U-net model. *Remote Sens. Environ.* 305 (May), 114099. doi:10.1016/j.rse.2024.114099

Wang, C., Elmore, A. J., Numata, I., Cochrane, M. A., Lei, S., Hakkenberg, C. R., et al. (2022). A framework for improving wall-to-wall canopy height mapping by integrating GEDI LiDAR. *Remote Sens.* 14 (15), 3618. doi:10.3390/rs14153618

Watanabe, M., Koyama, C. N., Hayashi, M., Nagatani, I., and Shimada, M. (2018). Early-stage deforestation detection in the tropics with *L* -band SAR. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 11 (6), 2127–2133. doi:10.1109/JSTARS.2018.2810857

Wijaya, A., Liesenberg, V., Susanti, A., Karyanto, O., and Verchot, L. V. (2015). Estimation of biomass carbon stocks over peat swamp forests using multi-temporal and multi-polratizations SAR data. *Int. Archives Photogrammetry, Remote Sens. Spatial Inf. Sci. XL-7/W3 (April)*, 551–556. doi:10.5194/isprsarchives-XL-7-W3-551-2015

Yang, Y., Saatchi, S., Xu, L., Yu, Y., Lefsky, M., White, L., et al. (2016). Abiotic controls on macroscale variations of humid tropical forest height. *Remote Sens.* 8 (6), 494. doi:10.3390/rs8060494

Yuan, X., Shi, J., and Gu, L. (2021). A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* 169 (May), 114417. doi:10.1016/j.eswa.2020.114417

Zhao, Z., Jiang, B., Wang, H., and Wang, C. (2024). Forest canopy height retrieval model based on a dual attention mechanism deep network. *Forests* 15 (7), 1132. doi:10.3390/f15071132

Chapter 5

Addressing GEDI-Induced Anisotropies in Residual Kriging to Improve Multi-Source Canopy Height Estimates in French Guiana

Using the high-resolution canopy height map for French Guiana presented in the previous study, this chapter addresses the fourth and final research question of this dissertation: how do the acquisition parameters of GEDI data influence the results of canopy height mapping? With the availability of a comprehensive canopy height map, the objectives are to examine the impact of GEDI's acquisition features on the canopy height estimates and to explore methods to enhance these results.

In this study, two different canopy height maps built from distinct modeling approaches and with the same input data are used as basis for the analysis: the U-Net architecture presented previously (*CHNET*) and a Random Forest algorithm (*RFH*). The U-Net method is designed to handle sparse reference data and incorporates spatial information in its design which is why it was chosen as the main framework to produce a canopy height map. In contrast, the Random Forest model is a non-spatial method that treats each data sample independently without accounting for spatial relationships. By comparing these two methodologies, the study assesses the significance of incorporating spatial autocorrelation in the mapping process of a continuous variable like canopy height. The final goal is to enhance predictions through residual kriging by integrating spatial components in the modeling process.

The spatial analysis of the residuals from both models reveals the presence of anisotropies that do not represent the actual canopy horizontal structure and dynamics. These anisotropies are explained by the GEDI sensor's acquisition characteristics, including variations in laser power and the regular spatial pattern of GEDI's ground tracks. Differences is laser penetration of the vegetation caused by varying power levels result in inconsistent measurements over the study area. These effects coupled with the spatial pattern of GEDI's acquisition (regularly

spaced beam ground transects) introduce systematic periodic anisotropies in the measurements. These sensor-induced artifacts need to be addressed to achieve more reliable canopy height estimates.

To refine the canopy height predictions, a residual kriging procedure is integrated into the modeling process. In the light of the preliminary spatial analysis of the residuals conducted, the spatial information is incorporated in a way that addresses and mitigates the sensor-induced anisotropies identified earlier. This is achieved by analyzing subsets of the original GEDI database to produce an accurate and faithful representation of the actual spatial autocorrelation. This spatial structure computed under conditions that eliminate sources of sensor-induced anisotropies is considered a reliable representation of the canopy's spatial variability in an isotropic manner. The results demonstrate that integrating spatial components in the modeling process through residual kriging improves the accuracy of canopy height estimations.

In summary, this chapter presents a framework to address sensor-induced anisotropies in GEDI measurements through the implementation of residual kriging. The comparative analysis of *CHNET* and *RFH* highlights the importance of accounting for spatial autocorrelation when mapping a continuous variable like canopy height whose dynamics are driven by many complex factors. As underlined in the other chapters, considering GEDI specific characteristics is key to improve the reliability of canopy height estimates. This study refines the original methodological framework presented in the previous chapter by integrating additional spatial information in an optimal way.

Addressing GEDI-Induced Anisotropies in Residual Kriging to Improve Multi-Source Canopy Height Estimates in French Guiana

Kamel Lahssini, Guerric le Maire, Nicolas Baghdadi, and Ibrahim Fayad

***Abstract*—Quantifying aboveground biomass (AGB) is essential in the context of global climate change. Canopy height, which is related to AGB, can be mapped using machine learning models trained with multi-source spatial data and Global Ecosystem Dynamics Investigation (GEDI) measurements. In this study, a comparative analysis of canopy height estimates derived from two models is presented: a U-Net deep learning model (*CHNET*) and a Random Forest algorithm (*RFH*). Both models were trained using GEDI LiDAR data and utilized multi-source inputs, including optical, radar, and environmental data. By conducting a spatial analysis of the models' residuals, we identified that GEDI data acquisition parameters, particularly the variability in laser beam energy combined with the azimuthal directions of the observation tracks, introduce spatial inconsistencies in the semivariograms in the form of periodic patterns. To address these anisotropies, we considered exclusively GEDI power beams, and we conducted our spatial autocorrelation analysis in the GEDI track azimuthal direction. Next, we employed the residual kriging (RK) spatial interpolation technique to account for the spatial autocorrelation of canopy heights and improve the accuracies of *CHNET* and *RFH* estimates. Adding RK corrections improved the performance of both *CHNET* and *RFH*, with more substantial gains observed for *RFH*. The corrections appeared to be localized around the GEDI sample points and the density of usable GEDI information is therefore an important factor in the effectiveness of RK. Furthermore, our findings reveal that a Random Forest model combined with spatial interpolation can deliver performance comparable to that of a U-Net model alone.**

***Index Terms*—Canopy height, deep learning, French Guiana, GEDI, LiDAR, Random Forest, residual kriging, spatial interpolation, tropical forest, U-Net.**

I. Introduction

TROPICAL forests are among the most important natural ecosystems on Earth as they play a crucial role in regulating the global climate. These forests are not only biodiversity centers but also act as significant carbon sinks, absorbing large amounts of carbon dioxide from the atmosphere [1], [2], [3]. The preservation of the standing aboveground biomass (AGB) within these ecosystems is key in the context of global warming and climate change. By storing carbon, tropical forests mitigate the greenhouse effect and thus help to stabilize global temperatures. However, the ongoing deforestation and degradation of these forests threaten their ability to function as carbon sinks, making it imperative to monitor accurately AGB levels.

To effectively quantify AGB, studies commonly rely on allometric equations that link the structural characteristics of a forest, such as tree height, diameter at breast height, and wood density, to its biomass [4], [5]. These relationships can be applied either at the individual tree level or at stand level. Precise estimates of AGB can be obtained from in situ measurements but they are limited to small and accessible areas. Canopy height is of paramount importance in AGB estimation models. This variable can be estimated using remote sensing techniques. Accurate canopy height measurements are essential for deriving reliable AGB estimates through allometric models that rely on canopy height. The United Nations' REDD+ program emphasizes the importance of precision in AGB estimates, recommending a relative error of 20% [6], [7]. Therefore, improving the accuracies of canopy height estimates is needed to meet these international standards and to ensure the effectiveness of global forest conservation efforts.

In this context, Light Detection and Ranging (LiDAR) systems have emerged as powerful tools for characterizing vegetation profiles and structural parameters. Airborne LiDAR scanning (ALS) offers higher spatial resolution than spaceborne systems. It provides more precise measurements of vegetation structure over smaller scales. These characteristics represent an intermediary option between detailed ground-based data and the broad coverage of satellite observations. Spaceborne LiDAR sensors offer the advantages of frequent revisit times and global coverage, which makes them an invaluable source of data for large-scale and continuous monitoring of forest ecosystems. These sensors have demonstrated their capability to generate regional and global forest products, and to effectively map canopy height [8], [9]. However, challenges such as the low density of information due to sparse and unevenly distributed footprints, as well as their high operational altitudes that increase sensitivity to environmental conditions, limit their effectiveness [10], [11], [12]. The Global Ecosystem Dynamics Investigation (GEDI), a system specifically designed to measure vegetation structure on a global scale, represents the latest advancement in spaceborne LiDAR technology [13]. As a full-waveform

Kamel Lahssini and Nicolas Baghdadi are with INRAE, UMR TETIS, Université de Montpellier, 34000 Montpellier, France (e-mail: kamel.lahssini@inrae.fr; nicolas.baghdadi@inrae.fr).

Guerric le Maire is with CIRAD, UMR Eco&Sols, Université de Montpellier, Institut Agro, IRD, INRAE, 34000 Montpellier, France (e-mail: guerric.le_maire@cirad.fr).

Ibrahim Fayad is with IPSL, LSCE, Université Paris-Saclay, 91191 Gif-sur-Yvette, France, and with Kayrros SAS, 75009 Paris, France (e-mail: ibrahim.fayad@lsce.ipsl.fr).

system, GEDI captures waveforms that directly represent the vegetation's vertical structure. From these raw waveforms, various descriptive metrics can be extracted to characterize canopy height. The simplest and most straightforward method is to use a single metric for direct canopy height estimation. More accurate results can also be achieved by building regression-based models from multiple GEDI metrics and parameters [14].

Since canopy height estimation is only performed at GEDI footprint locations given the sparse nature of GEDI information, it is necessary to combine these point-based estimates with continuous bidimensional remote sensing data to generate comprehensive and continuous canopy height maps. GEDI data are used as reference canopy heights to train and validate prediction models that usually integrate optical and radar information. In this perspective, statistical and machine learning models, such as stepwise regression or Random Forest, have been widely employed for accurate canopy height mapping at various scales [15], [16]. Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have significantly improved the integration of complementary data sources [17]. In the context of canopy height mapping from multiple data sources, CNNs are able to integrate spatial and textural information at both local and global scales [18], [19], [20]. Consequently, convolutional architectures have been effective in combining GEDI metrics with optical and radar images to produce extensive and continuous canopy height maps [21], [22]. The incorporation of additional ancillary environmental parameters related to canopy structure can also further enhance the accuracy of canopy height estimates [23].

Estimating forest parameters and especially canopy height from multiple remote sensing data is challenging in tropical biomes, which are characterized by dense vegetation and tall canopies. A major issue affecting canopy height estimates in these regions is the saturation of sensors in areas with high AGB levels [24]. For GEDI, a key factor impacting height estimates is the sensor's ability to penetrate through dense vegetation to accurately capture the whole vertical structure from the canopy top to the ground [25]. Multispectral satellite data also face limitations, as the reflectance of tropical rainforest canopies tends to saturate beyond a certain height, making it difficult to distinguish vegetation heights above that threshold [26], [27]. In the case of radar information, even L-band signals experience difficulties penetrating canopies when AGB levels exceed 150 Mg/ha [28]. Moreover, canopy height is a continuous variable influenced by various environmental factors, and traditional approaches such as linear regressions or Random Forest algorithms often lack the inherent incorporation of spatial correlation in their design [29]. As a result, these methods may not fully explain the spatial autocorrelation of canopy heights and leave some unexplained variance which could be linked to forest endogenous processes. Regarding CNNs, despite being able to extract spatial and textural features, the question of their ability to fully exploit the spatial information contained in raw input data sources and reference data is still an open research question.

In this regard, spatial interpolation techniques can be employed alongside canopy height regression models to take into account the spatial autocorrelation of the data and refine canopy height estimates. In particular, residual kriging (RK) is a geostatistical technique that relies on the intrinsic stochastic properties of a dataset [30]. Contrary to deterministic interpolation techniques, which use mathematical functions to calculate the values at unsampled locations based on the degree of similarity with respect to known points at sampled locations, RK uses both analytical and statistical methods to predict unknown values based on the spatial autocorrelation of the data [31]. RK is widely employed in climatic, water, and soil applications [30], [31], [32], yet its use in forestry and particularly for estimating canopy height remains relatively rare [33]. By integrating RK with canopy height regression models, it is possible to better capture the spatial patterns and variations in canopy height. In the context of canopy height mapping with sparse GEDI reference data, RK consists in an ordinary kriging (OK) procedure applied to the residuals of GEDI-based regression models. The resulting interpolated residuals, known as kriged residuals, can then be added to the regression results. This technique allows for the adjustment of canopy height estimates by accounting for the spatial structure of the residual errors that are not explained by the regression model. To accurately describe this spatial structure, it is essential to consider the specificities of the GEDI sensor, as its acquisition process introduces spatial anisotropies that may not represent the true spatial autocorrelation of canopy heights [14].

In this study, we conduct a detailed spatial analysis of two canopy height maps of French Guiana at a 10-m resolution, generated from the same input sources (optical, radar, and environmental parameters) and trained on GEDI reference data using two different methodologies: *CHNET* (Canopy Height estimation NETwork), a convolutional U-Net architecture, and *RFH* (Random Forest for canopy Height estimation), a more classical Random Forest algorithm. By comparing these two frameworks, we aim at understanding how their specific designs influence their ability to account for the spatial autocorrelation of canopy heights, and thus, the accuracy of their estimates. A spatial analysis of model residuals is conducted to identify and mitigate GEDI sensor-induced anisotropies observed during the spatial model calibration. Furthermore, using the insights gained from this preliminary spatial analysis, we implement a RK spatial interpolation technique to refine and enhance the results of both models. Our main goal is to determine whether incorporating a spatial residual autocorrelation component to canopy height modeling can be relevant and improve the accuracies of the estimates, which are tested against an independent ALS dataset. More precisely, we aim to test the hypothesis that the *CHNET* model may already incorporate some of the spatial correlations present in GEDI measurements, potentially limiting the improvement achieved through RK. Conversely, we hypothesize that *RFH* does not account for the spatial correlation in GEDI data, and thus could benefit more significantly from RK. By validating these hypotheses, we can explore if some of the unexplained variance in the initial regression models could be attributed to spatial correlations.

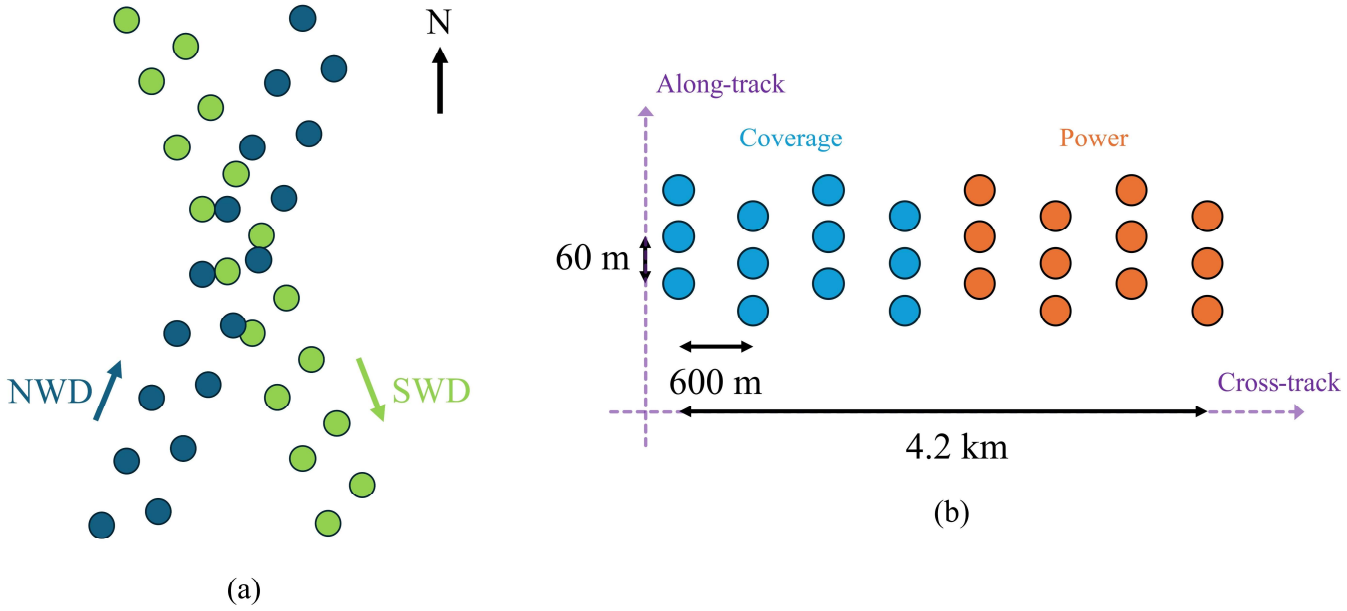


Fig. 1. GEDI acquisition track azimuthal directions NWD and SWD (a). GEDI ground sampling pattern (b).

