# Spécification de paramètres techniques et stratégie d'échantillonnage pour la conception de nouveaux capteurs lidars dédiés à la cartographie de forêts

Tristan Allouis

*École Nationale du Génie Rural des Eaux et des Forêts — AgroParisTech ENGREF*

# Thèse

Pour obtenir le grade de docteur délivré par :

**L'Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech)**

École doctorale SIBAGHE
(Systèmes Intégrés en Biologie, Agronomie, Géosciences, Hydrosciences, Environnement)
Spécialité : Eaux Continentales et Société

# Spécification de paramètres techniques et stratégie d'échantillonnage pour la conception de nouveaux capteurs lidars dédiés à la cartographie de forêts

Présentée et soutenue publiquement par

**Tristan Allouis**

Le 14 décembre 2011

***Jury :***

**Pierre Couteron** Directeur de thèse, *Directeur de recherche*, IRD
**Pierre Flamant**, Co-directeur de thèse (membre invité), *Directeur de recherche*, CNRS
**Sylvie Durrieu**, Encadrante principale, *Chargée de recherche*, Irstea
**Jean-François Dhôte**, Rapporteur, *Directeur de recherche*, INRA
**Laurent Polidori**, Rapporteur, *Professeur*, CNAM
**Jean-Stéphane Bailly**, Membre du jury, *Enseignant-Chercheur*, AgroParisTech
**Xavier Briottet**, Membre du jury, *Directeur de recherche*, ONERA

Thèse préparée à l'unité mixte de recherche « Territoires, environnement, télédétection et information spatiale ».
UMR TETIS, Maison de la télédétection, 500 rue Jean-François Breton, 34093 Montpellier Cedex 5, France

# Résumé

Les forestiers ont besoin d'outils permettant de cartographier les essences, la hauteur, la structure ou la biomasse des peuplements. Même si le lidar aéroporté (light detection and ranging, détection et télémétrie par laser) ne permet pas de mesurer l'ensemble de ces variables, il permet d'accéder aux hauteurs et à la biomasse de manière rapide, précise et surtout spatialisée sur de grandes surfaces. Cependant, les lidars aujourd'hui utilisés en forêts n'ont pas été conçus spécifiquement pour étudier la végétation, et l'ajustement de leurs caractéristiques techniques est supposé permettre d'améliorer la précision des mesures.

L'objectif de cette thèse est de déterminer des configurations de capteurs lidars dédiés à l'étude de la végétation forestière, et de proposer des méthodes d'extraction de paramètres forestiers adaptés aux différentes configurations. La capacité de différentes résolutions (taille d'empreinte et échantillonnage spatial), longueurs d'onde et modes d'enregistrement du signal retour à mesurer des paramètres forestiers (hauteurs et densité d'arbres, taille des couronnes et indirectement volume et biomasse) a été évaluée. Les études ont été menées de l'échelle de l'arbre jusqu'à celle du peuplement, sur des données expérimentales ou simulées.

Dans une première partie, des méthodes de traitement de données lidars aéroportés classiques (scanneur, largeur du faisceau décimétrique, 5 mesures/$m^2$, laser proche infrarouge) ont été développées pour estimation la biomasse d'arbres individuels. Dans cette étude, l'apport de nouvelles données dites *full-waveform* (enregistrement du signal complet) a été démontré par rapport aux traditionnelles données multiéchos (extraction des échos les plus significatifs).

Dans une seconde partie, une expérimentation avec un prototype lidar du Commissariat à l'Énergie Atomique (CEA) embarqué sur un ULM a été réalisée (profileur, largeur de faisceau de 2,4 m, espacement de 2,4 m entre deux mesures successives le long de la ligne de vol, laser ultraviolet). Sans permettre des mesures d'arbres individuels, cette configuration de capteur a permis d'étudier les variations de la structure des arbres à l'intérieur d'une placette forestière (30 m de diamètre). En validant l'utilisation d'un laser ultraviolet pour l'étude de la végétation, les résultats de cette expérience permettent d'envisager le développement de capteurs bifonctions atmosphère/végétation.

Dans une troisième partie, des signaux lidars à larges empreintes (plusieurs dizaines de mètres au sol) ont été simulés par agrégation de signaux lidars aéroportés classiques. Une méthode de modélisation permettant d'étudier la dynamique supposée d'un signal lidar satellitaire dans différents types de forêts a été proposée. Elle apporte des informations utiles à la calibration de l'énergie à émettre pour une future mission satellitaire dédiée à la cartographie de forêts. Le principal problème avec les données à larges d'empreintes est la forte influence de la topographie sur la précision des mesures de hauteurs d'arbres en zones pentues. En conséquence, une méthode de correction de cet effet a été élaborée, permettant ainsi de mesurer les hauteurs d'arbres avec une précision jusqu'alors inenvisageable. Cette méthode a de plus ouvert de nouvelles perspectives dans l'estimation de la topographie sous la forêt, à partir de données lidars à larges empreintes.

**Mots clés :** Télédétection, Laser, Protocole de Kyoto, Biomasse, Végétation, Foresterie.

# Abstract

Foresters need tools to map the tree species, tree heights, stand structure and biomass. Although the airborne lidar (Light detection and ranging) technology does not give access to all these variables, it can provide quick, accurate and spatially explicit measurements of tree heights and biomass over large surfaces. However, lidar systems currently used have not been specially designed to performed vegetation studies. The adjustment of the technical characteristics of such systems is expected to improve the accuracy of retrieved forest parameters.

Consequently, the objective of this thesis is to determine configurations of lidar sensors dedicated to the study of forest vegetation, and to propose methods designed to extract forest parameters depending on the different configurations. The ability of different resolutions (footprint size and spatial sampling), wavelengths and sampling modes of the backscattered signal to measure forest parameters (canopy height and density, crown size and also volume and biomass) was evaluated. The studies were conducted from tree to stand level, on experimental or simulated data.

In a first part, we developed methods to process classic airborne lidar data (scanner system, tens of centimeter footprint, 5 measurements/$m^2$, near-infrared laser) for the estimation of the biomass of individual trees. In this study, we demonstrated the contribution of new data called "Full-waveform" (recording the entire signal) compared to traditional multi-echoes data (extraction of the most significant echoes).

In a second part, we performed an experiment using a Comissariat of Energy Atomique (CEA)'s lidar prototype onboard an ultra-light aircraft (profiler system, 2.4 m footprint, 2.4 m spacing between two measurements along the flight line, ultraviolet laser). Such a configuration did not allow to measure individual trees, but we were able to study variations in forest structure at the plot level (30 m diameter). Having demonstrated the ability of an ultraviolet lidar to perform vegetation studies, this sensor opens the way to the development of bi-functional lidar for both atmosphere and vegetation remote sensing.