II. Materials and Methods

A. Study Area

French Guiana, an overseas territory of France, is situated within the Amazon biome on the northern coast of South America. It covers an area of 83,534 km², with more than 80,000 km² of forests [34]. The primary forest type is mature old-growth tropical rainforest, while some areas contain secondary forests [35]. The coastal zones also include savannas and mangroves, but rainforest covers over 90% of the territory. Timber extraction and agricultural activities are largely concentrated in the sub-coastal regions close to major towns and along the main roads [36]. The region's terrain is mostly flat, with ground elevations rarely surpassing 200 m, though some small hills and mountains can be found in the landscape [35]. Approximately 70% of the slopes are less than 5° [34]. French Guiana has a hot, tropical climate, classified as tropical rainforest (Af) under the Köppen climate classification [37]. The average annual temperature is around 26°C, while rainfall varies significantly, reaching up to 4,000 mm per year in the northeast and about 2,000 mm in the southern and western areas [36]. The combination of climatic, geological, and geomorphological factors creates favorable conditions for the development of diverse forest structures across the region. In French Guiana's tropical forests, canopy heights typically range from 20 to 40 m, with some emergent trees reaching as tall as 60 m [8]. AGB varies widely, generally between 150 Mg/ha and over 600 Mg/ha, with higher values found in mature and undisturbed forests [34].

B. Datasets

1) GEDI Reference Data: The Global Ecosystem Dynamics Investigation (GEDI) instrument is a spaceborne LiDAR system mounted on the International Space Station (ISS), specifically designed to characterize the structure and dynamics of forest ecosystems. It is a joint mission between NASA and the University of Maryland that has acquired and processed data for the period between March 2019 and March 2023. The system utilizes three 1064 nm lasers, which emit 242 pulses per second, to produce energy return waveforms (L1B product) and waveform-derived height metrics (L2A product) within circular footprints of 25 meters in diameter. Consequently, the data is sparsely and unevenly distributed, with GEDI covering only about 4% of Earth's surface [13].

The ISS orbits Earth at a Low Earth Orbit (LEO), at an altitude of approximately 400 km and an inclination of 51.6° relative to the equator. This inclination allows GEDI to cover latitudes between approximately 51.6°N and 51.6°S. The ISS follows a near-circular prograde orbit, meaning it travels in the same direction as Earth's rotation (from the west to the east), completing one full revolution around Earth every 90 minutes. Due to its inclined orbit, the ISS's projected ground track exhibits a sinusoidal shape and therefore the subsequent GEDI's ground tracks show variability in azimuthal direction. Specifically, in regions near the equator, such as French Guiana, the ground tracks are distributed in two main azimuthal directions (Fig. 1a). In the context of this study, these two configurations are referred to as northward pass (NWD) and southward pass (SWD). A northward pass occurs when the ISS ground track moves from lower to higher latitudes (i.e., south to north), while a southward pass occurs when the ISS ground track moves from higher to lower latitudes (i.e., north to south). In French Guiana, the NWD azimuth relative to true north is about 36°, while the SWD azimuth is approximately 144°. Of the three 1064 nm lasers employed by GEDI, one is split into two half-power beams (coverage beams), while the other two remain at full power (power beams). The beams are then slightly dithered, and this setup generates eight parallel ground tracks along the ISS's orbital path: four from the coverage beams and four from the

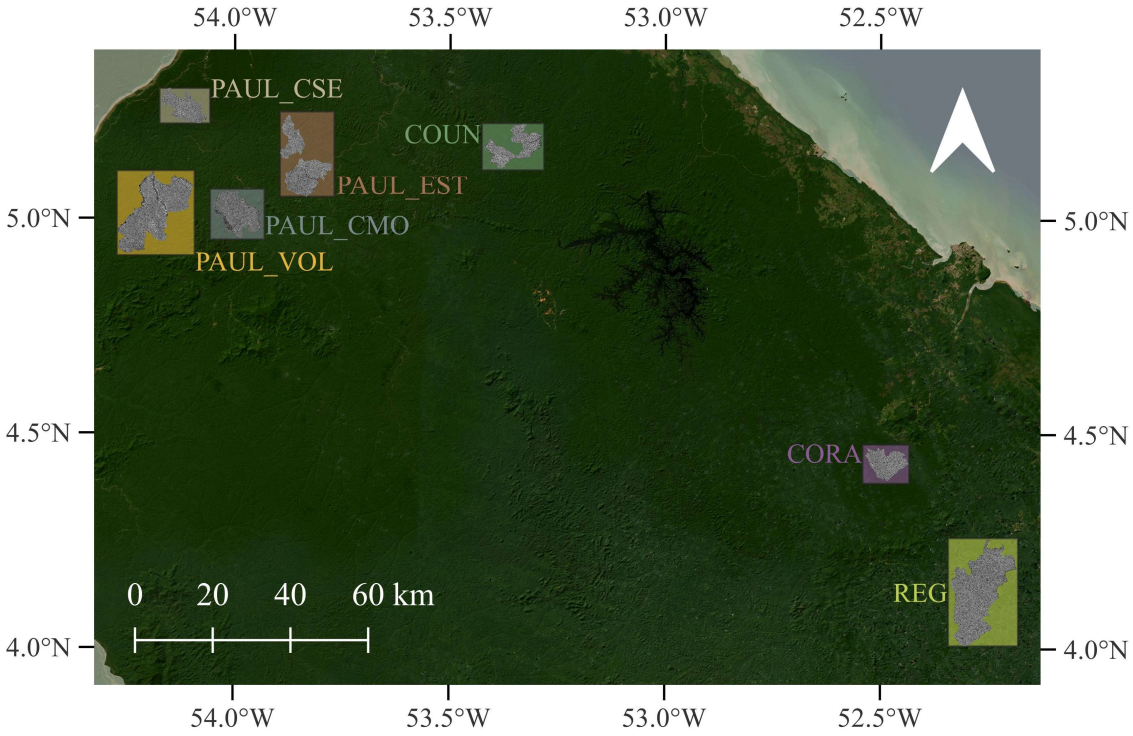


Fig. 2. ALS acquisition areas of the seven French Guiana study sites (ESRI Satellite®).

power beams. For a given pass, whether NWD or SWD, the eight parallel beam ground transects cover a swath of 4.2 km, with footprint samples spaced approximately every 60 m along-track and beam transects spaced approximately 600 m apart on the Earth's surface in the cross-track direction (Fig. 1b).

The GEDI L2A product provides elevation and height metrics derived from the GEDI L1B product's geolocated and smoothed waveforms [38]. These metrics are obtained from six different signal processing configurations or algorithm setting group, with group number 5 generally offering the highest accuracy in tropical areas due to its lower waveform signal end threshold, which better distinguishes weak ground returns in the waveforms [14]. Given the dense vegetation in French Guiana, the *rh_95* metric extracted under the conditions of algorithm setting group number 5 was therefore selected as the direct reliable proxy for canopy height. Between April 2019 and May 2022, 11,798,179 GEDI shots over French Guiana were collected from NASA's GEDI Level 2A Geolocated Elevation and Height Metrics product. These data were filtered to remove irrelevant or erroneous entries, including shots with no detected modes, pure noise signals, and incomplete waveforms, and other criteria [14]. Additionally, only GEDI shots with a signal-to-noise ratio (SNR) greater than 10 dB were retained [39]. After filtering, 2,127,076 waveforms remained, which amounts to about 18% of the original dataset. The corresponding *rh_95* values were rasterized on a 10-m grid. For each GEDI shot, the associated *rh_95* value was assigned to a unique pixel corresponding to the center of the footprint.

2) Satellite Remote Sensing Input Data: The canopy height prediction frameworks analyzed in this study leverage both optical and radar remote sensing data to derive canopy height estimates at a spatial resolution of 10 m. Specifically, optical data from the Sentinel-2 (S2) satellite and radar data from Sentinel-1 (S1) and ALOS-2 PALSAR-2 were utilized. The S2 data provided spectral information crucial for understanding vegetation characteristics, through the inclusion of ten spectral bands: B2 (Blue), B3 (Green), B4 (Red), B5-B6-B7 (Vegetation Red Edge), B8 (Near Infrared, NIR), B8A (Narrow NIR), and B11-B12 (Short Wave Infrared). A relative reflectance normalization was conducted for each spectral band to harmonize reflectance values between orbits [40]. The bands originally at a 20-m resolution were resampled to 10 m. Meanwhile, the radar data from S1 (C-band) and ALOS-2 (L-band) offered complementary structural information on the canopy. The S1 data included four layers based on polarization and orbit: VV ascending, VH ascending, VV descending, and VH descending, each with a pixel size of 10 m. ALOS-2 PALSAR-2 provided dual polarization observations (HH and HV), initially at a 25-m resolution, resampled to 10 m. In total, the integration of these datasets resulted in 16 input layers.

3) Environmental Input Data: Additionally, two environmental descriptors, specifically the height above nearest drainage (HAND) and forest landscape types (FLT) both resampled to 10 m, were integrated into the models' inputs. HAND is a normalized digital elevation model which normalizes topography to the relative heights along the drainage network [41]. FLT provide information on the broader ecological context of French Guiana by describing 20 different forest classes [42]. These descriptors provide valuable contextual information that classical remote sensing data cannot fully reveal, thus enhancing the canopy height

estimates by incorporating aspects related to hydrological and geomorphological characteristics of the landscape that directly impact forest dynamics.

4) ALS Test Data: ALS data were utilized as reference ground truth to assess the accuracy of the canopy height maps both before and after RK. These data were collected by the French National Forests Office (ONF) during several surveys conducted across various study sites in French Guiana between 2016 and 2018. The ALS acquisitions had technical specifications that included an average point density of 10-12 points per m², with each laser pulse covering an approximate diameter of 20 cm and a scan angle of ± 30°. From the raw tridimensional point clouds, canopy height models (CHMs) were generated by ONF at a 1-m resolution (maximum height per 1-m grid cell). To align with the resolution of our canopy height maps, these CHMs were resampled to a 10-m resolution using the maximum value per 10-m grid cell. The maximum value was chosen because it better represents the top-of-canopy signal captured in GEDI waveforms [43], [44], [45].

The ALS acquisitions were conducted over seven distinct areas of interest (Fig. 2): Paul Isnard – Crique Serpent (PAUL_CSE), Paul Isnard – Crique Mousse (PAUL_CMO), Paul Isnard – Voltaire (PAUL_VOL), Paul Isnard – Est (PAUL_EST), Counamama (COUN), Coralie (CORA), and Régina (REG). In this study, localized RK was performed separately on each of the seven individual areas highlighted in Fig. 2, and the results were subsequently aggregated to compute overall accuracy and error metrics.

C. Predicted Canopy Height Maps

1) U-Net Canopy Height Map: The first canopy height map utilized in this study was generated using a U-Net deep learning model known as *CHNET* [23]. The model was developed to produce a 10-m canopy height map from multi-source remote sensing data as well as ancillary environmental parameters. Overall, the *CHNET* model's inputs consist of 18 layers, incorporating both remote sensing (S2, S1, and ALOS-2) and environmental (HAND and FLT) data to produce canopy height estimates. The reference data for model training and validation consisted of the rasterized GEDI *rh_95* metric values.

The U-Net architecture, known for its encoder-decoder structure, was chosen for its ability to leverage multimodal data and to capture both the local and global features of the input data [22], [46], [47], [48]. The encoder part of the network consists of multiple convolutional layers followed by max pooling, which progressively reduce the spatial dimensions while increasing the depth of the feature maps. The decoder part then upsamples these feature maps, combining them with corresponding feature maps from the encoder through skip connections, which help retain spatial information and improve the localization of features. The *CHNET* model was trained using a sequential scenario approach: (1) Initial training with raw GEDI data, which involved training the model using the original *rh_95* dataset as reference heights; (2) Enhanced GEDI data integration, which involved using a refined GEDI dataset to remove data points deemed likely to be unreliable; (3) Incorporation of environmental descriptors, which added hydrological and geomorphological descriptors HAND and FLT; (4) Geolocation correction of GEDI footprints, which incorporated an iterative geo-correction process to address spatial inaccuracies in the geolocations of GEDI waveforms. Each scenario brought improvements in model accuracy and this process finally resulted in the most accurate canopy height predictions with reduced error and minimal bias. The *CHNET* model's performance was rigorously evaluated using an independent GEDI test set to ensure the generalizability of the results. Additionally, the canopy height estimates were assessed against ALS data to confirm the robustness and accuracy of the model as well as the relevancy of GEDI data for canopy height estimation [23].

2) Random Forest Canopy Height Map: In addition to the *CHNET* model, this study also exploits an alternative canopy height map that was obtained from a Random Forest algorithm, hereby referred to as *RFH* (Random Forest for canopy Height estimation). The primary purpose of this additional map is to serve as benchmark for comparison with the *CHNET* model. By including a classical machine learning approach to canopy height mapping, we aim to provide a reference point to better evaluate the performances and advantages of the *CHNET* framework. The Random Forest is a widely used ensemble learning method that builds multiple decision trees during training and outputs the average prediction for the individual trees [49]. This method is particularly robust to overfitting while being able to model non-linear relationships for large datasets with many features, making it suitable for our canopy height estimation task.

For consistency, *RFH* was implemented and tested under the same conditions as *CHNET*, using the same input data and the same geo-corrected GEDI reference dataset. This input data comprised 18 layers, including optical data from S2 (10 layers), radar data from S1 (4 layers) and ALOS-2 (2 layers), as well as two environmental descriptors HAND (1 layer) and FLT (1 layer). Each pixel where a valid *rh_95* value was available was treated as an individual object described by these 18 variables. Therefore, the dataset used to train and validate *RFH* is a table of points (i.e., pixels), where each row represents an individual pixel with its associated GEDI *rh_95* reference canopy height, and the columns represent the 18 descriptive variables. To ensure consistency in our experimental setup, we maintained the same train-validation-test split that was used in the *CHNET* model. This approach ensures that any differences in model performance can be attributed to the algorithms themselves rather than differences in the data used for training and validation. A grid search with cross-validation was also performed to optimize the hyperparameters of *RFH*. After the training process, *RFH* was applied to predict the canopy height for each 10-m pixel across French Guiana. This procedure involved using the trained and validated *RFH* to output a predicted canopy height value at each geospatial location based on the 18 input descriptors.

D. Residual Kriging

1) Ordinary Kriging of Model Residuals: When mapping a continuous variable such as canopy height across a large study

site, understanding the spatial behavior of the reference data is crucial for better interpretation of the final predictions. Since *RFH* treats each prediction independently, any spatial correlation in the results can only be due to existing spatial correlations in the input variables. Nonetheless, some of the unexplained variance in *RFH* predictions may be attributable to spatial correlations within GEDI canopy heights. Therefore, a spatial prediction model is insightful when dealing with spatially dependent data. In contrast, the U-Net architecture utilized in the *CHNET* framework leverages local and global features of the input predictors, offering some capability to better capture spatial dependencies. Unlike *RFH*, which treats each prediction independently, the *CHNET* model's convolutional layers can implicitly identify and learn spatial structures from the input data, which enhances the model's performance in a spatially dependent regression task. However, despite these advantages, *CHNET* does not explicitly model the spatial correlation of GEDI reference canopy heights. Consequently, there is also a need to assess if *CHNET* residuals are spatially structured, which would offer the possibility to explicitly model the spatial correlation of these residuals.

Kriging is a group of geostatistical techniques designed to predict the optimal estimate of a spatially distributed variable V at any unsampled location. This estimate is computed as a weighted average of observed values at surrounding points, with the weights being determined by the spatial covariance of the observations. Various kriging methods exist, and the most relevant for this study is RK [50]. RK posits that the target variable V measured at a location y can be represented as the sum of a deterministic trend d and stochastic residuals s:

$$V(y) = d(y) + s(y) \quad (1)$$

The deterministic drift d represents the large-scale trend or systematic part of the spatial variable. It is a predictable function that can be modeled using known covariates or polynomial functions. Essentially, it captures the overall structure of the target variable, and it only captures the general trend of the spatial data. It does not account for the more detailed and localized spatial relationships. Conversely, the stochastic residuals s correspond to more detailed local-scale variations that cannot be captured by the deterministic drift. They represent spatially correlated deviations from the large-scale trend, and they contain information that reflects local variability and spatial autocorrelation. Residuals can be estimated by means of a standard kriging procedure. In this study, the residuals, defined as the difference between the observed (i.e., the GEDI *rh_95* metric values) and the predicted (i.e., the outputs of either *CHNET* or *RFH*) canopy heights, display local second order stationarity and can therefore be estimated by an ordinary kriging (OK) algorithm. The key assumption of OK is that there is a spatial autocorrelation between nearby values, meaning that nearby locations are likely to have similar values. The estimated residual $\hat{s}$ at an unsampled location y_0 is obtained using a weighted average of n known residual values s at nearby sampled locations y_i:

$$\hat{s}(y_0) = \sum_{i=1}^{n} \lambda_i s(y_i) \quad (2)$$

The weights λ_i are calculated such that the estimation is unbiased, and they are determined by minimizing the estimation variance, under the unbiasedness condition. The values of the weights depend on the spatial structure of the residuals, which is described by means of a semivariogram function [51]. The semivariogram function γ quantifies the spatial dependence of the residuals s by measuring the average dissimilarity between pairs of points as a function of the distance h (lag) between them:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [s(y_i + h) - s(y_i)]^2 \quad (3)$$

$\gamma(h)$ represents the semivariance at a lag distance of h, while $N(h)$ is the number of pairs of points separated by a distance of h and s are the canopy height prediction model residuals at locations y_i and $y_i + h$.

Empirical semivariograms are a fundamental tool in geostatistics used to characterize the degree of spatial dependence between observations as a function of distance. As highlighted in (3), the empirical semivariogram of the residuals is computed using pairs of points within the study area. Specifically, for each pair of points, the semivariance is calculated as half the average squared difference between the values at the two points. The data is then aggregated into distance bins to average the semivariance values within each bin to smooth out variability and reveal more meaningful patterns. In our study, the width of distance intervals into which data point pairs are grouped for semivariance estimates is set to 100 m. Next, a mathematical function is fitted to the empirical semivariogram to model the spatial structure of the residuals. This fitted function allows the empirical semivariogram of the residuals to be represented mathematically, providing a basis for spatial application in any coordinates of the area. The fitted semivariogram has three key parameters: (1) Nugget, which represents the semivariance at a lag distance approaching zero, indicating measurement error or small-scale variability; (2) Sill, which is the semivariance value where the semivariogram reaches a plateau, beyond which there is no longer spatial correlation; (3) Range, which is the distance at which the semivariance reaches the sill, marking the extent of spatial correlation.

Using the fitted semivariogram, an OK procedure is performed on a 10-m grid to predict residuals at unsampled locations, producing an interpolated map of the model residuals, referred to as kriged residuals. The final step of RK consists in combining the regression predictions $\hat{d}$ with the kriged residuals $\hat{s}$ to produce new canopy height estimates $\hat{V}$ at every unsampled location y_0:

$$\hat{V}(y_0) = \hat{d}(y_0) + \hat{s}(y_0) \quad (4)$$

This whole process, comprising the computation of the empirical semivariogram, the fitting of a mathematical function, and the OK procedure, is applied to both *CHNET* and *RFH* residuals. By integrating the kriged residuals back into the initial canopy height predictions, we aim to account for the spatial dependencies not captured by the original models, and thus incorporate both the deterministic and spatial components of the data. *RFH* can only capture the spatial autocorrelation present in the input variables, but not that of

GEDI data. Therefore, we use RK to account for the spatial autocorrelation of canopy heights, resulting in an enhanced model referred to as *RFH-RK*. For *CHNET*, although it inherently integrates spatial information through its architecture, we apply RK to evaluate whether additional spatial modeling offers any further benefits, resulting in a new model called *CHNET-RK*. RK allows us to produce new and potentially enhanced canopy height maps on one hand, and to determine the effectiveness of adding a spatial component to both models on the other hand. Ultimately, our analysis includes a comprehensive comparison of different canopy height maps produced with different methodological approaches and subsequently refined with RK.

2) Spatial Analysis of Model Residuals: To explore the spatial dependencies of GEDI measurements and understand whether they are accounted for in the regression models, we perform a spatial analysis of GEDI *rh_95* values, regression results, and models' residuals. The primary focus is on the residuals, with the goal of producing for each model a corrected map that incorporates the spatial correlations of GEDI measurements that were not captured by the regression models. The analysis of GEDI *rh_95* spatial correlations helps confirm that the observed correlations in the residuals are indeed associated with the GEDI data themselves, rather than being a result of the regression models. Similarly, the predictions from *CHNET* and *RFH* are also analyzed. These spatial analyses are based on the computation of empirical semivariograms to describe spatial patterns. These empirical semivariograms provide valuable information about the spatial structure and variability of the data, revealing patterns such as spatial continuity and potential anisotropies.

The specific parameters of the GEDI acquisition, i.e., sparse measurements along azimuthal tracks and laser beam type, introduce a spatial correlation in the regression results that does not reflect the actual canopy structure and dynamics. We hypothesize that the acquisition configuration as well as the characteristics of the GEDI sensor introduce anisotropies originating from the measurements themselves rather than the canopy. To perform effective RK based on a representative spatial structure, it is crucial to differentiate and account for the spatial correlations of forest canopy heights, while minimizing or correcting for the spatial correlations introduced by the instrument or measurement methodology.

Given the specificities of the GEDI data used as reference canopy height and their potential spatial effects we aim to observe, we first compute empirical semivariograms at different levels of GEDI data filtering with respect to beam type, including all data, power-only, and coverage-only. This approach isolates each particularity of the GEDI acquisition in terms of laser energy to understand its impact on the spatial autocorrelation of the measurements. Additionally, directional semivariograms are computed to study the spatial correlation in specific directions. We hypothesize that the azimuthal configurations of GEDI tracks introduce anisotropic effects, particularly in the cross-track directions. Indeed, each ground track corresponds to either power or coverage beams. Moreover, they are acquired at different periods, and the atmospheric conditions at the time of the acquisitions may differ, resulting in different received waveforms, and in turn eventual changes in the *rh_95* values across tracks. To validate this hypothesis, we perform a spatial analysis through empirical semivariograms computed from different levels of GEDI data filtering with respect to track azimuthal direction. Specifically, semivariograms are computed using data from NWD tracks on one hand, and SWD tracks on the other, which are sub-datasets containing parallel tracks only. We hypothesize that the true forest canopy heights should exhibit minimal anisotropy, meaning that there should be no significant change in the spatial correlation of canopy height with azimuthal direction when averaged over a large area such as French Guiana. Consequently, any observed anisotropy in the residuals (or in *rh_95*) is likely attributable to sensor-related effects rather than natural properties of the forest. Eventually, we consider that the spatial structure of the residuals characterized under conditions that mitigate anisotropic effects reflects the reality of the horizontal structuring of canopy height in all directions in an isotropic manner.

In the end, these spatial analyses are a crucial step to obtain a relevant and reliable representation of the spatial structure of the residuals, with the final goal of performing RK. In this study, using the insights gained from these preliminary analyses, the empirical semivariograms of the models' residuals are computed under conditions designed to mitigate GEDI sensor-induced anisotropies and to represent the true spatial correlations of forest canopy heights.

III. Results

A. Semivariogram Analysis

1) CHNET Predictions and Residuals: To confirm that the specificities of the GEDI acquisition induce a sensor-related spatial correlation in the measurements, we first computed and plotted the empirical semivariogram of the reference heights used for the *CHNET* model's training and validation, specifically the *rh_95* values (Fig. 3a). The semivariogram of the GEDI *rh_95* data exhibits a pronounced spatial periodicity of approximately 600 m. This periodicity is particularly strong at the 600-m mark, where the semivariance peaks significantly. This trend continues, although with a diminishing effect, at subsequent multiples of 600 m. This 600-m periodicity is consistent with the distance between two adjacent GEDI ground tracks, indicating a strong spatial dependence corresponding to the sensor's acquisition pattern. At the scale of French Guiana, there should be no periodicity related to canopy height in the spatial correlation, nor should there be any anisotropies. Therefore, we hypothesize that any observed periodicity or anisotropy in the *rh_95* values results from sensor measurements artifacts rather than true spatial correlations of canopy height.

Similarly, we computed the semivariogram for *CHNET* predictions of canopy height and corresponding model residuals (Fig. 3b). The predictions display a clear spatial structure, which means that *CHNET* accounts for some spatial correlation in its predictions. The semivariogram of the residuals also shows a degree of spatial structure, with slightly lower semivariance values than the predictions. This implies that while *CHNET* accounts for some spatial patterns, a

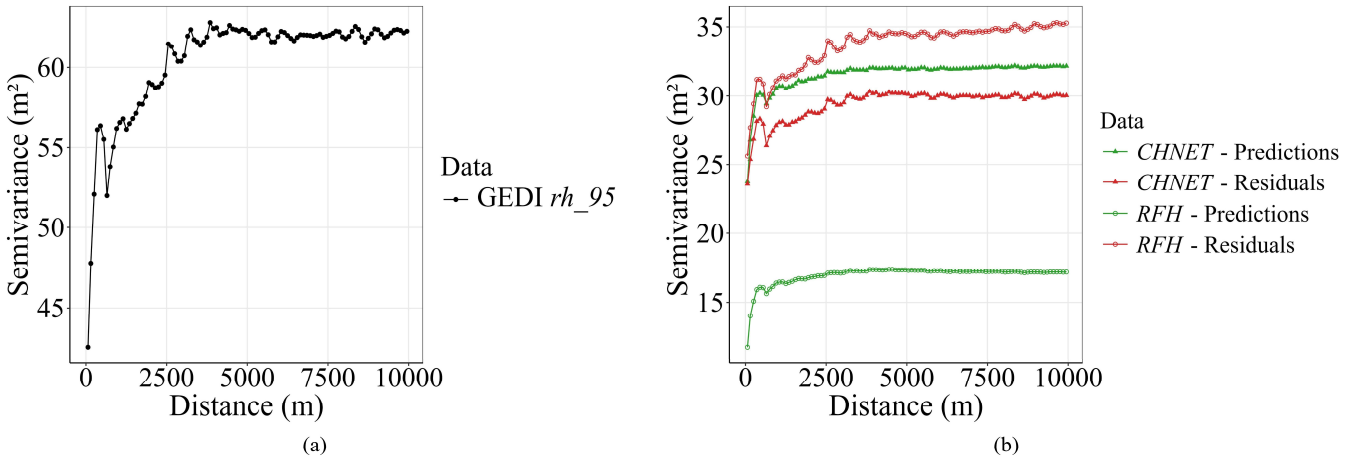


Fig. 3. Omnidirectional semivariogram of GEDI *rh_95* values (a). Omnidirectional semivariograms of *CHNET* and *RFH* predictions and residuals (b).

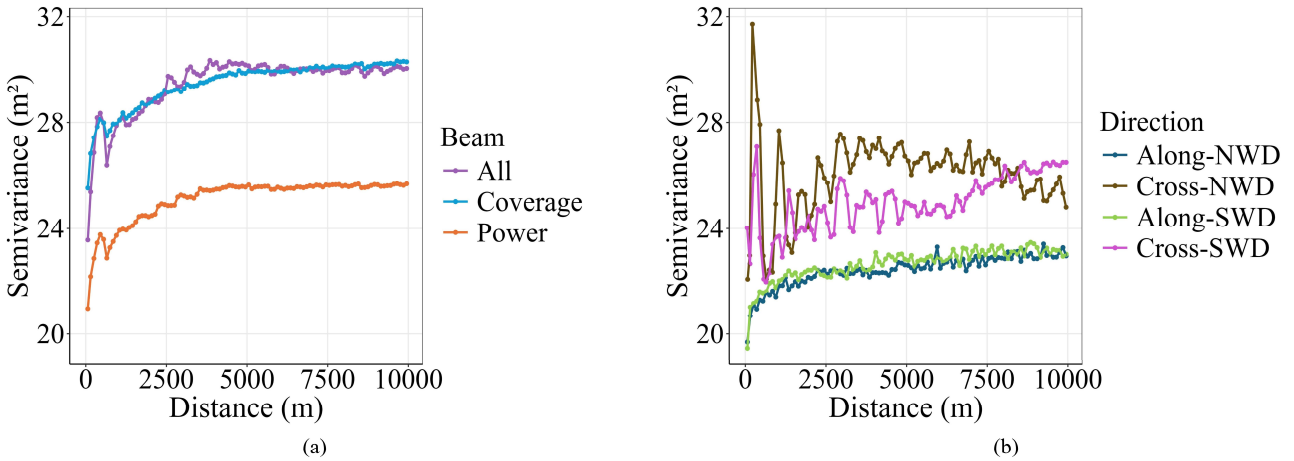


Fig. 4. Omnidirectional semivariograms of *CHNET* residuals depending on beam type (a). Directional semivariograms of *CHNET* residuals in the along-track and cross-track directions (b).

portion of the spatial autocorrelation remains unexplained by the model and is contained within the residuals. The results also reveal the same periodic pattern for the residuals as the one observed in the semivariogram of GEDI *rh_95* data. Observing the same periodicity in the semivariograms of both GEDI *rh_95* values and the residuals supports the idea that this periodicity is linked to GEDI and is not an artifact from the *CHNET* model. This idea is further confirmed by the semivariogram of the predictions, which indicates that the model performs some level of smoothing on the periodicity present in the semivariogram of GEDI *rh_95* data. Therefore, we conclude that the observed periodicity is sensor-related and is not linked to a specific spatial organization of the forest canopy height. Specifically, the 600-m spatial pattern reflects the characteristics of the GEDI sensor and its data acquisition process rather than the behavior of the *CHNET* model itself.

Among the specific characteristics of GEDI acquisition, laser beam type can be a significant source of anisotropy and inconsistency in the measurements, as laser energy is halved depending on beam type. To study this impact, we plotted the empirical semivariograms of *CHNET* residuals based on beam type (Fig. 4a). We found that the magnitude of semivariance values differs substantially depending on beam type, with significantly lower values for power beams compared to coverage beams. Additionally, when considering a single beam type in the spatial analysis, the observed periodicity is notably less pronounced than when all beams are included. These findings indicate that differences in laser beam type can be considered as a source of anisotropy in GEDI measurements, and that the associated spatial autocorrelation does not reflect the actual spatial structure of the forest's canopy.

Given these results, only GEDI shots of full power (i.e., power beams) are retained for the subsequent steps of the study, as these shots generally provide more accurate and reliable canopy height predictions. From the preceding analyses, it appears that the anisotropic effects in GEDI data result from a combination of track geometry and beam type. To further our understanding of anisotropy sources, we divided our GEDI power beam dataset into two subdatasets corresponding to the GEDI shots of each of the two azimuthal directions NWD and SWD. Specifically, one subdataset contains only parallel power beam ground transects of either NWD or SWD azimuth. Directional semivariograms are an

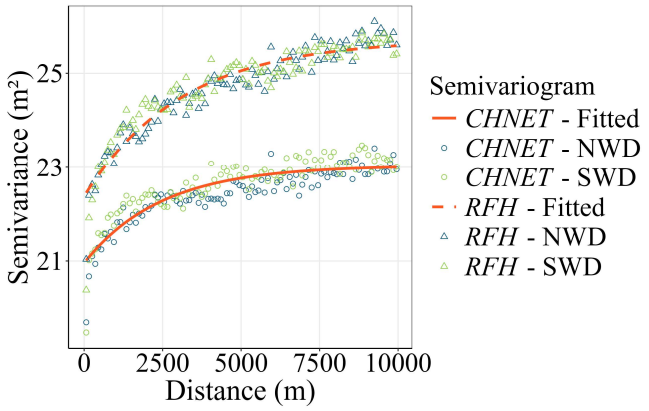


Fig. 5. Directional semivariograms (empirical and fitted) of *CHNET* power beam residuals for NWD and SWD configurations.

effective tool to study spatial patterns in specific directions. Therefore, we plotted the directional semivariograms of *CHNET* residuals in the along-track and cross-track directions for each subdataset separately. Fig. 4b shows the directional semivariograms in the along-NWD and cross-NWD directions, which correspond to 36° and 126° relative to true north, respectively, and in the along-SWD and cross-SWD directions, which correspond to 324° and 234° relative to true north, respectively. Notably, the results obtained for NWD and SWD azimuths are equivalent. Cross-track semivariograms exhibit a high periodicity, while the periodicity is minimal in the along-track directions. These directional semivariograms clearly demonstrate that the spatial periodicity previously observed in Fig. 3 and 4a is linked to the pattern of GEDI beam ground transects. The 600-m period visible in the cross-track directions corresponds to the spacing of GEDI tracks.