In a third part, large footprint lidar signals (tens of meters on the ground) were simulated from the aggregation of classical airborne lidar signals. We proposed a method for modeling the signal dynamics of satellite lidars in different forest types, in order to calibrate the energy to emit for a future space-borne mission. The main problem with large-footprint size is the strong influence of topography on accurate measurements of tree heights in steep areas. We consequently developed a method to correct this effect, thus increasing the accuracy of tree height retrieval. This approach also opened new perspectives in topography assessment from large-footprint data in forest environments.

**Key words:** Remote Sensing, Laser, Kyoto Protocol, Biomass, Vegetation, Forestry.

Je dédie ce mémoire au docteur William Horatio Bates (1860 – 1931), éminent ophtalmologiste new-yorkais, professeur et chirurgien. Ses travaux de recherche sur l'amélioration naturelle de la vue m'ont accompagné et inspiré durant l'élaboration de cette thèse.

# Remerciements

Je tiens tout d'abord à remercier mon encadrante, Sylvie Durrieu, avec qui j'ai eu plaisir à travailler durant ces trois années de thèse. J'ai appris à rester précis dans mon raisonnement et ma rédaction (un peu trop parfois !). Merci d'avoir accepté de m'envoyer à l'autre bout du monde pour présenter mes travaux.

Les premiers remerciements vont aussi à l'Irstea et de la région Languedoc-Roussillon, qui m'ont permis de réaliser ce doctorat grâce à leur financement. Je remercie Pierre Couteron d'avoir accepté de diriger la thèse, ainsi que pour son aide précieuse au cours de mes travaux de recherche et lors de la rédaction du présent mémoire. Un grand merci à Jean-Stéphane Bailly, celui sans qui je ne serais probablement jamais arrivé à la Maison de la télédétection, pour son soutien et ses tuyaux sur R.

Une des parties les plus intéressantes de la thèse s'est réalisée hors du bureau, sur le terrain. J'ai pris un grand plaisir à expérimenter des prototypes lidars avec Patrick Chazette, entre les aérodromes, la forêt et les discussions au restaurant. La collecte des mesures de référence utilisées dans cette thèse n'aurait pas été possible sans l'aide précieuse de Cédric Véga, Laurent Albrech, Jonathan Baron, Marc Bouvier et Sylvie.

Merci à mes coéquipiers et colocataires de bureaux successifs, Cédric, Hani et Marc. Je remercie de manière générale tous les collègues de la MTD, pour l'ambiance qui y règne. Elle a été décisive dans mon choix de prolonger mon séjour pendant trois ans. Une pensée particulière va à ceux avec qui j'ai partagé le quotidien, notamment lors des déjeuners à la cantine : Isa, Gab, Nat, Roberto, Clotilde, Silvan, Jenny et tant d'autres...

Un clin d'œil pour ceux, issus de la MTD, avec qui j'ai profité d'une partie de mes RTT et de mon temps libre : Bruno Roux et Denis Feurer pour les bonnes sessions de planche à voile. Ma dernière pensée va à Clémence, ma compagne, pour sa présence à mes côtés.

# Table des matières

# Introduction

La forêt occupe 30% de la superficie totale des terres. Écosystème riche et complexe, elle abrite une grande diversité d'espèces végétales, animales, fongiques et microbiennes qui, en interagissant, maintiennent le cycle de la vie. En particulier, la structure de la forêt – agencement des arbres dans les trois dimensions – a un impact sur le fonctionnement de l'écosystème, le cycle du carbone, de l'eau et des nutriments (*Shugart et al., 2010*). Connaître la structure forestière est donc nécessaire pour comprendre la dynamique des écosystèmes ainsi que les relations complexes liant la biosphère et l'atmosphère.

## Contexte sociétal

La forêt est depuis toujours considérée comme une source de revenus pour l'homme. Jusqu'à la fin du XX$^{e}$ siècle, la régulation de son exploitation avait pour but d'éviter l'épuisement des ressources pour des raisons économiques (fourniture de bois d'œuvre) et vivrières (bois de chauffe, produits forestiers non ligneux). Ce n'est que depuis la conférence des Nations unies sur l'environnement et le développement (ou Sommet de Rio) en 1992 que de nouvelles préoccupations sont reconnues et formalisées au travers d'une convention internationale. Ainsi, le terme « gestion durable des forêts » apparaît pour que des indicateurs sociaux et environnementaux soient pris en compte au côté d'objectifs économiques. Des éléments clés de cette gestion durable sont la conservation de la diversité biologique, le maintien de la santé et de la vitalité des forêts, la protection des ressources en sol et en eau ainsi que la conservation du patrimoine culturel (*Prabhu et al., 1999*).

La Convention-cadre des Nations unies sur les changements climatiques est aussi créée lors du sommet de Rio. L'objectif de ce traité est d'étudier les mécanismes des changements climatiques, au travers de leurs causes et de leurs conséquences, et de mettre en œuvre des solutions pour y remédier (*Bruce et al., 1996*). En 1997, le protocole de Kyoto est signé. Il vise à réduire les émissions de gaz à effet de serre. En 2001, les accords de Bonn instaurent des avantages pour les pays dont certaines activités absorbent et stockent le carbone présent dans l'atmosphère, tels les reboisements (*Gouvernement Français, 2005*).

À l'opposé, la déforestation est la seconde source anthropique d'émission de carbone dans l'atmosphère, après la combustion du pétrole (*van der Werf et al., 2009*). Malgré les avantages accordés depuis les accords de Bonn, elle reste une activité très lucrative pour certaines entreprises ou responsables politiques locaux, en particulier dans la zone équatoriale aussi appelée tropicale humide (Figures 1 et 2). La forêt y est souvent pillée avant d'être convertie en terre agricole, généralement de manière illégale. La destruction de l'écosystème mène en quelques années à l'épuisement des terres, à l'érosion de la biodiversité et à la privation des populations des ressources de la forêt et parfois de leurs terres (*Le Guen, 2010*). Pour lutter contre la déforestation et favoriser les zones à forte biodiversité, l'initiative REDD (Reducing Emissions from Deforestation and Forest Degradation) et ses élargissements REDD+ et REDD++ s'appuient sur des incitations financières à destination des états (*Minang et al., 2009*).

De forts enjeux géopolitiques ont donc émergé, notamment en 2005 au travers de la création du Marché européen d'échanges des quotas d'émissions de carbone. Dans ce marché, les pays et les entreprises peuvent s'échanger leurs droits d'émissions de gaz à effet de serre (ou crédits-carbone) afin d'atteindre les objectifs fixés par le protocole de Kyoto (*Ellerman and Buchner, 2007*) ou par l'accord qui devra le prolonger. Une entreprise ou un pays peut alors monnayer ses droits d'émissions non consommés, de la même façon qu'un autre pays peut les acheter pour augmenter sa capacité d'émission. L'inventaire et le suivi de la biomasse forestière, vue comme une ressource en tant que telle, sont donc des enjeux mondiaux dans les domaines économique, environnemental et social.