From the analyses performed at this point, we validate that the combination of laser beam type and GEDI track acquisition introduce sensor-related anisotropies in the spatial distribution of the measurements. The spatial autocorrelation of the residuals is almost equivalent in the along-track directions of both NWD and SWD subdatasets taken independently. Hence, we posit that these along-track semivariograms represent the underlying isotropic (i.e., omnidirectional) spatial autocorrelation of model residuals without sensor-induced anisotropies. These semivariograms obtained along-track accurately reflect the actual spatial autocorrelation structure of the residuals. In the end, the corresponding empirical semivariogram can be used isotropically with GEDI power beams in the RK process. Fig. 5 shows the associated directional semivariograms of *CHNET* residuals for each subdataset in the corresponding along-track direction. It is important to note that the directional semivariograms presented in Fig. 4b and 5 are noisier due to the strict tolerance angle of 1° used during their computation. This tolerance was chosen to ensure that the semivariograms are as directional as possible, which consequently reduces the number of point pairs used for semivariance calculation. This reduced number of pairs introduces additional noise into the results. However, despite this increase in noise, the spatial correlation remains observable, and the validity of the observed trends is not compromised.

2) RFH Predictions and Residuals: A comprehensive analysis identical to that performed for the *CHNET* model was also conducted for the *RFH* model. The empirical semivariogram of *RFH* residuals (computed from all available GEDI data) reveals a similar periodic pattern (Fig. 3b), which further confirms our earlier findings that the observed anisotropies are attributable to the GEDI sensor itself. Furthermore, the semivariogram of *RFH* predictions (Fig. 3b) shows almost no spatial structure. *RFH* being a non-spatial method by design, it implies that the input data provided for model calibration (i.e., optical, radar, and environmental descriptors) do not contain strong spatial information or are insufficiently spatially correlated with GEDI canopy heights. As a result, the model's predictions exhibit minimal spatial autocorrelation. In contrast, the residuals present a significant spatial structure, with significantly higher semivariance values compared to those of the predictions. This indicates that the residuals integrate the remaining spatial autocorrelation that is not accounted for by the model, which makes them the most spatially structured among all configurations presented in Fig. 3b.

To derive a representative spatial structure of *RFH* residuals, we utilized the same configuration identified as optimal in the previous analyses of *CHNET* residuals. Fig. 5 presents the directional semivariograms of *RFH* residuals in the corresponding along-track direction for each GEDI power beam subdataset, either from NWD or SWD tracks. The results corroborate the findings from the *CHNET* residuals analysis. Indeed, to perform relevant RK of the *RFH* model's predictions, the optimal approach also involves using the directional semivariogram of power beam residuals belonging to a single azimuthal direction, which is computed in the corresponding along-track direction.

B. Residual Kriging

1) Semivariogram Fitting: After understanding the sources of anisotropies in GEDI measurements and mitigating them to produce relevant empirical semivariograms of the models' residuals that accurately represents the actual spatial correlation of GEDI canopy heights in French Guiana, the next step was to fit mathematical functions to these empirical representations to perform RK. In this purpose, many classical functions exist, such as linear, spherical, circular, gaussian, and exponential models, each characterizing different types of spatial correlation. A linear model means that the spatial autocorrelation increases linearly with distance. The spherical model is a modified quadratic equation where the spatial dependence levels off at the sill and range values. Circular models resemble the spherical ones, with spatial dependence diminishing to an asymptotic level. The gaussian models use a normal probability distribution curve and have an inflection point. For exponential models, the spatial autocorrelation gradually reaches the sill, with the relationship between two distance bins decaying progressively until spatial dependence dissipates at an infinite distance.

As highlighted in Subsection III-A, for each regression model, we generated two empirical semivariograms of the residuals to model the spatial autocorrelation of canopy

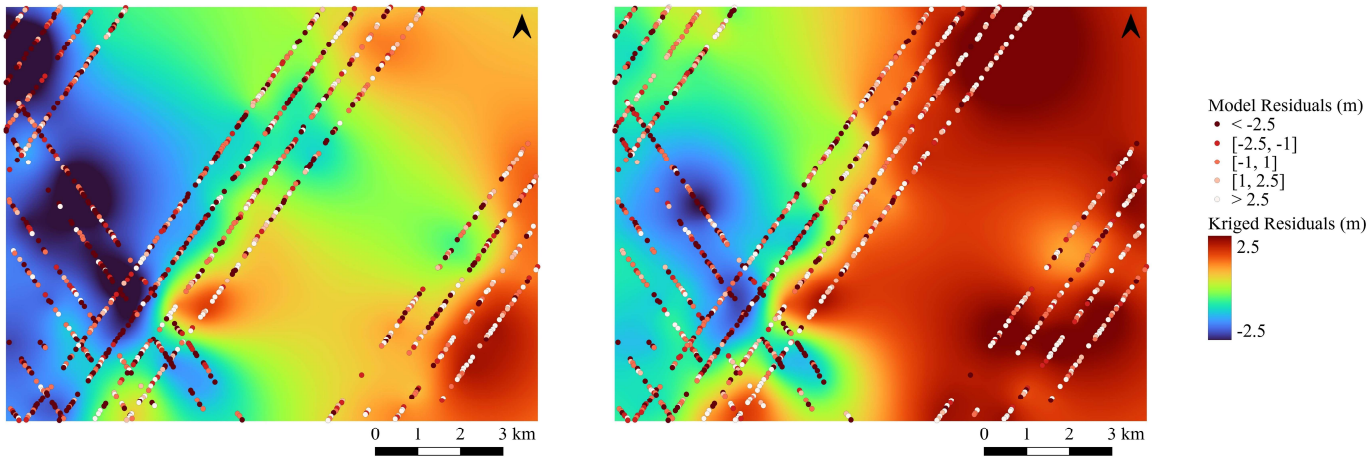


Fig. 6. Kriged residuals and model residuals of GEDI power beam footprints for *CHNET* (a) and *RFH* (b) in a window within the PAUL_EST area.

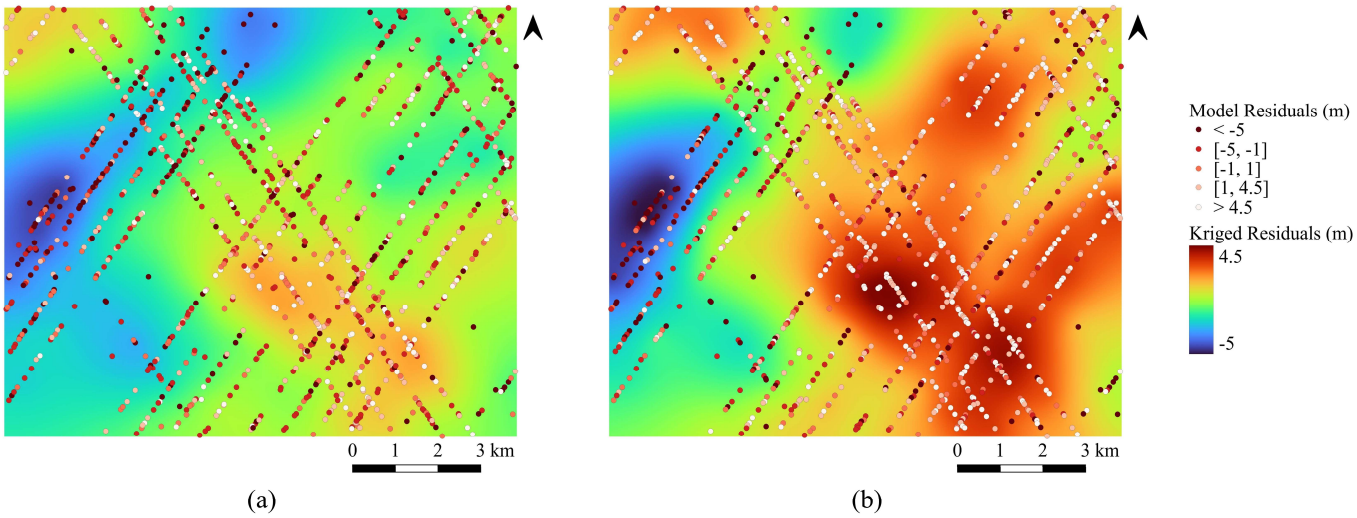


Fig. 7. Kriged residuals and model residuals of GEDI power beam footprints for *CHNET* (a) and *RFH* (b) in the CORA area.

heights: one for the NWD configuration and another for the SWD configuration. Since both semivariograms displayed equivalent behaviors, which indicates consistent spatial patterns in both configurations, we proceeded by fitting a single semivariogram model to the combined empirical data. We fitted the functions described above and observed that the exponential function exhibited the best correlation with the empirical data, achieving R^2 scores of 0.87 and 0.93 for *CHNET* and *RFH*, respectively. As shown in Fig. 5, the fitting process resulted in exponential models characterized by the parameters described in Table I. The residuals of *RFH* appear to exhibit more spatial structure than those of *CHNET*, as highlighted by the greater sill value. The range of the spatial autocorrelation is also larger by more than 500 m for *RFH* compared to *CHNET*. Spatial dependencies persist over greater distances in the *RFH* residuals, which suggests that the *RFH* model leaves more of the spatial correlation unaccounted for. In the end, these fitted semivariograms served as the covariance models used to perform RK for *CHNET* and *RFH*.

2) Kriged Residuals: To explore the practical application of RK and its effectiveness in correcting the regression models' predictions, we first conducted localized computations rather than applying the method across the entire study area. This approach allowed us to evaluate the capabilities and potential of the RK procedure in smaller zones before scaling up. By initially focusing on localized areas, we were able to gain insights into the kriging process and draw preliminary observations. RK was executed separately on several localized areas using the fitted exponential semivariograms, and with sampled points consisting of the GEDI power beam footprints in those areas. It was performed on a 10-m resolution grid to produce a continuous map of kriged residuals. Fig. 6 shows a representative example of kriged residuals for *CHNET* (Fig. 6a) and *RFH* (Fig. 6b) after RK was applied to a localized window of 10×12.5 km in the PAUL_EST area, out of several windows analyzed.

TABLE I
PARAMETERS OF THE FITTED EXPONENTIAL SEMIVARIOGRAMS FOR *CHNET* AND *RFH* RESIDUALS

Model	R^2	Nugget (m^2)	Sill (m^2)	Range (m)
CHNET	0.87	21.0	23.0	2466.0
RFH	0.93	22.4	25.7	3096.2

Several key observations emerge from this analysis. Firstly, the corrections provided by RK are predominantly localized around GEDI measurement points. These localized corrections appear as patches in the kriged residuals maps (Fig. 6), highlighting areas where the kriging process has estimated the residuals based on nearby GEDI data. Secondly, the kriged residuals show a tendency to reproduce the spatial pattern of model residuals. Specifically, in regions where residuals are high and positive, the kriged residuals also tend to be positive. Conversely, in areas where residuals are highly negative, the kriged residuals generally fall within the negative range. This indicates that RK is effectively capturing and adjusting for local variations in the residuals. Thirdly, regions with few or no GEDI footprints tend to have kriged residuals that are closer to the average of the range of values, rather than exhibiting extreme values. This observation underscores that, contrary to regions with many GEDI measurements, areas lacking GEDI information do not experience a significant correction of their predictions. Finally, we note that the magnitude of the corrections applied by RK in this area is similar for *CHNET* and *RFH*, with kriged residuals varying between -2.5 and 2.5 m.

A localized RK procedure was also applied to the entire CORA area (study site shown in Fig. 2), with the goal of progressively scaling the analysis up. CORA area is characterized by a greater density of GEDI points than PAUL_EST. Fig. 7 shows the kriged residuals of *CHNET* (Fig. 7a) and *RFH* (Fig. 7b) after performing RK. The results allow drawing the same conclusions as before. Specifically, we observe that the magnitude of the corrections applied by RK is less pronounced for *CHNET* than for *RFH*, with kriged residuals varying between -4.5 and 2 m for the former, and between -5 and 4.5 m for the latter. This wider level of correction suggests that a more significant adjustment is needed for *RFH* predictions compared to those of *CHNET* in the CORA area. This observation is consistent with the higher model residuals observed for *RFH* (Fig. 7), which indicates that its predictions require a more pronounced spatial correction.

3) Enhanced Predictions and Performance Assessment: Following the localized analyses, we scaled up the RK process to apply it separately to each of the seven ALS survey areas (Fig. 2), and then aggregate the results for performance assessment. This approach is possible because the distance between the different areas of interest is greater than the range of the fitted semivariograms, which allows the application of independent RK procedures to each area. Furthermore, to mitigate edge effects when comparing the spatially corrected predictions with ALS canopy heights, we introduced an additional buffer zone of 3 km around each area shown in Fig. 2, which is approximately equal to the range of the fitted semivariograms, to extend the areas on which RK is performed. Indeed, spatial interpolation near the edges can be less reliable due to the lack of nearby sampled points. Performing RK on buffered areas ensures that values on the edges do not compromise the comparison with the test data.

TABLE II
ACCURACY METRICS OF *CHNET* AND *RFH* BEFORE AND AFTER RK COMPARED TO GEDI AND ALS TEST DATA

Test Set	Model	Bias (m)	RMSE (m)
GEDI	*CHNET*	0.6	4.7
	CHNET-RK	0.0	4.2
	RFH	-1.3	5.3
	RFH-RK	0.0	4.4
ALS	*CHNET*	-0.2	5.8
	CHNET-RK	-0.1	5.8
	RFH	-1.8	6.1
	RFH-RK	0.5	6.1

For each of the seven areas of interest, and for both *CHNET* and *RFH,* a RK procedure was applied using the fitted exponential semivariograms, and with sampled points consisting of the GEDI power beam footprints in those areas. The obtained kriged residuals were integrated to the original predictions to produce spatially corrected predictions. Specifically, for both *CHNET* and *RFH*, we computed the kriged residuals over a 10-m grid to add them to the corresponding 10-m canopy height maps. The enhanced canopy height maps produced through RK were then evaluated against GEDI power beam data and ALS-derived reference heights. Table II presents a comparison of the four distinct canopy height products: the initial predictions from both *CHNET* and *RFH*, and the spatially corrected predictions obtained from RK for each model (*CHNET-RK* and *RFH-RK*).

The evaluation of the canopy height maps produced through RK for both *CHNET* and *RFH* reveals several key insights. The primary observation is that RK has a positive impact on the accuracy of canopy height estimates with respect to GEDI data, as it allows reducing the RMSE by half a meter for *CHNET* and about a meter for *RFH*. More specifically, in the case of *RFH*, the RK procedure results in a reduction in bias, as we observe a shift from negatively biased estimates (with a bias of -1.3 m for *RFH*) to unbiased estimates (no bias for *RFH-RK*). This indicates that RK tends to shift *RFH* estimates upwards to align them more closely with GEDI measurements. The same trend is observed regarding *CHNET* predictions, although in a less notable way, with a bias value going from 0.6 m for *CHNET* to a null bias value for *CHNET-RK*. Overall, the integration of kriged residuals through a RK procedure proves to enhance both *CHNET* and *RFH* predictions when assessed against GEDI reference canopy heights.

Conversely, a striking observation is that RK seems to have no discernible impact on the general accuracy of canopy height estimates when compared to ALS data. Indeed, *CHNET* and *CHNET-RK* exhibit almost identical performances. In the

TABLE III
RMSE VALUES OF *CHNET-RK* AND *RFH-RK* COMPARED TO ALS DATA WITHIN CIRCULAR WINDOWS AROUND GEDI POINTS

Radius (m)	RMSE (m)	
	CHNET-RK	***RFH-RK***
0	4.9	5.0
250	5.2	5.4
500	5.3	5.6
1000	5.3	5.7
Range	5.8	6.1
∞	5.8	6.1

case of *RFH*, the application of RK still results in an upward shift in the estimates, which is evident in the change from a strong negative bias of -1.8 m for *RFH* to a slightly positive bias of 0.5 m for *RFH-RK*. Despite this adjustment, no improvement in RMSE is observed. However, this apparent lack of improvement may be explained by the different natures of the performance assessments. As previously observed, the RK process provides more significant and tangible corrections around GEDI points. When comparing the results of *CHNET-RK* and *RFH-RK* with GEDI data, the evaluation is performed at discrete points, where there is available GEDI information and where RK can provide targeted corrections. Conversely, assessing the results against ALS data involves the use of continuous maps, which include areas with few GEDI footprints that may not have been significantly influenced by the corrections from the RK process. Consequently, any improvements introduced by RK might be less apparent in an overall comparison with ALS data, as the enhancements could be less discernible when averaged across the broader spatial extent of the ALS acquisitions. Therefore, to accurately investigate the impact of RK in relation to ALS data, we computed errors within circular neighborhoods around GEDI points with varying radius values. The goal was to assess the localized effects of RK more effectively and potentially highlight improvements that are less evident in an overall comparison.

Table III presents the RMSE values of *CHNET-RK* and *RFH-RK* compared to ALS data within circular windows centered around GEDI points. A radius of 0 m represents the exact location of the GEDI points, while an infinite radius corresponds to the full spatial extent of the ALS data. Range is a radius value that corresponds to the range of the fitted semivariogram for each model (see Table II), i.e., 2466.0 m for *CHNET* residuals and 3096.2 m for *RFH* residuals. As the radius decreases, both *CHNET-RK* and *RFH-RK* exhibit a progressive decrease in RMSE values, which suggests that the RK process is more effective in regions closer to GEDI points, where more significant corrections may have been applied. The trends are consistent for both *CHNET-RK* and *RFH-RK*, with an overall advantage for *CHNET-RK* across all configurations. Notably, between the point comparison (at a radius of 0 m) and the full ALS extent (infinite radius), there is an approximate RMSE difference of 1 m for both models, which indicates the localized effectiveness of RK at GEDI points. Additionally, at a radius equal to the range of the fitted semivariograms, the RMSEs for both models are the same as the infinite radius. This is due to the spatial density of the GEDI database. Indeed, such radius values around GEDI points almost amount to the extent of the ALS acquisitions, which makes these two configurations identical.

IV. DISCUSSION

A. On Sensor-Induced Anisotropies

The results of our study demonstrated that GEDI's data acquisition parameters introduce sources of spatial anisotropies that are not representative of the actual horizontal structure of the canopy. These anisotropies are linked to the pattern of the GEDI sensor acquisition as well as signal physical parameters such as laser energy. In French Guiana, and more generally in tropical biomes, the penetration of vegetation by the signal is a major challenge, as these ecosystems are characterized by tall and dense canopies. To accurately measure canopy height, the signal must reach the ground, and the capability of the signal to penetrate through the forest and detect the ground is closely related to the laser's physical properties.

Numerous studies have highlighted the strong relationship between signal penetration capabilities and signal properties, which directly impacts the measurements performed by GEDI and the derived canopy heights [43], [52], [53]. Specifically, GEDI's power beams are twice as powerful as coverage beams in terms of laser energy, significantly influencing their ability to penetrate dense canopies. For example, Fayad et al. [25] noted that coverage lasers exhibited substantially lower performances for tree height estimation compared to full power configurations like those of GEDI and NASA's Land Vegetation and Ice Sensor (LVIS). In an analysis over protected study sites in French Guiana, Lahssini et al. [14] observed that power beams produced significantly better measurements for canopy height estimation compared to coverage beams. They found that coverage beams tended to underestimate tree heights, whereas power beams showed a strong linear correlation with ALS reference canopy heights. Coverage lasers generally face more difficulty in reaching the ground because of the dense vegetation, leading to an overestimation of ground elevation and an underestimation of canopy heights. Conversely, power beams can penetrate the vegetation more effectively, and produce return waveforms with a recognizable lowest mode corresponding to the actual ground. Other studies have also confirmed that beam strength significantly impacts the results and have recommended using only power beam data for more accurate canopy height estimates [54]. Nevertheless, GEDI power beams still exhibit a tendency to underestimate canopy height in dense tropical forests with high AGB [25].

These differences in measurements introduce anisotropies because the different beams do not produce the same measurements. For example, the same location in a dense tropical forest observed by a power or coverage laser would produce different waveforms, and consequently, different reference heights for that particular location. When these effects are combined with the ground track pattern of GEDI, which includes straight and parallel beam transects of either power or coverage laser in two azimuthal directions, it creates

periodic anisotropy patterns in given directions. Our semivariogram analysis demonstrated that depending on the beams considered and the direction of the analysis, the subsequent spatial autocorrelation extracted was not consistent. Therefore, it is crucial to account for anisotropies that are due to sensor-related effects to analyze the semivariograms of model residuals and eventually perform RK. In this study, the GEDI dataset was gradually refined to achieve a configuration that provided an accurate and artifact-free representation of the spatial correlation of GEDI canopy heights. In tropical contexts, laser power is of paramount importance, which is why we chose to use power beams exclusively to obtain a consistent representation of the spatial structure of canopy heights. We also chose to incorporate a directional aspect in the spatial analysis to account for the spatial sampling of GEDI measurements. An alternative approach to addressing sensor-induced anisotropies would be to handle these issues earlier in the regression process by implementing a preliminary step to refine the GEDI data used as reference for model training. In this approach, the regression model is not responsible alone for managing sensor-related discrepancies, as it is fed with preprocessed data. Corrected GEDI-derived heights could be obtained by constructing linear or non-linear models based on GEDI metrics, signal parameters, and other environmental factors, like terrain that can also introduce anisotropies in GEDI measurements [45]. Provided there is enough ground truth data, such as ALS data, to build these models, they could be applied to correct GEDI relative height metrics and produce new reference heights that are free from anisotropy biases. In our study, however, we chose to filter GEDI data to retain only the most relevant measurements. Coverage beams present significant challenges, particularly in dense tropical environments, due to their limited penetration of the vegetation. Accurately determining the true values of the associated metrics is difficult, and attempting to correct them would likely introduce additional noise and uncertainty. Furthermore, addressing the 600-m periodicity is even more challenging, as the differences between tracks arise from complex atmospheric factors that are difficult to model and correct in a reliable way. Consequently, it is more effective in our case to filter the data and focus on usable measurements rather than introduce corrections that have their own uncertainties.

B. On Spatial Information for Canopy Height Mapping

When mapping a continuous variable like canopy height, it is essential to understand that it exhibits spatial characteristics influenced by numerous environmental and ecological factors. Canopy height is not uniformly distributed across landscapes due to the complex interactions between many different environmental parameters. For example, topography can play a vital role in forest growth by influencing water availability and hydrological networks. In this perspective, the height above nearest drainage (HAND) descriptor proved to provide essential contextual information that is directly related to canopy state, structure, and dynamics. Numerous studies have emphasized the significance of the HAND grid when studying tropical forest ecosystems. For instance, Schietti et al. [55] found a correlation between HAND and changes in floristic composition in the Amazon. Regarding structural parameters, the relationship between HAND and biomass in Eucalyptus plantations in Brazil demonstrated that the functioning and dynamics of the same tree species can vary significantly depending on drainage availability [56]. Therefore, the spatial variability of canopy height is a complex phenomenon, and this spatial dependence needs to be considered in any predictive approach.

In this perspective, the U-Net architecture on which *CHNET* is built allows accounting for both local and global contexts when mapping canopy height. This architecture enables the model to capture small-scale details and broader spatial patterns in the input data. This ability of U-Net to integrate information from multiple scales through its encoder-decoder structure is particularly beneficial in learning complex spatial relationships in the input data. Consequently, *CHNET* predictions exhibit a spatial structure that is well retrieved by the model (Fig. 3b). This explains why the residuals alone do not contain all the spatial autocorrelation, as some of it is successfully captured by the model. However, despite this capacity for spatial feature learning, *CHNET* does not directly incorporate all the spatial dependencies of GEDI reference canopy heights. These canopy heights are only used as reference for training and are not fed into the model as part of the input data that *CHNET* uses to make predictions. As a result, the model can learn spatial patterns indirectly but cannot fully represent the spatial autocorrelation of the canopy heights in its predictions. This limitation leaves some unexplained variance linked to the spatial autocorrelation of canopy heights. This remaining spatial variance is indeed found in the spatial structure observed in the residuals (Fig. 3b). Implementing a RK procedure can therefore still be relevant and useful for a spatial method like *CHNET*, which is highlighted by the improvements in accuracies presented in this study. Other studies have demonstrated the benefits of combining deep learning frameworks with spatial considerations. Liu et al. [57] employed an interpolation-based mapping strategy, combining multi-layer perceptrons with kriging interpolation, to produce a 30-m canopy height map of China, and they noted that their method allowed reducing the saturation effect of estimates in tall forests. In their canopy height model of the Earth, Lang et al. [21] used CNNs trained on S2 data and encoded geographical coordinates, and they observed that their model yielded far better results when this geographical information was incorporated compared to S2 data alone. To integrate spatial information, they designed their model with the ability to learn geographical priors, by feeding it geographical coordinates (in a suitable cyclic encoding) as additional input channels. By doing so, the model could better understand spatial relationships and contextual information in relation to S2 reflectance values, leading to more accurate and reliable canopy height predictions.

Conversely, *RFH* treats each data point independently and does not incorporate any spatial correlations beyond those already present in the input data. These limitations are apparent in the initial accuracies of *RFH*, which is outperformed by *CHNET* (Table II). This is also particularly evident in the representation of the spatial autocorrelation of

RFH predictions, which exhibit almost no spatial structure (Fig. 3b). This indicates that the model fails to capture the spatial correlations in GEDI reference canopy heights. These spatial correlations can be more complex than simply linking raw input data to canopy heights. In the analysis of their global canopy height model, Lang et al. [21] observed that although there is a clear correlation between classical vegetation indexes like NDVI and the estimated canopy heights, the relationships between image features and canopy height are much more complex. In another study for country-scale canopy height mapping using 10-m resolution S2 images, Lang et al. [56] demonstrated the important role of textural features that correlate with vegetation height. To confirm their hypothesis, they implemented their CNN model by setting the spatial size of all convolution kernels to 1×1, thus forcing it to treat each pixel independently and preventing it from learning any texture or spatial context. In this configuration, the reported errors grew significantly, especially for high canopy heights in the range of 40-60 m. These results demonstrate the benefit of considering textual and spatial features when dealing with high-resolution images, particularly in areas of very high vegetation like French Guiana. These features are not directly leveraged by a technique like Random Forest in the context of this study. As a result, the remaining spatial correlation, which is not accounted for by *RFH*, is instead in the residuals, where the most pronounced spatial structure can be observed (Fig. 3b). Contrary to *CHNET* predictions, where some spatial autocorrelation is found, it is entirely contained in the residuals for *RFH*. Implementing a spatial interpolation technique such as RK is therefore even more important for a non-spatial method like *RFH*. Several studies have demonstrated the importance of integrating spatial information for more accurate and reliable predictions. For example, Wang et al. [29] implemented a spatially-weighted geographical Random Forest model which outperformed the traditional Random Forest method, reaching the conclusion that the effects of spatial non-stationarity need to be accounted for in the modeling process.

In our study, the importance of integrating spatial information in canopy height mapping is clear when examining the spatial behaviors of both *CHNET* and *RFH*. The fitted semivariograms of the residuals provide useful insights in this regard. For *RFH*, the residuals show more spatial structure than those of *CHNET* in general. Specifically, the range for *RFH* is larger by more than half a kilometer compared to *CHNET*, indicating that spatial dependencies persist over greater distances for the residuals of *RFH*. This is because *RFH* leaves more spatial correlation unaccounted for in its predictions. Additionally, the semivariogram of *RFH* residuals displays more dynamics, as highlighted by its partial sill (i.e., the difference between sill and nugget) of 3.3 m² compared to 2.0 m² for *CHNET*. This further illustrates how *RFH* residuals contain more spatial structure. Nevertheless, RK provides improvements for both methods. When evaluating the impact of RK, the first observation is that it is effective at unbiasing the results, particularly for *RFH*. Since RK adjusts canopy height predictions based on the residuals, it allows reducing overestimation and underestimation, which are common issues in studies using GEDI data for canopy height mapping in tropical ecosystems. By adding kriged residuals to the initial predictions, underestimated values are increased and overestimated values are reduced, which tends to reduce the bias. Another key aspect of RK corrections is that they are primarily localized around the available GEDI sample points. Conversely, areas lacking GEDI information tend to benefit less from the corrections. As RK is an interpolation method, its effectiveness strongly depends on the density of sample points. Fayad et al. [33] observed in their study using airborne LiDAR for canopy height mapping that spatial sampling had a significant impact on the results obtained with kriging. Indeed, they found that the accuracy of their canopy height map of French Guiana decreased as the spacing between LiDAR flight lines increased. Their results proved to be highly sensitive to the spatial sampling of the reference LiDAR dataset. Similarly in our study, because GEDI measurements are sparse and unevenly distributed, the spatial density of GEDI points is a key factor affecting the overall results obtained with RK.

In the end, integrating kriged residuals through RK allows reaching improvements in the accuracies of canopy height estimates. These improvements are more substantial for *RFH*, as it inherently lacks spatial awareness by itself, but they are still interesting for *CHNET*. Even if CNNs inherently capture some spatial patterns, their architecture alone is not sufficient to fully address the spatial complexity of the data. RK can therefore be a valuable addition, even if it is less impactful than it is for Random Forest. Essentially, when assessed against GEDI data, the combination of *RFH* with RK (*RFH-RK*) produces slightly better results than those obtained from the *CHNET* model alone. This indicates that adding spatial information to Random Forest can bridge the performance gap between a non-spatial method and a convolutional approach that integrates spatial context by design. Other studies have shown that RK can improve the results of Random Forest models for canopy height estimation [33], [58]. This leads to the conclusion that a Random Forest model augmented with spatial information can emulate the capabilities of a convolutional approach in terms of accuracy. This improvement is only possible if the GEDI data are properly distributed, meaning at distances smaller than the range of the spatial autocorrelation and with sufficient spatial density. Kriging is thus a workaround for an imperfect model when it comes to accounting for the spatial autocorrelation in the reference data, but it is only truly effective when there are enough data points. This argues in favor of more usable GEDI measurements than are currently available to achieve truly effective RK for *CHNET* and *RFH*.

V. Conclusion

This study analyzes and addresses the impacts that sensor-induced anisotropies can have on canopy height estimation using GEDI data, particularly in complex tropical environments like French Guiana. These anisotropies arise from the GEDI sensor's acquisition configuration, mainly from the differing laser beam strengths combined with the spatial sampling of the measurements, which introduces spatial inconsistencies that do not reflect the true spatial

autocorrelation of canopy height. Specifically, the variation in measurements between power and coverage beams, given the fact that power beams offer superior penetration through dense vegetation, leads to discrepancies in the spatial data, which appear as periodic patterns in the residuals of the canopy height prediction models.

Ensuring the spatial consistency of canopy height measurements is essential to derive accurate estimates. It is also important to account for the spatial behavior of canopy height, especially since it is a continuous and environment-sensitive variable. Our analyses revealed that refining GEDI reference data to focus exclusively on power beams belonging to a given azimuthal direction, in order to perform the subsequent spatial analysis along that corresponding direction, removed anisotropic patterns and resulted in a more reliable representation of the true spatial autocorrelation of canopy heights. However, a more systematic approach to handling these sensor-related effects from the beginning, possibly through pre-processing techniques, could further enhance the quality of the reference data used to calibrate regression models.

Regarding the integration of spatial correlation for canopy height estimation, our study showed that regression approaches can benefit from spatial interpolation techniques to improve their estimates. We explored the addition of RK to both *CHNET* and *RFH* models and the results indicated that the incorporation of kriged residuals allowed improving the accuracies of the estimates. These improvements were more noticeable for the Random Forest algorithm compared to the U-Net architecture. However, for both methods, the corrections are localized around GEDI sample points. The density of available GEDI information is therefore a major determinant in the effectiveness of RK. Finally, our findings suggest that integrating spatial information into non-spatial models like Random Forest can yield results comparable to those achieved by inherently spatial architectures like CNNs. Consequently, a combination of Random Forest with spatial integration methods could serve as an alternative to CNNs for canopy height estimation.

ACKNOWLEDGMENT

The authors would like to thank the GEDI team and NASA's LP DAAC (Land Processes Distributed Active Archive Center) for proving the GEDI data. The authors also acknowledge Caroline Bedeau (ONF) and Ludovic Villard (CESBIO) for providing additional data used in this study.

REFERENCES

[1] Y. Pan *et al.*, "A Large and Persistent Carbon Sink in the World's Forests," *Science*, vol. 333, no. 6045, pp. 988–993, Aug. 2011, doi: 10.1126/science.1201609.

[2] N. L. Harris *et al.*, "Global maps of twenty-first century forest carbon fluxes," *Nat. Clim. Chang.*, vol. 11, no. 3, pp. 234–240, Mar. 2021, doi: 10.1038/s41558-020-00976-6.

[3] P. Friedlingstein *et al.*, "Global Carbon Budget 2023," *Earth Syst. Sci. Data*, vol. 15, no. 12, pp. 5301–5369, Dec. 2023, doi: 10.5194/essd-15-5301-2023.

[4] J. Chave *et al.*, "Tree allometry and improved estimation of carbon stocks and balance in tropical forests," *Oecologia*, vol. 145, no. 1, pp. 87–99, Aug. 2005, doi: 10.1007/s00442-005-0100-x.

[5] G. P. Asner and J. Mascaro, "Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric," *Remote Sensing of Environment*, vol. 140, pp. 614–624, Jan. 2014, doi: 10.1016/j.rse.2013.09.023.

[6] M. Bayrak and L. Marafa, "Ten Years of REDD+: A Critical Review of the Impact of REDD+ on Forest-Dependent Communities," *Sustainability*, vol. 8, no. 7, p. 620, Jul. 2016, doi: 10.3390/su8070620.