## Inventaire de la ressource forestière

Un inventaire et un suivi de la dynamique des écosystèmes sont requis par les institutions internationales pour tenter d'enrayer la destruction de la forêt. Un inventaire mondial des ressources forestières est régulièrement effectué par la FAO (Food and Agricultural Organisation, United Nations), sur la base d'estimations nationales effectuées par les pays membres (*FAO, 2006*). Cet inventaire fait principalement le bilan de l'étendue des forêts, de leur gestion et de leur utilisation.

L'inventaire forestier est traditionnellement effectué depuis le sol en mesurant les arbres sur de petites surfaces carrées, rectangulaires ou circulaires de quelques centaines de mètres carrés appelées placettes. Sur ces placettes, les diamètres des troncs à 1,3 m (diameter at breast height ou DBH, diamètre à hauteur de poitrine) sont mesurés. Les hauteurs d'une partie ou de tous les arbres peuvent aussi l'être, en fonction du protocole utilisé. Certains protocoles particuliers, liés à des fins de recherche, envisagent la mesure de la dimension des houppiers et de la position des arbres dans la placette. Les placettes sont distribuées dans l'espace selon un plan d'échantillonnage puis la ressource forestière est extrapolée à l'ensemble du territoire, en utilisant différentes techniques statistiques (*IFN, 2010*).

La mesure de la biomasse, qui peut être vue comme la masse de carbone emmagasinée par les forêts, nécessite de peser la quantité de matériel végétal déshydraté. Des techniques destructives appliquées à un petit nombre d'arbres (coupe, déshydratation, pesée) sont donc utilisées pour construire des relations allométriques à partir des variables mesurées lors des inventaires forestiers (DBH et éventuellement hauteurs). Ces relations sont ensuite utilisées pour déduire la biomasse à partir de mesures effectuées sur d'autres arbres. Ces relations sont susceptibles d'être différentes pour chaque essence dans chaque condition de sol ou de climat, mais elles sont souvent construites et approximées à des échelles régionale ou nationale.

L'inventaire forestier mondial (*FAO, 2006*), établi à partir de mesures de terrain, rencontre certaines difficultés. De nombreux pays – et pas seulement les moins développés – ne possèdent pas d'organisme chargé des inventaires forestiers à l'échelle nationale. Aussi, l'hétérogénéité des méthodes d'évaluation et des intervalles de temps de surveillance peut mener à des évaluations imprécises, en particulier pour les forêts transcendant les frontières nationales. De plus, la mesure de certains paramètres telle que la hauteur des arbres est difficilement réalisable depuis le sol en forêt tropicale. Même dans le cas de forêts tempérées, une étude a montré une erreur aléatoire de 1,6 m sur les hauteurs d'arbres comprises entre 10 et 45 m et mesurées par deux équipes différentes (*Brassel, 2001*).

Faire appel à la télédétection prend ici tout son sens, puisqu'elle permet d'acquérir des informations par l'intermédiaire de capteurs aéroportés ou satellitaires, à différentes échelles de temps et d'espace, et selon des procédures qui peuvent rester homogènes au-delà des frontières nationales. Les méthodes d'inventaire développées à partir de données de télédétection sont vérifiables et régulièrement améliorables. Elles sont moins coûteuses que des opérations de terrain, surtout dans le cas d'une cartographie exhaustive à l'échelle d'un territoire.

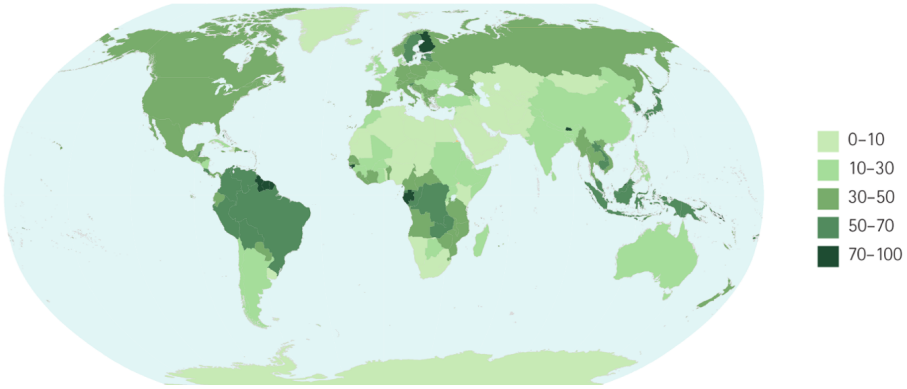



**Figure 1** – Superficie forestière par pays en pourcentage de la superficie totale des terres, 2010.
*Source : FAO, Évaluation des ressources forestières mondiales 2010*

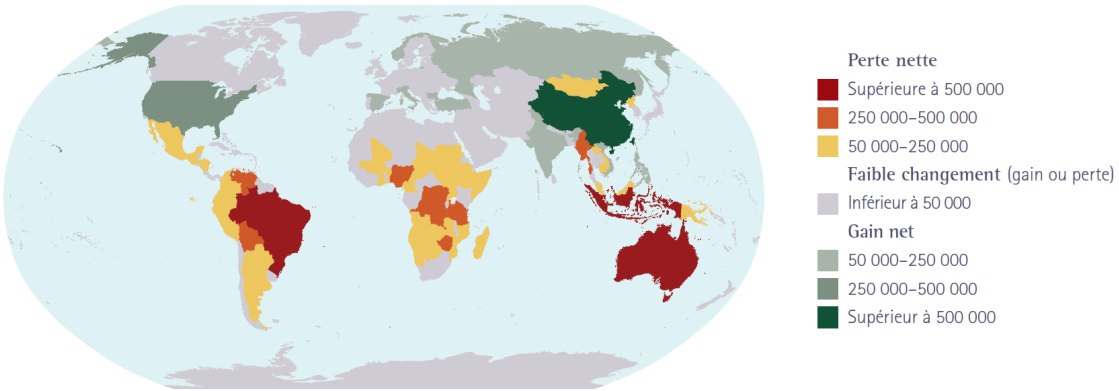



**Figure 2** – Changement net de superficie forestière par pays, 2005–2010 *(ha/an)*.
*Source : FAO, Évaluation des ressources forestières mondiales 2010*

## Télédétection et forêt

Les capteurs de télédétection se répartissent selon deux grandes familles : les capteurs passifs qui reçoivent l'énergie sans en émettre eux-mêmes, et les capteurs actifs qui enregistrent la quantité d'énergie qu'ils ont eux-mêmes émise et qui a été réfléchie par la cible (Figure 3).