[7] A. E. Duchelle, G. Simonet, W. D. Sunderlin, and S. Wunder, "What is REDD+ achieving on the ground?," *Current Opinion in Environmental Sustainability*, vol. 32, pp. 134–140, Jun. 2018, doi: 10.1016/j.cosust.2018.07.001.

[8] I. Fayad *et al.*, "Canopy Height Estimation in French Guiana with LiDAR ICESat/GLAS Data Using Principal Component Analysis and Random Forest Regressions," *Remote Sensing*, vol. 6, no. 12, pp. 11883–11914, Nov. 2014, doi: 10.3390/rs61211883.

[9] M. R. Pourrahmati *et al.*, "Capability of GLAS/ICESat Data to Estimate Forest Canopy Height and Volume in Mountainous Forests of Iran," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 8, no. 11, pp. 5246–5261, Nov. 2015, doi: 10.1109/JSTARS.2015.2478478.

[10] N. N. Baghdadi, M. El Hajj, J.-S. Bailly, and F. Fabre, "Viability Statistics of GLAS/ICESat Data Acquired Over Tropical Forests," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 7, no. 5, pp. 1658–1664, May 2014, doi: 10.1109/JSTARS.2013.2273563.

[11] I. Fayad, N. Baghdadi, and J. Riedi, "Quality Assessment of Acquired GEDI Waveforms: Case Study over France, Tunisia and French Guiana," *Remote Sensing*, vol. 13, no. 16, p. 3144, Aug. 2021, doi: 10.3390/rs13163144.

[12] M. Rajab Pourrahmati *et al.*, "Effects of Eucalyptus plantation characteristics and environmental factors on GEDI waveform metrics," *International Journal of Remote Sensing*, vol. 45, no. 11, pp. 3737–3763, Jun. 2024, doi: 10.1080/01431161.2024.2354072.

[13] R. Dubayah *et al.*, "The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography," *Science of Remote Sensing*, vol. 1, p. 100002, Jun. 2020, doi: 10.1016/j.srs.2020.100002.

[14] K. Lahssini, N. Baghdadi, G. Le Maire, and I. Fayad, "Influence of GEDI Acquisition and Processing Parameters on Canopy Height Estimates over Tropical Forests," *Remote Sensing*, vol. 14, no. 24, p. 6264, Dec. 2022, doi: 10.3390/rs14246264.

[15] P. Potapov *et al.*, "Mapping global forest canopy height through integration of GEDI and Landsat data," *Remote Sensing of Environment*, vol. 253, p. 112165, Feb. 2021, doi: 10.1016/j.rse.2020.112165.

[16] D. Morin, M. Planells, S. Mermoz, and F. Mouret, "Estimation of forest height and biomass from open-

access multi-sensor satellite imagery and GEDI Lidar data: high-resolution maps of metropolitan France," 2023, *arXiv*. doi: 10.48550/ARXIV.2310.14662.

[17] D. Hong *et al.*, "More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 59, no. 5, pp. 4340–4354, May 2021, doi: 10.1109/TGRS.2020.3016820.

[18] S. A. A. Shah, M. A. Manzoor, and A. Bais, "Canopy Height Estimation at Landsat Resolution Using Convolutional Neural Networks," *MAKE*, vol. 2, no. 1, pp. 23–36, Feb. 2020, doi: 10.3390/make2010003.

[19] I. Fayad *et al.*, "A CNN-based approach for the estimation of canopy heights and wood volume from GEDI waveforms," *Remote Sensing of Environment*, vol. 265, p. 112652, Nov. 2021, doi: 10.1016/j.rse.2021.112652.

[20] S. Illarionova, D. Shadrin, V. Ignatiev, S. Shayakhmetov, A. Trekin, and I. Oseledets, "Estimation of the Canopy Height Model From Multispectral Satellite Imagery With Convolutional Neural Networks," *IEEE Access*, vol. 10, pp. 34116–34132, 2022, doi: 10.1109/ACCESS.2022.3161568.

[21] N. Lang, W. Jetz, K. Schindler, and J. D. Wegner, "A high-resolution canopy height model of the Earth," *Nat Ecol Evol*, vol. 7, no. 11, pp. 1778–1789, Sep. 2023, doi: 10.1038/s41559-023-02206-6.

[22] M. Schwartz *et al.*, "High-resolution canopy height map in the Landes forest (France) based on GEDI, Sentinel-1, and Sentinel-2 data with a deep learning approach," *International Journal of Applied Earth Observation and Geoinformation*, vol. 128, p. 103711, Apr. 2024, doi: 10.1016/j.jag.2024.103711.

[23] K. Lahssini, N. Baghdadi, G. L. Maire, I. Fayad, and G. Vincent, "Integrating Multi-Source Satellite Data and Environmental Information in a U-Net Architecture for Canopy Height Mapping in French Guiana," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, Athens, Greece: IEEE, Jul. 2024, pp. 5300–5304. doi: 10.1109/IGARSS53475.2024.10641577.

[24] E. T. A. Mitchard *et al.*, "Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud," *Biogeosciences*, vol. 9, no. 1, pp. 179–191, Jan. 2012, doi: 10.5194/bg-9-179-2012.

[25] I. Fayad, N. Baghdadi, and K. Lahssini, "An Assessment of the GEDI Lasers' Capabilities in Detecting Canopy Tops and Their Penetration in a Densely Vegetated, Tropical Area," *Remote Sensing*, vol. 14, no. 13, p. 2969, Jun. 2022, doi: 10.3390/rs14132969.

[26] M. Simard, N. Pinto, J. B. Fisher, and A. Baccini, "Mapping forest canopy height globally with spaceborne lidar," *J. Geophys. Res.*, vol. 116, no. G4, p. G04021, Nov. 2011, doi: 10.1029/2011JG001708.

[27] M. C. Hansen *et al.*, "Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data," *Remote Sensing of Environment*, vol. 185, pp. 221–232, Nov. 2016, doi: 10.1016/j.rse.2016.02.023.

[28] S. Mermoz, T. Le Toan, L. Villard, M. Réjou-Méchain, and J. Seifert-Granzin, "Biomass assessment in the Cameroon savanna using ALOS PALSAR data," *Remote Sensing of Environment*, vol. 155, pp. 109–119, Dec. 2014, doi: 10.1016/j.rse.2014.01.029.

[29] H. Wang, T. Seaborn, Z. Wang, C. C. Caudill, and T. E. Link, "Modeling tree canopy height using machine learning over mixed vegetation landscapes," *International Journal of Applied Earth Observation and Geoinformation*, vol. 101, p. 102353, Sep. 2021, doi: 10.1016/j.jag.2021.102353.

[30] L. Laiti, D. Zardi, M. De Franceschi, and G. Rampanelli, "Residual kriging analysis of airborne measurements: application to the mapping of atmospheric boundary-layer thermal structures in a mountain valley," *Atmospheric Science Letters*, vol. 14, no. 2, pp. 79–85, Apr. 2013, doi: 10.1002/asl2.420.

[31] H. Alsamamra, J. A. Ruiz-Arias, D. Pozo-Vázquez, and J. Tovar-Pescador, "A comparative study of ordinary and residual kriging techniques for mapping global solar radiation over southern Spain," *Agricultural and Forest Meteorology*, vol. 149, no. 8, pp. 1343–1357, Aug. 2009, doi: 10.1016/j.agrformet.2009.03.005.

[32] T. Gia Pham, M. Kappas, C. Van Huynh, and L. Hoang Khanh Nguyen, "Application of Ordinary Kriging and Regression Kriging Method for Soil Properties Mapping in Hilly Region of Central Vietnam," *IJGI*, vol. 8, no. 3, p. 147, Mar. 2019, doi: 10.3390/ijgi8030147.

[33] I. Fayad *et al.*, "Regional Scale Rain-Forest Height Mapping Using Regression-Kriging of Spaceborne and Airborne LiDAR Data: Application on French Guiana," *Remote Sensing*, vol. 8, no. 3, p. 240, Mar. 2016, doi: 10.3390/rs8030240.

[34] I. Fayad *et al.*, "Aboveground biomass mapping in French Guiana by combining remote sensing, forest inventories and environmental data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 52, pp. 502–514, Oct. 2016, doi: 10.1016/j.jag.2016.07.015.

[35] S. Guitet, J.-F. Cornu, O. Brunaux, J. Betbeder, J.-M. Carozza, and C. Richard-Hansen, "Landform and landscape mapping, French Guiana (South America)," *Journal of Maps*, vol. 9, no. 3, pp. 325–335, Sep. 2013, doi: 10.1080/17445647.2013.785371.

[36] S. Guitet, R. Pélissier, O. Brunaux, G. Jaouen, and D. Sabatier, "Geomorphological landscape features explain floristic patterns in French Guiana rainforest," *Biodivers Conserv*, vol. 24, no. 5, pp. 1215–1237, May 2015, doi: 10.1007/s10531-014-0854-8.

[37] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, "Present and future Köppen-Geiger climate classification maps at 1-km resolution," *Sci Data*, vol. 5, no. 1, p. 180214, Oct. 2018, doi: 10.1038/sdata.2018.214.

[38] R. Dubayah, M. Hofton, J. Blair, J. Armston, H. Tang, and S. Luthcke, "GEDI L2A Elevation and Height Metrics Data Global Footprint Level V002." NASA

EOSDIS Land Processes Distributed Active Archive Center, 2021. doi: 10.5067/GEDI/GEDI02_A.002.

[39] I. Fayad, N. Baghdadi, and F. Frappart, "Comparative Analysis of GEDI's Elevation Accuracy from the First and Second Data Product Releases over Inland Waterbodies," *Remote Sensing*, vol. 14, no. 2, p. 340, Jan. 2022, doi: 10.3390/rs14020340.

[40] R. Cresson and N. Saint-Geours, "Natural Color Satellite Image Mosaicking Using Quadratic Programming in Decorrelated Color Space," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 8, no. 8, pp. 4151–4162, Aug. 2015, doi: 10.1109/JSTARS.2015.2449233.

[41] C. D. Rennó *et al.*, "HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia," *Remote Sensing of Environment*, vol. 112, no. 9, pp. 3469–3481, Sep. 2008, doi: 10.1016/j.rse.2008.03.018.

[42] *Catalogue des habitats forestiers de Guyane*. Cayenne (Guyane): Direction régionale ONF Guyane, 2015.

[43] M. Adam, M. Urbazaev, C. Dubois, and C. Schmullius, "Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters," *Remote Sensing*, vol. 12, no. 23, p. 3948, Dec. 2020, doi: 10.3390/rs12233948.

[44] C. Hilbert and C. Schmullius, "Influence of Surface Topography on ICESat/GLAS Forest Height Estimation and Waveform Shape," *Remote Sensing*, vol. 4, no. 8, pp. 2210–2235, Jul. 2012, doi: 10.3390/rs4082210.

[45] K. Lahssini, N. Baghdadi, G. Le Maire, S. Dupuy, and I. Fayad, "Use of GEDI Signal and Environmental Parameters to Improve Canopy Height Estimation over Tropical Forest Ecosystems in Mayotte Island," *Canadian Journal of Remote Sensing*, vol. 50, no. 1, p. 2351004, Dec. 2024, doi: 10.1080/07038992.2024.2351004.

[46] M. Gazzea, A. Solheim, and R. Arghandeh, "High-resolution mapping of forest structure from integrated SAR and optical images using an enhanced U-net method," *Science of Remote Sensing*, vol. 8, p. 100093, Dec. 2023, doi: 10.1016/j.srs.2023.100093.

[47] F. H. Wagner *et al.*, "Sub-meter tree height mapping of California using aerial images and LiDAR-informed U-Net model," *Remote Sensing of Environment*, vol. 305, p. 114099, May 2024, doi: 10.1016/j.rse.2024.114099.

[48] S. Ge, H. Gu, W. Su, J. Praks, and O. Antropov, "Improved Semisupervised UNet Deep Learning Model for Forest Height Mapping With Satellite SAR and Optical Data," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 15, pp. 5776–5787, 2022, doi: 10.1109/JSTARS.2022.3188201.

[49] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[50] S. Ahmed and G. De Marsily, "Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity," *Water Resources Research*, vol. 23, no. 9, pp. 1717–1737, Sep. 1987, doi: 10.1029/WR023i009p01717.

[51] P. Goovaerts, *Geostatistics for natural resources evaluation*, vol. 483. Oxford University Press, 1997.

[52] E. Kutchartt, M. Pedron, and F. Pirotti, "ASSESSMENT OF CANOPY AND GROUND HEIGHT ACCURACY FROM GEDI LIDAR OVER STEEP MOUNTAIN AREAS," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. V-3–2022, pp. 431–438, May 2022, doi: 10.5194/isprs-annals-V-3-2022-431-2022.

[53] C. Wang *et al.*, "Factors affecting relative height and ground elevation estimations of GEDI among forest types across the conterminous USA," *GIScience & Remote Sensing*, vol. 59, no. 1, pp. 975–999, Dec. 2022, doi: 10.1080/15481603.2022.2085354.

[54] A. Liu, X. Cheng, and Z. Chen, "Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals," *Remote Sensing of Environment*, vol. 264, p. 112571, Oct. 2021, doi: 10.1016/j.rse.2021.112571.

[55] J. Schietti *et al.*, "Vertical distance from drainage drives floristic composition changes in an Amazonian rainforest," *Plant Ecology & Diversity*, vol. 7, no. 1–2, pp. 241–253, Apr. 2014, doi: 10.1080/17550874.2013.783642.

[56] J. L. Stape and C. A. Alvares, "Physiographic position drives Eucalyptus productivity in Mato Grosso do Sul, Brazil," *ST IPEF*, vol. 26, no. 48, pp. 289–293, 2023, doi: 10.18671/sertec.v26n48.057.

[57] X. Liu *et al.*, "Neural network guided interpolation for mapping canopy height of China's forests by integrating GEDI and ICESat-2 data," *Remote Sensing of Environment*, vol. 269, p. 112844, Feb. 2022, doi: 10.1016/j.rse.2021.112844.

[58] J. Zhao, L. Zhao, E. Chen, Z. Li, K. Xu, and X. Ding, "An Improved Generalized Hierarchical Estimation Framework with Geostatistics for Mapping Forest Parameters and Its Uncertainty: A Case Study of Forest Canopy Height," *Remote Sensing*, vol. 14, no. 3, p. 568, Jan. 2022, doi: 10.3390/rs14030568.

Chapter 6

Conclusions and Perspectives

6.1 Conclusions

The research presented in this dissertation demonstrates the potential of spaceborne LiDAR, particularly GEDI, in estimating canopy height in tropical forests. However, several challenges and limitations also emerged and provide important insights for future studies. This section summarizes the key findings in terms of potential and challenges of GEDI data for canopy height estimation, focusing on the main results from the four studies conducted during this doctoral thesis.

6.1.1 Potential

The potential of GEDI data to derive accurate canopy height estimates in tropical forests globally has been highlighted through multiple aspects of the data explored across the studies.

First, GEDI's ability to retrieve data over dense canopies is a significant advantage in tropical ecosystems, where ground measurements are difficult or impossible to obtain due to inaccessibility. This feature, which was assessed in both Chapters 2 and 3, allows for reliable canopy height measurements when appropriate settings are used. GEDI power beams, as opposed to coverage beams, offer a higher degree of vegetation penetration and accuracy, especially in contexts characterized by high AGB levels. This is particularly important in dense tropical ecosystems such as those of French Guiana, where the vertical structure of the vegetation is complex. The results indicate that the choice of beam type and algorithm setting group (with group number 5 being optimal) substantially enhance the precision of canopy height estimation through a single GEDI metric such as *rh_95*. Additionally, statistical models, as implemented in Chapters 2 and 3, prove to be useful in improving the accuracy of canopy height estimates beyond the utilization of a single GEDI relative height metric.

Moreover, in Chapter 4, the use of deep learning models (U-Net) demonstrates how multi-source data fusion with GEDI can be used to optimize canopy height mapping at a regional scale. By integrating GEDI data with other remote sensing sources like optical and radar information, and considering ancillary geomorphological and hydrological factors, the accuracy of canopy height mapping can be significantly enhanced. This approach enabled the

production of a comprehensive canopy height map of French Guiana at a 10-m resolution. These results underscore the benefits of GEDI measurements when they are combined with robust models and multimodal data. Accurate and reliable canopy height maps are crucial for quantifying carbon stocks and monitoring forest dynamics on a larger scale, particularly in the context of the REDD+ initiative and global forest management efforts.