Les capteurs passifs les plus utilisés sont des capteurs optiques, qui prennent des images dans le spectre électromagnétique ultraviolet, visible, proche infrarouge ou infrarouge. Le canal proche infrarouge est le plus couramment utilisé pour étudier la végétation, puisqu'il permet de suivre l'activité végétale et sa santé grâce à la forte réflectance de la chlorophylle dans cette gamme spectrale (*Asner, 1998*). Certaines informations de structure de la forêt telles la taille de la couronne, la hauteur d'arbre et la biomasse peuvent parfois être indirectement mesurées à partir de la réflectance et de ses variations dans l'image (texture), grâce à des modèles statistiques (*e.g. Proisy et al. (2007)*). Ces modèles demandent cependant d'être étalonnés dans chaque contexte local sur la base de données de terrain. De plus, les capteurs passifs n'apportent pas de mesure directe de la structure en trois dimensions de la végétation. Encore, l'utilisation de l'imagerie optique est perturbée par la présence de nuages qui limitent considérablement son utilisation en zones tropicales humides, où la couverture nuageuse est quasi permanente. Enfin, la réflectance mesurée est soumise à une grande variabilité en fonction des conditions atmosphériques ou de l'éclairement de la scène. Elle nécessite donc d'être corrigée de ces effets (*Nadeau, 2000*).

Parmi les capteurs actifs, les radars à synthèse d'ouverture (aussi appelés RSO, ou SAR en anglais) émettent et reçoivent des micro-ondes radio dont la longueur d'onde conditionne leur pouvoir de pénétration à l'intérieur du couvert forestier. Plus la longueur d'onde est élevée (jusqu'à plus d'1 m pour les radars en bande P), plus l'onde va pénétrer profondément dans le couvert forestier (*Sexton et al., 2009*). Ces capteurs ouvrent ainsi davantage de perspectives pour la caractérisation de la structure verticale de la végétation et l'estimation de la biomasse. En effet, la mesure de la quantité d'énergie réfléchie par la végétation est corrélée avec la géométrie, l'humidité ainsi que l'orientation de la cible (*Shugart et al., 2010*). La densité des composantes ligneuses aériennes de la végétation peut aussi être déduite, permettant ainsi d'accéder à des mesures de la biomasse et du volume de bois (*Le Toan et al., 2011*). Le radar est peu perturbé par les conditions atmosphériques et fonctionne de jour comme de nuit. Cependant, lorsque la biomasse augmente, la mesure y est de moins en moins sensible jusqu'à atteindre un seuil de saturation. Ce seuil est atteint d'autant plus rapidement que la longueur d'onde est courte. Dans le meilleur des cas (bande P), le SAR ne permet de mesurer des biomasses aériennes que jusqu'à 300 tonnes par hectare (*Shugart et al., 2010*).

## Caractéristiques des capteurs lidar

Le lidar (« light detection and ranging ») est un capteur actif particulièrement adapté à la mesure de la structure en trois dimensions de la forêt. Il fonctionne sur le principe d'émission/réception d'impulsions laser (Figure 4). Le temps de parcours de l'onde, entre son émission et sa réception après réflexion sur la scène étudiée, est mesuré par le capteur et converti en distance en multipliant la durée par la vitesse de propagation de la lumière. En couplant cette mesure de distance avec les données de navigation issues d'un GPS et d'une centrale inertielle, les différentes cibles ayant réfléchi le laser sont positionnées dans un référentiel global.

La plupart des lidars fonctionnent dans le proche infrarouge, à la longueur d'onde d'émission des laser Nd:YAG (grenat d'yttrium aluminium dopé au néodyme) les plus couramment utilisées (1064 nm). Cette longueur d'onde est particulièrement adaptée aux applications forestières, car la végétation possède dans ce domaine une transmittance et une réflectance importantes (*Asner, 1998*). Il est cependant possible pour le lidar d'émettre des photons ayant la moitié de la longueur d'onde des photons initiaux (seconde

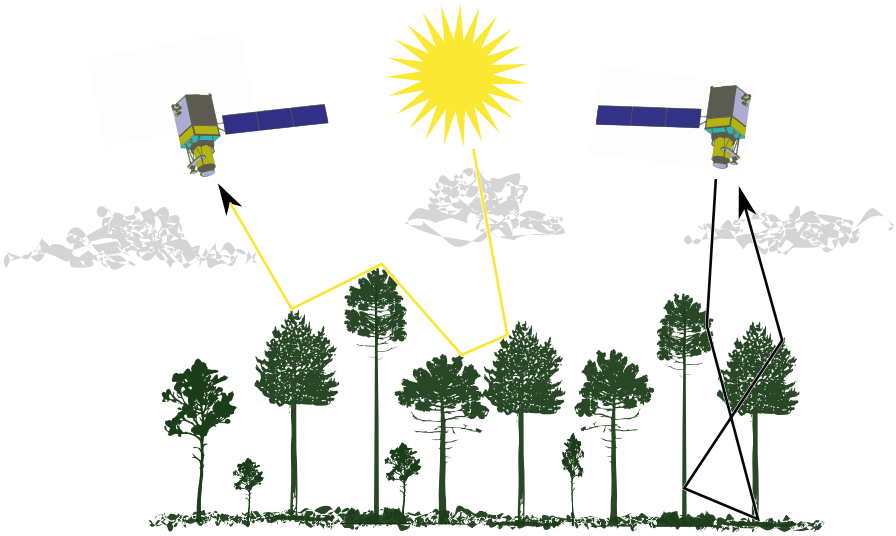


**Figure 3** – Interactions entre les capteurs passifs (gauche), actifs (droite) et la forêt.

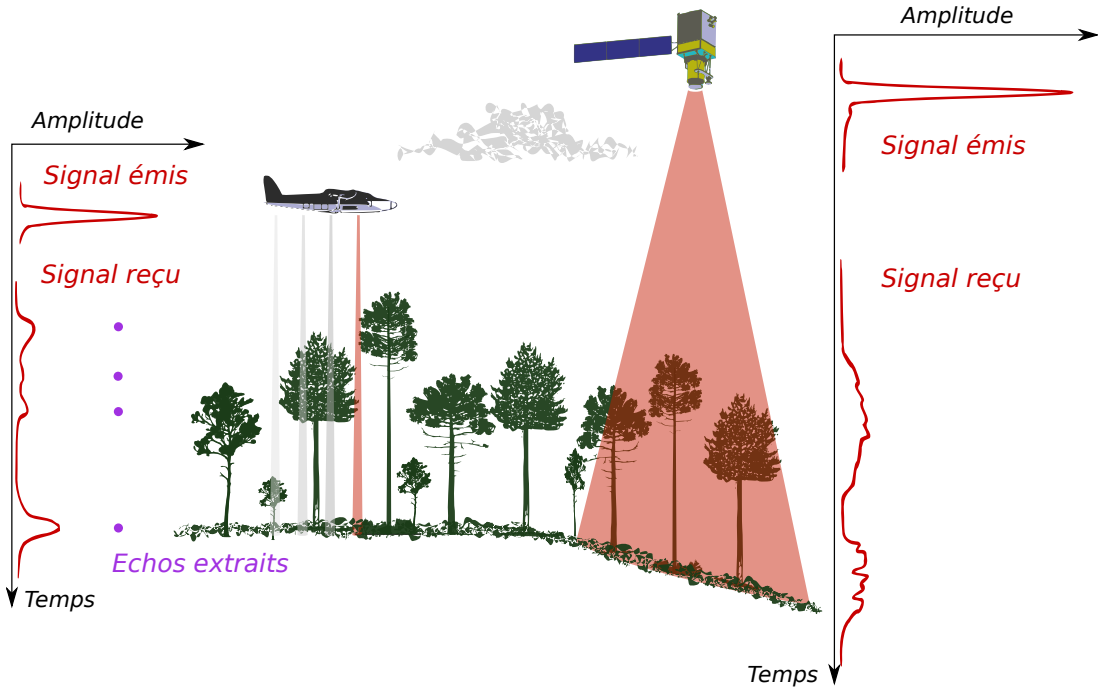



**Figure 4** – Principe du lidar et effet de la taille du faisceau laser sur la forme du signal retour.

harmonique) grâce à un cristal non linéaire. Ce processus permet d'obtenir des lidars verts (532 nm) utilisés pour la bathymétrie (*Allouis et al., 2010*), ou ultraviolets (355 nm) utilisés pour la mesure des concentrations d'aérosols dans l'atmosphère (*Chazette et al., 2007*).

Un système lidar est aussi caractérisé par le mode d'enregistrement de son signal retour. Alors que le lidar satellitaire GLAS (Geoscience Laser Altimeter System) embarqué sur ICESat (Ice, Cloud,and land Elevation Satellite) enregistre la totalité du signal à un pas vertical de 15 cm (1 GHz), ce mode n'est que récemment apparu sur les systèmes aéroportés. Jusqu'à peu, seuls les échos les plus significatifs étaient enregistrés (systèmes multiéchos). La disponibilité de nouveaux types de données a permis d'extraire un nombre plus important d'échos qu'avec les méthodes classiques. De nouvelles informations sur la nature de la cible sont aussi extraites, telles que l'intensité de l'écho ou sa forme. Elles se sont montrées complémentaires aux données classiques dans des études récentes pour la discrimination des échos sol et végétation ainsi que pour la reconnaissance des essences d'arbres (*Heinzel and Koch, 2011*).

Une autre caractéristique importante des lidars est la dimension du faisceau laser au sol (taille d'empreinte), qui dépend de la divergence du faisceau et de l'altitude de vol. Les systèmes lidars aéroportés fonctionnent généralement avec des tailles d'empreintes de quelques dizaines de centimètres, alors que les lidars satellitaires GLAS et CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) possèdent des empreintes respectives de 70 m et 90 m (Figure 4). La taille d'empreinte est l'un des deux paramètres dimensionnant la résolution de l'appareil et donc la finesse de mesure. Avec de larges empreintes, il n'est pas possible de détecter des arbres individuels, ni d'étudier des variations de structure à l'intérieur d'une placette forestière. Bien que la taille d'empreinte soit en théorie ajustable, par modification des optiques d'émission, elle doit être choisie afin que le lidar respecte les normes de sécurité oculaire. Une empreinte de quelques dizaines de centimètres émise depuis un satellite concentrerait l'énergie à un taux dangereux pour l'oeil humain.

La résolution est aussi déterminée par l'espacement entre deux mesures successives. Les capteurs à petites empreintes effectuent généralement un balayage de la scène grâce à un miroir oscillant et à une fréquence de répétition de plusieurs centaines de kilohertz. En fonction de la vitesse de l'avion et de l'altitude de vol, ces scanneurs peuvent mener à plusieurs dizaines de mesures par mètre carré. Un balayage est difficilement envisageable pour les systèmes satellitaires pour des raisons techniques, puisque le déplacement du miroir agirait sur la stabilité du satellite et serait un élément supplémentaire pouvant causer des problèmes de fiabilité. Les lidars satellitaires sont donc des systèmes profileurs dont la fréquence de répétition et la vitesse de déplacement conduisent à un espacement entre deux mesures successives le long de la trajectoire de l'ordre de 170 m pour GLAS et 330 m pour CALIOP. Il est techniquement possible d'augmenter la cadence d'émission, mais il est nécessaire de maximiser la durée de vie du laser (9 millions de tirs) pour que le satellite puisse couvrir plusieurs fois l'ensemble du globe.

Le choix de la taille d'empreinte et de l'espacement entre les mesures doivent être étudiés en détail avant le lancement d'un lidar satellitaire. En effet, contrairement aux prototypes lidar aéroportés pour lesquels il existe une marge de maneuvre dans la modifications des paramètres, un lidar satellitaire une fois lancé ne peut plus subir de modifications. Afin de calibrer correctement la puissance d'émission du laser, il est aussi nécessaire d'étudier la dynamique du signal à larges empreintes dans différents contextes forestiers (méditerranéen, tropical, etc.). Différentes stratégies de simulation de signaux peuvent être mises en œuvre. Elles peuvent être réalisées grâce à la modélisation des interactions entre les photons et la végétation, au travers de programmes et de maquettes informatiques de plantes (*Disney et al., 2000*). Cette solution requiert cependant de lourds investissements pour pouvoir simuler différents contextes forestiers, des maquettes réalistes devant être mises au point pour chaque essence et pour différents stades de développement des arbres. Une autre solution est de modéliser les interactions entre le signal et la végétation, au travers d'approches optico géométriques et de transfert radiatif (*Ni-Meister et al., 2001*). Enfin, des données à larges empreintes peuvent être simulées par agrégation de données à petites ou moyennes empreintes (*Sun et al., 2008; Muss et al., 2011*). Cette dernière approche ne prend pas

en compte l'influence de l'atmosphère, mais elle permet d'appréhender la forme que pourrait avoir un signal à larges empreintes.

Depuis vingt-cinq ans, le lidar aéroporté a été utilisé avec succès pour étudier le volume et la biomasse des forêts, avec des empreintes de 20 m pour les premiers systèmes (*Nelson et al., 1988*). L'évolution des capteurs, avec les tailles d'empreintes de plus en plus petites et des densités de mesures de plus en plus fortes, a permis des études de plus en plus détaillées sur la structure en trois dimensions des forêts, telles les hauteurs d'arbres et la distribution verticale du matériel végétal (*Lefsky et al., 1999; Means et al., 2000*). Les données lidars peuvent être couplées à des données apportant des informations complémentaires, telles que des images multispectrales (*Hyde et al., 2006*) ou radar (*Hall et al., 2011*). Ces dernières, acquises en vue oblique, fournissent des informations de structure de la forêt complémentaires à celles acquises verticalement par les lidars. Le lancement en 2003 du satellite ICESat a permis quant à lui d'obtenir des signaux lidars sur l'ensemble du globe.