Finally, the incorporation of spatial techniques like residual kriging, as explored in Chapter 5, adds a relevant dimension to GEDI-based models. These methods improve the spatial consistency of canopy height estimates, especially in tropical environments where the spatial variability of canopy height is complex and influenced by many factors. Classical non-spatial models like Random Forest particularly benefit from such spatial corrections which allow them to achieve comparable performance to more complex CNN-based architectures. These results show that GEDI, when combined with spatial interpolation methods, holds the potential for accurate and region-wide canopy height predictions, suitable for operational forest monitoring.

6.1.2 Challenges

Despite its considerable potential, the use of GEDI for canopy height estimation also faces several limitations, both inherent to the technology and related to the challenges of working in tropical forests.

A key challenge, as discussed in Chapters 2 and 3, is GEDI's sensitivity to acquisition parameters and environmental conditions. GEDI's physical signal settings, particularly beam type and sensitivity, significantly influence the accuracy of canopy height retrievals. For instance, coverage beams, with their lower power, are less capable of penetrating dense canopies and can fail to accurately detect ground returns, particularly in areas where AGB is high. Consequently, estimates derived from these beams are generally less reliable and the use of power beams yields more consistent and accurate canopy height measurements. Additionally, the GEDI acquisition process introduces geolocation errors that can negatively impact the canopy height measurements, especially in heterogeneous areas. For example, in the Mayotte study (Chapter 3), these errors accounted for approximately half of the total estimation error. Since Mayotte is also characterized by complex tropical terrains with steep slopes, these errors were exacerbated, which further reduced the precision of canopy height estimates.

Another limitation of GEDI is the reduced laser penetration depth in some tropical contexts. Chapter 3 highlighted that in the forests of Mayotte, laser penetration depth decreased from the expected 30 m to around 15 m. This limitation, which is linked to GEDI's signal properties,

indicates that the difficulty in capturing complete vertical profiles of tropical canopies is not only due to the density of the vegetation, but also to forest type and structure. Similarly, environmental factors like terrain slope can distort waveforms and lead to inaccurate canopy height estimates.

Additionally, spatial anisotropy induced by GEDI's acquisition configuration poses an important challenge, as presented in Chapter 5. The variations in laser power and the azimuthal orientation of the acquisition tracks introduce periodic spatial patterns in the canopy height measurements that do not represent the actual spatial variability of the canopy. These anisotropic effects must be mitigated to consider the true spatial autocorrelation of GEDI canopy heights when implementing post-processing correction methods like residual kriging. However, such techniques are computationally intensive and strongly dependent on the density of sample points, which limits their operational applicability at larger scales. The current density of usable GEDI points allows for only a slight improvement in accuracy. This argues in favor of a higher point density for kriging to substantially improve the accuracy of canopy height estimates.

Lastly, the availability of reference data remains essential for accurate canopy height estimation. Chapter 4 highlighted the dependence of deep learning models on sufficient high-quality reference data for model training. In regions where the amount of usable GEDI data is sparser, the performance of GEDI-based models and spatial interpolation techniques automatically deteriorates. Depending on the accuracy needed for operational GEDI-based applications, data filtering steps can result in significantly reduced amounts of usable data. Similarly, there is a need for high-quality independent validation datasets, such as airborne LiDAR, to assess the performance of canopy height prediction models and ensure that these models generalize well. Such datasets can also be leveraged to refine GEDI information and thus provide models with more reliable reference inputs for their training.

6.2 Perspectives

6.2.1 From canopy height to AGB

The main objective of this doctoral thesis was to map canopy height across French Guiana, and it was successfully achieved with promising results. However, canopy height itself is not the end goal, as it can be used to estimate other important forest attributes, such as AGB and aboveground carbon density (ACD). These variables are the ones used to assess carbon stocks in tropical forests and to understand the role of forests in climate change mitigation.

AGB is traditionally estimated using allometric equations that account for variables such as tree height, DBH, and wood density. These equations, applied at the individual tree level, are crucial for deriving AGB estimates in studies based on field measurements. However, the GEDI-derived canopy height estimates in this thesis refer to the maximum height of forest plots (i.e., GEDI footprints) rather than individual trees. This introduces some challenges because DBH and wood density cannot be directly extracted from spaceborne LiDAR data, which means that these parameters are absent from direct AGB estimation using LiDAR.

Despite these limitations, many studies have demonstrated strong correlations between AGB and LiDAR-derived canopy height estimates. Nevertheless, the remaining challenge is that these relationships often exhibit large uncertainties, especially in large-scale AGB estimation. This is due to the fact that AGB is influenced by a combination of factors beyond height, such as forest type, site-specific conditions, and structural complexity. As a result, using canopy height alone to estimate AGB may not produce accurate results across different regions. In order to enhance AGB estimation, it is necessary to develop approaches that account for other variables beyond canopy height. Studies have shown that different LiDAR metrics can yield varying correlations with AGB. Investigating the relationships between different LiDAR metrics rather than relying on a single estimation of canopy height could improve the accuracy of AGB estimates. Additionally, the incorporation of ancillary parameters such as topography, forest type, and tree cover could further enhance the precision of AGB models.

In conclusion, while canopy height serves as an effective proxy for AGB, future research should focus on refining allometric models that leverage canopy height by improving the integration of LiDAR-derived information with ancillary environmental variables.

6.2.2 Better characterizing dense forest ecosystems

One of the major challenges in accurately characterizing AGB in dense tropical forest ecosystems is the issue of sensor saturation. Sensor saturation occurs when a remote sensing system reaches its limit in detecting variations in the forest structure, particularly in areas with high AGB. This issue severely affects the ability to assess AGB in densely vegetated ecosystems, which represent a significant portion of the world's forests. As a result, the global understanding of forest carbon stocks is compromised, as much of the AGB information from these areas remains inaccessible due to the limitations of remote sensing technologies.

In this doctoral thesis, GEDI's laser penetration was a central focus. While GEDI offers significant advantages to measure canopy height and structure, it may struggle in very dense tropical ecosystems. Furthermore, saturation is not specific to LiDAR systems. Optical

sensors, which rely on reflectance data, also suffer from significant limitations when it comes to characterizing and discriminating dense canopies. Similarly, radar systems, while offering some penetration capabilities, experience saturation, particularly at lower wavelengths.

This saturation issue is particularly problematic because dense tropical forests contain some of the largest carbon reserves in the world. The inability to retrieve accurate AGB data from these regions introduces significant uncertainties into global AGB models and carbon stock estimates. The global AGB estimates used nowadays are therefore not as accurate or reliable as they need to be. Addressing the problem of sensor saturation is not easily achieved through modeling approaches. Since this issue is inherent to the data acquisition process itself, models cannot fully compensate by themselves for the saturated data resulting from sensor limitations. Thus, the solution to this problem lies mainly in the development and deployment of more advanced remote sensing technologies that can better capture the vertical structure of dense forests.

In this perspective, there are several planned and upcoming missions that aim to address these challenges and provide more accurate measurements over dense forest ecosystems. Notably, the BIOMASS mission (ESA) will use P-band radar, which has a much longer wavelength than current radar systems. This longer wavelength is expected to offer greater penetration capabilities and is promising for improving AGB estimates in tropical regions. Similarly, the NISAR mission (NASA and ISRO) will deploy a dual-frequency radar (L-band and S-band), which will provide additional data for monitoring forest structure. Although radar systems have limitations in dense environments, the complementary use of different frequency bands can help mitigate some of the saturation issues by providing a broader range of forest structure information.

In conclusion, although sensor saturation remains a significant challenge for accurately estimating AGB at a global scale, planned future remote sensing missions offer promising solutions. They are expected to provide new data that will improve the accuracy and reliability of AGB estimates globally, particularly in dense tropical regions.

References

Adam, M., Urbazaev, M., Dubois, C., & Schmullius, C. (2020). Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters. *Remote Sensing*, *12*(23), 3948. https://doi.org/10.3390/rs12233948

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

Baghdadi, N., Le Maire, G., Bailly, J.-S., Ose, K., Nouvellon, Y., Zribi, M., Lemos, C., & Hakamada, R. (2015). Evaluation of ALOS/PALSAR L-Band Data for the Estimation of *Eucalyptus* Plantations Aboveground Biomass in Brazil. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(8), 3802–3811. https://doi.org/10.1109/JSTARS.2014.2353661

Baraloto, C., Molto, Q., Rabaud, S., Hérault, B., Valencia, R., Blanc, L., Fine, P. V. A., & Thompson, J. (2013). Rapid Simultaneous Estimation of Aboveground Biomass and Tree Diversity Across Neotropical Forests: A Comparison of Field Inventory Methods. *Biotropica*, *45*(3), 288–298. https://doi.org/10.1111/btp.12006

Bayrak, M., & Marafa, L. (2016). Ten Years of REDD+: A Critical Review of the Impact of REDD+ on Forest-Dependent Communities. *Sustainability*, *8*(7), 620. https://doi.org/10.3390/su8070620

Bouvet, A., Mermoz, S., Ballère, M., Koleck, T., & Le Toan, T. (2018). Use of the SAR Shadowing Effect for Deforestation Detection with Sentinel-1 Time Series. *Remote Sensing*, *10*(8), 1250. https://doi.org/10.3390/rs10081250

Boyd, D. S., & Danson, F. M. (2005). Satellite remote sensing of forest resources: Three decades of research development. *Progress in Physical Geography: Earth and Environment*, *29*(1), 1–26. https://doi.org/10.1191/0309133305pp432ra

Brack, D. (2012). Excluding illegal timber and improving forest governance: The European Union's Forest Law Enforcement, Governance and Trade initiative. In *High-value natural resources and post-conflict peacebuilding* (pp. 211–220). Routledge.

Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P. W., Trisos, C., Romero, J., Aldunce, P., Barrett, K., Blanco, G., Cheung, W. W. L., Connors, S., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., … Péan, C. (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental*

Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. (First). Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647

Carcereri, D., Rizzoli, P., Ienco, D., & Bruzzone, L. (2023). A Deep Learning Framework for the Estimation of Forest Height From Bistatic TanDEM-X Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 8334–8352. https://doi.org/10.1109/JSTARS.2023.3310209

Carter, S., Herold, M., Avitabile, V., De Bruin, S., De Sy, V., Kooistra, L., & Rufino, M. C. (2018). Agriculture-driven deforestation in the tropics from 1990–2015: Emissions, trends and uncertainties. *Environmental Research Letters*, *13*(1), 014002. https://doi.org/10.1088/1748-9326/aa9ea4

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), 87–99. https://doi.org/10.1007/s00442-005-0100-x

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C., Henry, M., Martínez-Yrízar, A., Mugasha, W. A., Muller-Landau, H. C., Mencuccini, M., Nelson, B. W., Ngomanda, A., Nogueira, E. M., Ortiz-Malavassi, E., … Vieilledent, G. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, *20*(10), 3177–3190. https://doi.org/10.1111/gcb.12629

Cheng, D.-L., & Niklas, K. J. (2007). Above- and Below-ground Biomass Relationships Across 1534 Forested Communities. *Annals of Botany*, *99*(1), 95–102. https://doi.org/10.1093/aob/mcl206

Chokkalingam, U., & De Jong, W. (2001). Secondary forest: A working definition and typology. *The International Forestry Review*, 19–26.

Crowley, T. J. (2000). Causes of Climate Change Over the Past 1000 Years. *Science*, *289*(5477), 270–277. https://doi.org/10.1126/science.289.5477.270

Da Ponte, E., Fleckenstein, M., Leinenkugel, P., Parker, A., Oppelt, N., & Kuenzer, C. (2015). Tropical forest cover dynamics for Latin America using Earth observation data: A review covering the continental, regional, and local scale. *International Journal of Remote Sensing*, *36*(12), 3196–3242. https://doi.org/10.1080/01431161.2015.1058539

Dauber, E., Fredericksen, T. S., & Peña, M. (2005). Sustainability of timber harvesting in Bolivian tropical forests. *Forest Ecology and Management*, *214*(1–3), 294–304. https://doi.org/10.1016/j.foreco.2005.04.019

Dorado-Roda, I., Pascual, A., Godinho, S., Silva, C., Botequim, B., Rodríguez-Gonzálvez, P., González-Ferreiro, E., & Guerra-Hernández, J. (2021). Assessing the Accuracy of GEDI Data for Canopy Height and Aboveground Biomass Estimates in Mediterranean Forests. *Remote Sensing*, *13*(12), 2279. https://doi.org/10.3390/rs13122279

Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., & Silva, C. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, *1*, 100002. https://doi.org/10.1016/j.srs.2020.100002

Duchelle, A. E., Simonet, G., Sunderlin, W. D., & Wunder, S. (2018). What is REDD+ achieving on the ground? *Current Opinion in Environmental Sustainability*, *32*, 134–140. https://doi.org/10.1016/j.cosust.2018.07.001

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

Fagua, J. C., Jantz, P., Rodriguez-Buritica, S., Duncanson, L., & Goetz, S. J. (2019). Integrating LiDAR, Multispectral and SAR Data to Estimate and Map Canopy Height in Tropical Forests. *Remote Sensing*, *11*(22), 2697. https://doi.org/10.3390/rs11222697

Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Högberg, P., Linder, S., Mackenzie, F. T., Moore Iii, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., & Steffen, W. (2000). The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System. *Science*, *290*(5490), 291–296. https://doi.org/10.1126/science.290.5490.291

Fayad, I., Baghdadi, N., Bailly, J.-S., Barbier, N., Gond, V., Hajj, M., Fabre, F., & Bourgine, B. (2014). Canopy Height Estimation in French Guiana with LiDAR ICESat/GLAS Data Using Principal Component Analysis and Random Forest Regressions. *Remote Sensing*, *6*(12), 11883–11914. https://doi.org/10.3390/rs61211883

Fayad, I., Baghdadi, N., Bailly, J.-S., Barbier, N., Gond, V., Hérault, B., El Hajj, M., Fabre, F., & Perrin, J. (2016). Regional Scale Rain-Forest Height Mapping Using Regression-Kriging of Spaceborne and Airborne LiDAR Data: Application on French Guiana. *Remote Sensing*, *8*(3), 240. https://doi.org/10.3390/rs8030240

Fayad, I., Baghdadi, N., & Riedi, J. (2021). Quality Assessment of Acquired GEDI Waveforms: Case Study over France, Tunisia and French Guiana. *Remote Sensing*, *13*(16), 3144. https://doi.org/10.3390/rs13163144

Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J., & Millero, F. J. (2004). Impact of Anthropogenic CO_2 on the $CaCO_3$ System in the Oceans. *Science*, *305*(5682), 362–366. https://doi.org/10.1126/science.1097329

Feldpausch, T. R., Lloyd, J., Lewis, S. L., Brienen, 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., … Phillips, O. L. (2012). Tree height integrated into pantropical forest biomass estimates. *Biogeosciences*, *9*(8), 3381–3403. https://doi.org/10.5194/bg-9-3381-2012

Ferraz, A., Saatchi, S., Mallet, C., & Meyer, V. (2016). Lidar detection of individual tree size in tropical forests. *Remote Sensing of Environment*, *183*, 318–333. https://doi.org/10.1016/j.rse.2016.05.028

Ferraz, J. C., Santiago, J., & Ramos, L. (2023). Policy innovation for sustainable development: The case of the Amazon Fund. *Review of Evolutionary Political Economy*, *4*(1), 109–136. https://doi.org/10.1007/s43253-023-00092-z

Food and Agriculture Organization. (2011). *The state of forests in the Amazon Basin, Congo Basin and Southeast Asia: A report prepared for the Summit of the Three Rainforest Basins, Brazzaville, Republic of Congo, 31 May-3 June, 2011*. FAO.