Certaines solutions doivent cependant être trouvées afin de concevoir un lidar satellitaire réellement dédié à la cartographie de forêt. Bien que les signaux lidars de GLAS aient montré leur potentiel pour mesurer la hauteur de la canopée à l'échelle globale (*Lefsky et al., 2005*), ils se sont montrés imprécis en zone de pente à cause de problèmes de confusion entre le sol et la végétation provoqués par la taille de l'empreinte (70 m) (*Chen, 2010*). Les signaux de CALIOP sont quant à eux numérisés à une fréquence incompatible à l'étude de la forêt (résolution verticale de 30 m). La mesure précise des hauteurs d'arbres, en particulier pour les forêts denses, nécessite qu'une quantité suffisante d'énergie ait pénétré jusqu'au sol et y ait été réfléchie. Enfin, à défaut de pouvoir utiliser la technologie scanner sur satellite, il convient de réfléchir à une stratégie d'échantillonnage permettant d'obtenir des mesures à une résolution adaptée à un inventaire forestier global.

## Objectifs et structure de la thèse

Les nombreuses études menées par le passé ont cherché à estimer des paramètres forestiers à partir de capteurs lidars conçus initialement pour mesurer la topographie, à l'image des capteurs aéroportés ou satellitaires présentés précédemment. L'objectif de cette thèse est, à l'inverse, de déterminer des configurations de capteurs lidars dédiés à l'étude de la végétation forestière, et de proposer des méthodes d'extraction de paramètres forestiers adaptés aux différentes configurations, pour dépasser les limites identifiées ci-dessus.

Les caractéristiques étudiées dans cette thèse, sur la base desquelles ce mémoire a été structuré, sont la taille d'empreinte, l'espacement entre mesures successives, la longueur d'onde, ainsi que le mode d'enregistrement du signal retour. La performance des configurations testées a été déterminée en quantifiant la précision des mesures de la structure forestière extraites, de l'échelle de l'arbre jusqu'à l'échelle du peuplement, sur des données expérimentales ou simulées.

Dans la première partie de ce mémoire, nous avons étudié le potentiel des signaux complets par rapport aux données multiéchos traditionnelles pour l'estimation de la biomasse aérienne. Le capteur utilisé fournissant des données à très haute résolution (5 mesures/m$^2$ avec une taille d'empreinte décimétrique), nous avons travaillé à l'échelle de l'arbre individuel.

Cette partie consacrée à l'étude de la performance des capteurs à petites empreintes a été élaborée en deux articles. Le premier est un article de conférence, présenté à la conférence *IEEE International Geoscience and Remote Sensing Symposium* (IGARSS) en 2011. Il pose les bases de la méthode qui ont été ensuite reprises, précisées et améliorées dans un article plus conséquent soumis à la revue internationale *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* pour un numéro spécial IGARSS 2011.

Jusqu'à présent, aucune étude ne s'était intéressée à une taille d'empreinte intermédiaire à celles des données classiques issues de capteurs à haute résolution et des données satellitaires aujourd'hui utilisées (empreintes décamétriques). Des empreintes de 2,4 m ont donc été expérimentées sur un prototype lidar du Commissariat à l'Énergie Atomique (CEA) embarqué sur un ULM et effectuant des mesures jointives le long de sa ligne de vol (système profileur). Cette résolution était destinée à étudier des variations de la structure des arbres sur une parcelle sans aller jusqu'à la mesure d'arbres individuels. Elle semblait compatible avec les contraintes d'utilisation de systèmes satellitaires.

Le second objectif de cette étude était de valider l'utilisation d'un laser ultraviolet pour la mesure de la structure de la végétation. Une réponse différente de la végétation au laser ultraviolet était attendue par rapport au proche infrarouge couramment utilisé. Cependant, si l'ultraviolet permettait de mesurer correctement la structure des arbres il serait alors possible d'envisager le développement de capteurs bifonctions atmosphère/végétation. Cette expérimentation visait aussi à étudier l'impact d'un échantillonnage moins précis du signal retour (100 MHz, *i.e.* 1,5 m) pouvant être considéré comme fournissant des données multiéchos.

Cette étude s'articule en deux articles. Le premier, finaliste du concours de l'article étudiant, a été présenté à la conférence IGARSS 2010. Le second, qui se veut une étude plus poussée, a été publié dans un numéro spécial de la revue *ISPRS Journal of Photogrammetry and Remote Sensing* intitulé « Advancements in LiDAR data processing and applications ».

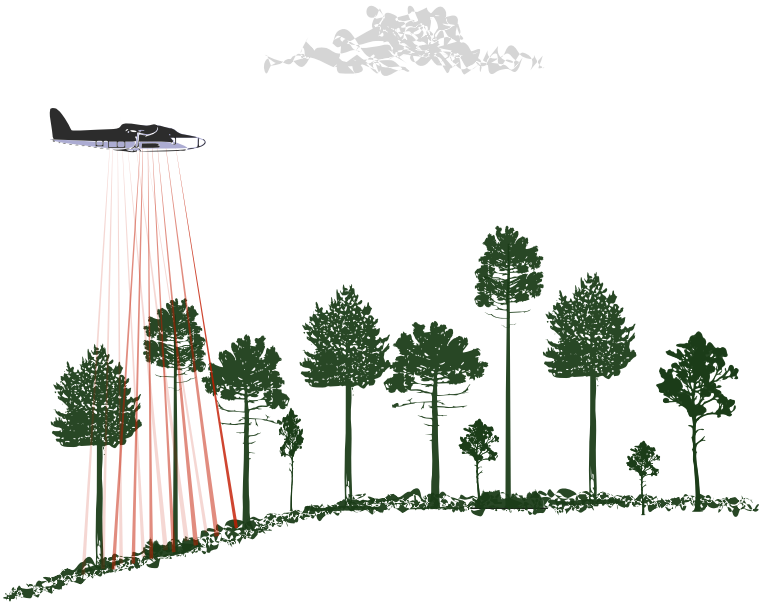
Pour la troisième partie consacrée aux données à larges empreinte, plusieurs modèles de simulation de signaux ont été développés. Ces modèles ont été élaborés à partir de données multiéchos possédant différents niveaux d'information. L'objectif était ici de proposer un outil permettant d'étudier la dynamique du signal dans différents contextes forestiers. La finalité de ces travaux était d'apporter des informations nécessaires à la calibration de l'énergie à émettre depuis un satellite. Cette étude fait suite à la proposition LEAF (Lidar for Earth And Forest) soumise à l'Agence Saptiale Européenne dans le cadre de l'appel à proposition de missions satellitaires pour Earth Explorer 8 (*Durrieu, 2010*).

Le second objectif de cette partie était d'améliorer les mesures de hauteur d'arbres dans des zones pentues à partir de signaux à larges empreintes. Pour cela une démarche inverse de modélisation a été mise en œuvre, permettant d'estimer et de séparer les contributions du sol et de la végétation dans les signaux.

Cette partie a donc été structurée en deux documents. Le premier est un rapport d'étude concernant des modèles de simulation. Le second article porte sur la correction de l'influence de la pente du terrain. Il a été publié dans la revue *IEEE Geoscience and Remote Sensing Letters*.

# Chapitre 1

# Traitement de données lidars à petites empreintes



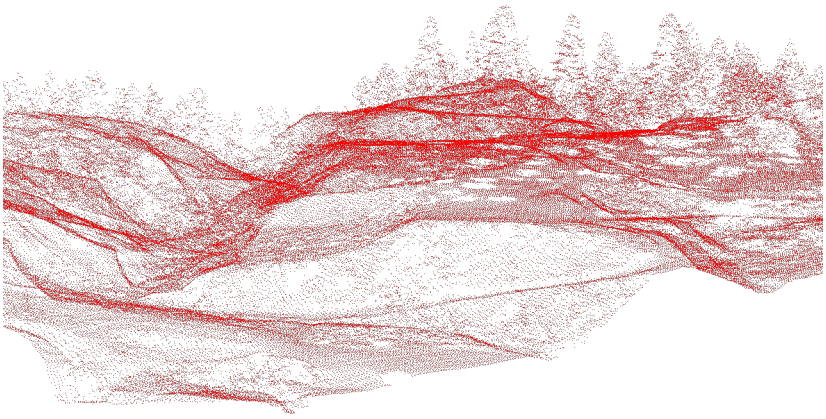


**Figure 1.1** – Exemple d'un nuage de points 3D issus de l'extraction des échos les plus significatifs

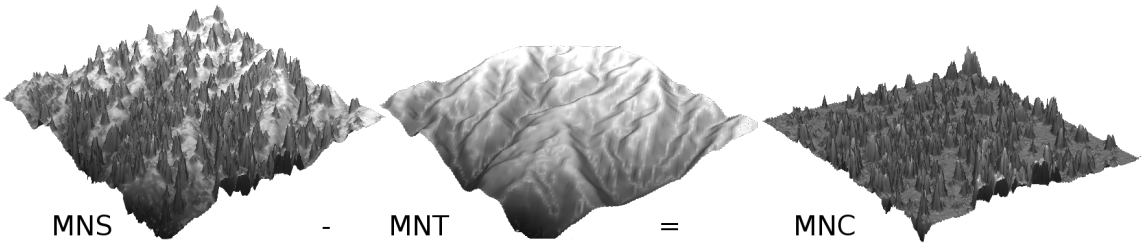



**Figure 1.2** – Modèle numérique de surface (MNS), modèle numérique de terrain (MNT) et modèle numérique de canopée (MNC)

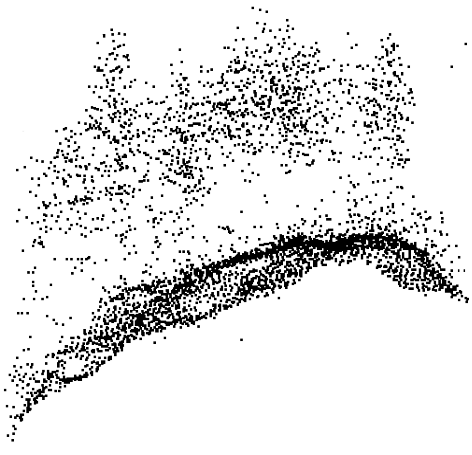


**Figure 1.3** – Données multi-échos

**Figure 1.4** – Données full-waveform

Les systèmes lidar à petites empreintes sont les plus utilisés depuis une quizaine d'années dans le domaine forestier (*Næsset, 1997; Dubayah and Drake, 2000*). Ces systèmes à balayage (scanners) utilisent un faisceau laser dont la tâche au sol (appelée empreinte) mesure quelques dizaines de centimètres de diamètre. Le paysage scanné est généralement restitué sous la forme d'un nuage de points 3D issu de l'extraction des échos les plus significatifs (Figure 1.1).

Le nuage de points peut être traité directement pour extraire des caractéristiques d'arbres (*Kato et al., 2009*), mais le processus suivant est généralement utilisé (Figure 1.2) : Un modèle numérique de surface (MNS) est calculé en interpolant les altitudes des échos de surface (premiers échos). Un modèle numérique de terrain (MNT) est calculé à partir des altitudes des échos classifiés sol. Le MNT est enfin soustrait au MNS pour produire un modèle numérique de canopée (MNC) qui fourni des informations sur la hauteur de la végétation. Ce MNC est ensuite traité pour extraire la densité d'arbres ainsi que leurs hauteurs (*Lim et al., 2003*), ces paramètres menant à l'estimation de la biomasse aérienne.
Pour obtenir des informations sur la structure verticale de la végétation, la distribution verticale de l'ensemble des échos (Figure 1.3) peut être étudiée (*Tonolli et al., 2010*). Cette technique permet notamment de travailler à l'échelle de la placette forestière, en se passant de la détection des arbres individuels.
Enfin, la récente apparition des capteurs full-waveform à petites empreintes a mis à disposition un nouveau type de données (enregistrement du signal complet et non plus seulement des échos significatifs, voir figure 1.4) dont il est nécessaire l'étudier le potentiel pour caractériser la structure verticale de la végétation.

Ainsi, l'objectif de cette partie est de déterminer le potentiel des données full-waveform pour l'amélioration des estimations de volume et biomasse d'arbres individuels. De nouvelles méthodes de traitement ont été développées pour travailler à l'échelle de l'arbre individuel, en s'insprirant des méthodes utilisées pour le traitement des signaux à plus larges empreintes (voir chapitres 2 et 3).

# EXPLOITING FULLWAVEFORM LIDAR SIGNALS TO ESTIMATE TIMBER VOLUME AND ABOVE-GROUND BIOMASS OF INDIVIDUAL TREES

*Tristan Allouis*[a,*], Student Member IEEE, *Sylvie Durrieu*[a], *Cédric Véga*[a,b], *Pierre Couteron*[c]

[a] Cemagref / AgroParisTech, UMR TETIS, Montpellier, France
[b] French Institute of Pondicherry, Pondicherry, India
[c] Institut de recherche pour le développement, UMR AMAP, Montpellier, France
[*] Corresponding author: tristan.allouis@teledetection.fr

## ABSTRACT

Volume and biomass are estimated mostly using tree diameter at breast height (BDH), which is not reachable using light detection and ranging (lidar) data. In this paper, we compare volume and biomass estimations performed using lidar metrics derived from a traditional canopy height model (CHM), such as total height and crown area, with estimations using additional metrics derived from full-waveform (FW) lidar signals related to the vertical distribution of vegetation. The study is performed on a selection of black pine trees (*Pinus nigra ssp. nigra* [Arn.]) in the French Alps. Predictive models are built on CHM-only and CHM+FW metrics, automatically selected using a stepwise algorithm, and the accuracy of the different models is compared. We find that FW metrics improve volume estimation of 3% and biomass estimations from 5% to 7%, with a slightly less spread distribution error.