Food and Agriculture Organization. (2020). *Global Forest Resources Assessment 2020*. FAO. https://doi.org/10.4060/ca8753en

Foody, G. M., Palubinskas, G., Lucas, R. M., Curran, P. J., & Honzak, M. (1996). Identifying terrestrial carbon sinks: Classification of successional stages in regenerating tropical forest from Landsat TM data. *Remote Sensing of Environment*, *55*(3), 205–216. https://doi.org/10.1016/S0034-4257(95)00196-4

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., … Zheng, B. (2023). Global Carbon Budget 2023. *Earth System Science Data*, *15*(12), 5301–5369. https://doi.org/10.5194/essd-15-5301-2023

Fuller, D. O. (2006). Tropical forest monitoring and remote sensing: A new era of transparency in forest governance? *Singapore Journal of Tropical Geography*, *27*(1), 15–29. https://doi.org/10.1111/j.1467-9493.2006.00237.x

Garestier, F., Dubois-Fernandez, P. C., & Champion, I. (2008). Forest Height Inversion Using High-Resolution P-Band Pol-InSAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(11), 3544–3559. https://doi.org/10.1109/TGRS.2008.922032

Grabska, E., Hostert, P., Pflugmacher, D., & Ostapowicz, K. (2019). Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing*, *11*(10), 1197. https://doi.org/10.3390/rs11101197

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamäki, J. V., Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campari, J., Conant, R. T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M. R., … Fargione, J. (2017). Natural climate solutions. *Proceedings of the National Academy of Sciences*, *114*(44), 11645–11650. https://doi.org/10.1073/pnas.1710465114

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Ho Tong Minh, D., Le Toan, T., Rocca, F., Tebaldini, S., Villard, L., Réjou-Méchain, M., Phillips, O. L., Feldpausch, T. R., Dubois-Fernandez, P., Scipal, K., & Chave, J. (2016). SAR tomography for the retrieval of forest biomass and height: Cross-validation at two tropical forest sites in French Guiana. *Remote Sensing of Environment*, *175*, 138–147. https://doi.org/10.1016/j.rse.2015.12.037

Holdridge, L. R., & Grenke, W. C. (1970). *Forest environments in tropical life zones: A pilot study.*

Houghton, R. A., House, J. I., Pongratz, J., Van Der Werf, G. R., DeFries, R. S., Hansen, M. C., Le Quéré, C., & Ramankutty, N. (2012). Carbon emissions from land use and land-cover change. *Biogeosciences*, *9*(12), 5125–5142. https://doi.org/10.5194/bg-9-5125-2012

Hudak, A. T., Lefsky, M. A., Cohen, W. B., & Berterretche, M. (2002). Integration of lidar and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, *82*(2–3), 397–416. https://doi.org/10.1016/S0034-4257(02)00056-1

Hunter, M. L. (1999). *Maintaining biodiversity in forest ecosystems*. Cambridge university press.

Huynh, T., Lee, D. J., Applegate, G., & Lewis, T. (2021). Field methods for above and belowground biomass estimation in plantation forests. *MethodsX*, *8*, 101192. https://doi.org/10.1016/j.mex.2020.101192

Jiang, F., Zhao, F., Ma, K., Li, D., & Sun, H. (2021). Mapping the Forest Canopy Height in Northern China by Synergizing ICESat-2 with Sentinel-2 Using a Stacking Algorithm. *Remote Sensing*, *13*(8), 1535. https://doi.org/10.3390/rs13081535

Koh, L. P., & Wilcove, D. S. (2008). Is oil palm agriculture really destroying tropical biodiversity? *Conservation Letters*, *1*(2), 60–64. https://doi.org/10.1111/j.1755-263X.2008.00011.x

Kunzmann, K. (2008). The Non-legally Binding Instrument on Sustainable Management of All Types of Forests—Towards a Legal Regime for Sustainable Forest Management? *German Law Journal*, *9*(8), 981–1006. https://doi.org/10.1017/S2071832200000250

Lahssini, K., Teste, F., Dayal, K. R., Durrieu, S., Ienco, D., & Monnet, J.-M. (2022). Combining LiDAR Metrics and Sentinel-2 Imagery to Estimate Basal Area and Wood Volume in Complex Forest Environment via Neural Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 4337–4348. https://doi.org/10.1109/JSTARS.2022.3175609

Lang, N., Jetz, W., Schindler, K., & Wegner, J. D. (2023). A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution*, *7*(11), 1778–1789. https://doi.org/10.1038/s41559-023-02206-6

Lefsky, M. A. (2010). A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System. *Geophysical Research Letters*, *37*(15), 2010GL043622. https://doi.org/10.1029/2010GL043622

Lefsky, M. A., Cohen, W. B., Harding, D. J., Parker, G. G., Acker, S. A., & Gower, S. T. (2002). Lidar remote sensing of above-ground biomass in three biomes. *Global Ecology and Biogeography*, *11*(5), 393–399. https://doi.org/10.1046/j.1466-822x.2002.00303.x

Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Del Bom Espirito-Santo, F., Hunter, M. O., & De Oliveira, R. (2005). Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, *32*(22), 2005GL023971. https://doi.org/10.1029/2005GL023971

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*(1), 3. https://doi.org/10.1186/s13021-015-0013-x

Lemaire, G., Francois, C., Soudani, K., Berveiller, D., Pontailler, J., Breda, N., Genet, H., Davi, H., & Dufrene, E. (2008). Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment*, *112*(10), 3846–3864. https://doi.org/10.1016/j.rse.2008.06.005

Li, W., Niu, Z., Shang, R., Qin, Y., Wang, L., & Chen, H. (2020). High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. *International Journal of Applied Earth Observation and Geoinformation*, *92*, 102163. https://doi.org/10.1016/j.jag.2020.102163

Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography: Earth and Environment*, *27*(1), 88–106. https://doi.org/10.1191/0309133303pp360ra

Lima, A. J. N., Suwa, R., De Mello Ribeiro, G. H. P., Kajimoto, T., Dos Santos, J., Da Silva, R. P., De Souza, C. A. S., De Barros, P. C., Noguchi, H., Ishizuka, M., & Higuchi, N. (2012). Allometric models for estimating above- and below-ground biomass in Amazonian forests at São Gabriel da Cachoeira in the upper Rio Negro, Brazil. *Forest Ecology and Management*, *277*, 163–172. https://doi.org/10.1016/j.foreco.2012.04.028

Martin, A. R., & Thomas, S. C. (2011). A Reassessment of Carbon Content in Tropical Trees. *PLoS ONE*, *6*(8), e23533. https://doi.org/10.1371/journal.pone.0023533

Masolele, R. N., De Sy, V., Herold, M., Marcos, D., Verbesselt, J., Gieseke, F., Mullissa, A. G., & Martius, C. (2021). Spatial and temporal deep learning methods for deriving land-use following deforestation: A pan-tropical case study using Landsat time series. *Remote Sensing of Environment*, *264*, 112600. https://doi.org/10.1016/j.rse.2021.112600

Matthews, G. (1993). *The carbon content of trees.* (Issue 4).

Maurer, B. A., & Whitmore, R. C. (1981). Foraging of five bird species in two forests with different vegetation structure. *The Wilson Bulletin*, 478–490.

McElhinny, C., Gibbons, P., Brack, C., & Bauhus, J. (2005). Forest and woodland stand structural complexity: Its definition and measurement. *Forest Ecology and Management*, *218*(1–3), 1–24.

Mermoz, S., Le Toan, T., Villard, L., Réjou-Méchain, M., & Seifert-Granzin, J. (2014). Biomass assessment in the Cameroon savanna using ALOS PALSAR data. *Remote Sensing of Environment*, *155*, 109–119. https://doi.org/10.1016/j.rse.2014.01.029

Mills, M. B., Malhi, Y., Ewers, R. M., Kho, L. K., Teh, Y. A., Both, S., Burslem, D. F. R. P., Majalap, N., Nilus, R., Huaraca Huasco, W., Cruz, R., Pillco, M. M., Turner, E. C., Reynolds, G., & Riutta, T. (2023). Tropical forests post-logging are a persistent net carbon source to the atmosphere. *Proceedings of the National Academy of Sciences*, *120*(3), e2214462120. https://doi.org/10.1073/pnas.2214462120

Mitchard, E. T. A. (2018). The tropical forest carbon cycle and climate change. *Nature*, *559*(7715), 527–534. https://doi.org/10.1038/s41586-018-0300-2

Molina-Valero, J. A., Camarero, J. J., Álvarez-González, J. G., Cerioni, M., Hevia, A., Sánchez-Salguero, R., Martin-Benito, D., & Pérez-Cruzado, C. (2021). Mature forests hold maximum live biomass stocks. *Forest Ecology and Management*, *480*, 118635. https://doi.org/10.1016/j.foreco.2020.118635

Murrins Misiukas, J., Carter, S., & Herold, M. (2021). Tropical Forest Monitoring: Challenges and Recent Progress in Research. *Remote Sensing*, *13*(12), 2252. https://doi.org/10.3390/rs13122252

Nakai, T., Sumida, A., Kodama, Y., & Hara, T. (2008). *Another definition of forest canopy height*. *2008*, B41A-0366. AGU Fall Meeting Abstracts.

Neuenschwander, A. L., & Magruder, L. A. (2019). Canopy and Terrain Height Retrievals with ICESat-2: A First Look. *Remote Sensing*, *11*(14), 1721. https://doi.org/10.3390/rs11141721

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), 988–993. https://doi.org/10.1126/science.1201609

Pendrill, F., Gardner, T. A., Meyfroidt, P., Persson, U. M., Adams, J., Azevedo, T., Bastos Lima, M. G., Baumann, M., Curtis, P. G., De Sy, V., Garrett, R., Godar, J., Goldman, E. D., Hansen, M. C., Heilmayr, R., Herold, M., Kuemmerle, T., Lathuillière, M. J., Ribeiro, V., … West, C. (2022). Disentangling the numbers behind agriculture-driven tropical deforestation. *Science*, *377*(6611), eabm9267. https://doi.org/10.1126/science.abm9267

Pillay, R., Venter, M., Aragon-Osejo, J., González-del-Pliego, P., Hansen, A. J., Watson, J. E., & Venter, O. (2022). Tropical forests are home to over half of the world's vertebrate species. *Frontiers in Ecology and the Environment*, *20*(1), 10–15. https://doi.org/10.1002/fee.2420

Pirotti, F. (2011). Analysis of full-waveform LiDAR data for forestry applications: A review of investigations and methods. *iForest - Biogeosciences and Forestry*, *4*(3), 100–106. https://doi.org/10.3832/ifor0562-004

Portillo-Quintero, C., Hernández-Stefanoni, J. L., Reyes-Palomeque, G., & Subedi, M. R. (2021). The Road to Operationalization of Effective Tropical Forest Monitoring Systems. *Remote Sensing*, *13*(7), 1370. https://doi.org/10.3390/rs13071370

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., & Hofton, M. (2021). Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*, 112165. https://doi.org/10.1016/j.rse.2020.112165

Pourrahmati, M. R., Baghdadi, N. N., Darvishsefat, A. A., Namiranian, M., Fayad, I., Bailly, J.-S., & Gond, V. (2015). Capability of GLAS/ICESat Data to Estimate Forest Canopy Height and Volume in Mountainous Forests of Iran. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(11), 5246–5261. https://doi.org/10.1109/JSTARS.2015.2478478

Prentice, I. C., Farquhar, G., Fasham, M., Goulden, M. L., Heimann, M., Jaramillo, V., Kheshgi, H., Le Quéré, C., Scholes, R. J., & Wallace, D. W. (2001). The carbon cycle and

atmospheric carbon dioxide. *Climate Change 2001: The Scientific Basis, Intergovernmental Panel on Climate Change*.

Quegan, S., Le Toan, T., Chave, J., Dall, J., Exbrayat, J.-F., Minh, D. H. T., Lomas, M., D'Alessandro, M. M., Paillou, P., Papathanassiou, K., Rocca, F., Saatchi, S., Scipal, K., Shugart, H., Smallman, T. L., Soja, M. J., Tebaldini, S., Ulander, L., Villard, L., & Williams, M. (2019). The European Space Agency BIOMASS mission: Measuring forest above-ground biomass from space. *Remote Sensing of Environment*, *227*, 44–60. https://doi.org/10.1016/j.rse.2019.03.032

Rahman, M. M., & Sumantyo, J. T. S. (2010). Mapping tropical forest cover and deforestation using synthetic aperture radar (SAR) images. *Applied Geomatics*, *2*(3), 113–121. https://doi.org/10.1007/s12518-010-0026-9

Rajab Pourrahmati, M., Baghdadi, N., & Fayad, I. (2023). Comparison of GEDI LiDAR Data Capability for Forest Canopy Height Estimation over Broadleaf and Needleleaf Forests. *Remote Sensing*, *15*(6), 1522. https://doi.org/10.3390/rs15061522

Robinson, S. K., & Holmes, R. T. (1982). Foraging behavior of forest birds: The relationships among search tactics, diet, and habitat structure. *Ecology*, *63*(6), 1918–1931.

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S., Wallace, D. W. R., Tilbrook, B., Millero, F. J., Peng, T.-H., Kozyr, A., Ono, T., & Rios, A. F. (2004). The Oceanic Sink for Anthropogenic CO_2. *Science*, *305*(5682), 367–371. https://doi.org/10.1126/science.1097403

Sanchez-Azofeifa, A., Antonio Guzmán, J., Campos, C. A., Castro, S., Garcia-Millan, V., Nightingale, J., & Rankine, C. (2017). Twenty-first century remote sensing technologies are revolutionizing the study of tropical forests. *Biotropica*, *49*(5), 604–619. https://doi.org/10.1111/btp.12454

Schwartz, M., Ciais, P., De Truchis, A., Chave, J., Ottlé, C., Vega, C., Wigneron, J.-P., Nicolas, M., Jouaber, S., Liu, S., Brandt, M., & Fayad, I. (2023). FORMS: Forest Multiple Source height, wood volume, and biomass maps in France at 10 to 30 m resolution based on Sentinel-1, Sentinel-2, and Global Ecosystem Dynamics Investigation (GEDI) data with a deep learning approach. *Earth System Science Data*, *15*(11), 4927–4945. https://doi.org/10.5194/essd-15-4927-2023

Seymour, F., & Busch, J. (2016). *Why forests? Why now?: The science, economics, and politics of tropical forests and climate change*. Brookings Institution Press.

Shah, S. A. A., Manzoor, M. A., & Bais, A. (2020). Canopy Height Estimation at Landsat Resolution Using Convolutional Neural Networks. *Machine Learning and Knowledge Extraction*, *2*(1), 23–36. https://doi.org/10.3390/make2010003

Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research*, *116*(G4), G04021. https://doi.org/10.1029/2011JG001708

Sothe, C., Gonsamo, A., Lourenço, R. B., Kurz, W. A., & Snider, J. (2022). Spatially Continuous Mapping of Forest Canopy Height in Canada by Combining GEDI and ICESat-2 with PALSAR and Sentinel. *Remote Sensing*, *14*(20), 5158. https://doi.org/10.3390/rs14205158

Thomas, S. C., & Martin, A. R. (2012). Carbon Content of Tree Tissues: A Synthesis. *Forests*, *3*(2), 332–352. https://doi.org/10.3390/f3020332

Trisasongko, B. H., & Paull, D. (2020). A review of remote sensing applications in tropical forestry with a particular emphasis in the plantation sector. *Geocarto International*, *35*(3), 317–339. https://doi.org/10.1080/10106049.2018.1516245

United Nations. (1973). *Report of the United Nations Conference on the Human Environment: Stockholm, 5-16 June 1972*.

United Nations. (1992). 21: The United Nations Plan of Action from Rio. *United Nations*.

United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development. *New York: United Nations, Department of Economic and Social Affairs*, *1*, 41.

Véga, C., Renaud, J.-P., Durrieu, S., & Bouvier, M. (2016). On the interest of penetration depth, canopy area and volume metrics to improve Lidar-based models of forest parameters. *Remote Sensing of Environment*, *175*, 32–42. https://doi.org/10.1016/j.rse.2015.12.039

Wang, D., Wan, B., Liu, J., Su, Y., Guo, Q., Qiu, P., & Wu, X. (2020). Estimating aboveground biomass of the mangrove forests on northeast Hainan Island in China using an upscaling method from field plots, UAV-LiDAR data and Sentinel-2 imagery. *International Journal of Applied Earth Observation and Geoinformation*, *85*, 101986. https://doi.org/10.1016/j.jag.2019.101986

Watanabe, M., Koyama, C. N., Hayashi, M., Nagatani, I., & Shimada, M. (2018). Early-Stage Deforestation Detection in the Tropics With *L* -band SAR. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(6), 2127–2133. https://doi.org/10.1109/JSTARS.2018.2810857

Watson, J. E. M., Evans, T., Venter, O., Williams, B., Tulloch, A., Stewart, C., Thompson, I., Ray, J. C., Murray, K., Salazar, A., McAlpine, C., Potapov, P., Walston, J., Robinson, J. G., Painter, M., Wilkie, D., Filardi, C., Laurance, W. F., Houghton, R. A., … Lindenmayer, D. (2018). The exceptional value of intact forest ecosystems. *Nature Ecology & Evolution*, *2*(4), 599–610. https://doi.org/10.1038/s41559-018-0490-x

Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., & Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, *39*(12), 4254–4284. https://doi.org/10.1080/01431161.2018.1452075

Yu, T., Wu, W., Gong, C., & Li, X. (2021). Residual Multi-Attention Classification Network for A Forest Dominated Tropical Landscape Using High-Resolution Remote Sensing Imagery. *ISPRS International Journal of Geo-Information*, *10*(1), 22. https://doi.org/10.3390/ijgi10010022

Zhu, X., Wang, C., Nie, S., Pan, F., Xi, X., & Hu, Z. (2020). Mapping forest height using photon-counting LiDAR data and Landsat 8 OLI data: A case study in Virginia and North Carolina, USA. *Ecological Indicators*, *114*, 106287. https://doi.org/10.1016/j.ecolind.2020.106287