***Index Terms—*** lidar, biomass, forest, signal processing, regression

## 1. INTRODUCTION

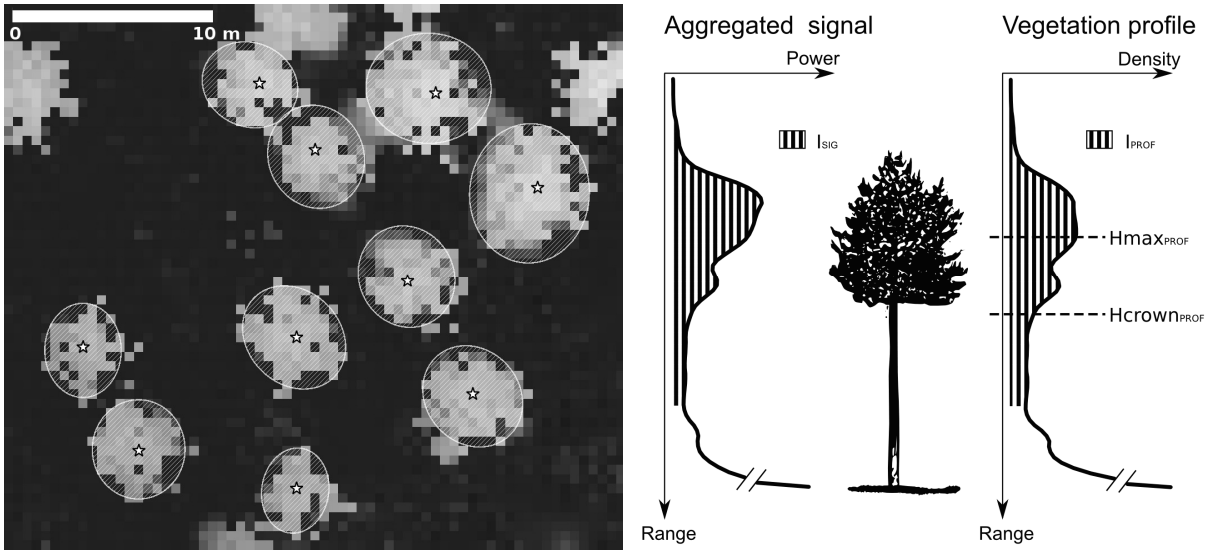
Timber volume is needed by forest managers to assess the productivity of their forest, and to plan silvicultural activities. As for biomass, it is a primary indication of carbon sequestration rate in forest ecosystems and has been identified as an Essential Climate Variable (ECV) needed to improve our knowledge on carbon cycle and define efficient strategies for climate mitigation. These two forest parameters are usually estimated through allometric relationships based on field-measured parameters such as tree diameter at breast height (DBH) and total height. However, field measurements are not spatially explicit. Even if lidar remote sensing data cannot be analyzed without field control measurements, they provide an effective solution for assessing forest parameters at various scales, from individual tree to global level. The recent development of small-footprint full-waveform airborne lidar sensors has provided new opportunities for vegetation studies [1] [2]. Unlike traditional small-footprint discrete-return lidars, where only the most significant echoes are recorded, this new generation of sensors gives access to the entire backscatter information through the vegetation column. Additional features related to echo intensity and shape are then available for refining vegetation structure measurements at individual tree level. In order to evaluate the potential of these new data to improve volume and biomass estimations at individual tree level, we propose in this paper a novel method based on the aggregation and process of lidar signals at tree level. Tree metrics are derived from Lidar data and used in a regression process for both timber volume and above ground biomass estimations. This process was first applied using only metrics obtained from the processing of a canopy height model (CHM) and then using additional metrics derived from aggregated full-waveform lidar signals. The accuracy of the obtained models is compared in order to evaluate the interest of full-waveform derived metrics for tree parameter assessment.

## 2. MATERIAL

The study area is located in the southern French Alps close to Digne-les-Bains (Alpes-de-Haute-Provence). This mountainous area was afforested in the end of the nineteenth century with black pine (*Pinus nigra ssp. nigra* [Arn.]) for fighting against erosion.

The data acquisition was performed using a RIEGL# LMS-Q560 system. This sensor is a small-footprint full-waveform airborne laser scanner. The lidar shot density was about 5 pts/m². The data set includes the raw signals on one side (1D intensity profiles), and the 3D points cloud resulting from the decomposition of raw signals by fitting sums of Gaussian [1] on the other side.

Field inventory data were collected on 6 circular plots of 15 m radius (61 trees). Within each plot, the accurate tree position, tree dbh (using a tape), total tree height and crown




(a) (b)

**Fig.1**. (a) Identification of tree crowns (shade) and total height (star) on a plot. (b) The aggregated signal (left), the corresponding vegetation profile after correction of the laser attenuation (right), and derived metrics.

base height (using a Vertex III clino meter, Haglöff, Sweden) were measured. The reference timber volume was calculated using the following equation (1) developed by INRA, the French National Institute for Agricultural Research, for Pinus nigra within France:

$$Volume = 34111.14 + 0.020833846 \times H \times C^2 - 1486.2307 \times C + 2.2695012 \times C \times H + 15.664201 \times C^2 - 56.250923 \times H - 0.0061317691 \times H^2 \quad (1)$$

where Volume is in cm3, H is the total height in cm, C is the trunk circumference at breast height in cm.

Due to the lack of equation for estimating above ground biomass for black pine in France, we chose to compare our estimates with two different reference models. The first one was used by Gil et al. [3] for Pinus nigra within Spain:

$$BiomassA = 0.6073 \times DBH^2 - 5.0998 \times DBH - 23.729 \quad (2)$$

where BiomassA is in kg, and DBH is the trunk diameter at breast height in cm.

The second reference model is more general. It was developed by Jenkins et al. [4] and used by Popescu [5] for pine trees within the USA:

$$BiomassB = \exp(-2.5356 + 2.4349 \times \log(DBH)) \quad (3)$$

where BiomassB is in kg, and DBH is the trunk diameter at breast height in cm.

## 3. METHOD

### 3.1. Deriving traditional Lidar metrics from CHM

A Canopy Height Model (CHM) was obtained by subtracting the Digital Terrain Model (DTM) to the Digital Surface Model (DSM), both resulting from point cloud processing. Segmentation was then performed on the CHM in order to identify two tree metrics [6]. The individual tree crowns were recognized as ellipses and each crown area ($Acrown_{CHM}$) was extracted from the attributes of the GIS vector layer. Total height ($Ht_{CHM}$) of each tree was considered to be the local maximum in each ellipse (see Fig. 1).

### 3.2. Deriving additional metrics from full-waveform signals

All the signals included into a tree crown identified in 3.1 were adjusted in altitude and summed. The resulting aggregated waveforms were then corrected from the laser beam attenuation using a logarithmic function [7] in order to produce vegetation profiles. Several new metrics were finally derived from both aggregated signals (SIG) and vegetation profiles (PROF), such as signal and profile integrals from the tree top to 3 m above ground ($I_{SIG}$, $I_{PROF}$), ratio between the whole integral and the ground component integral ($R_{SIG}$,$R_{PROF}$), height of the maximum profile amplitude in the crown ($Hmax_{PROF}$) and profile crown base height ($Hcrown_{PROF}$) (see Fig. 1). Because the vegetation profile corrects the signal from the light attenuation, it

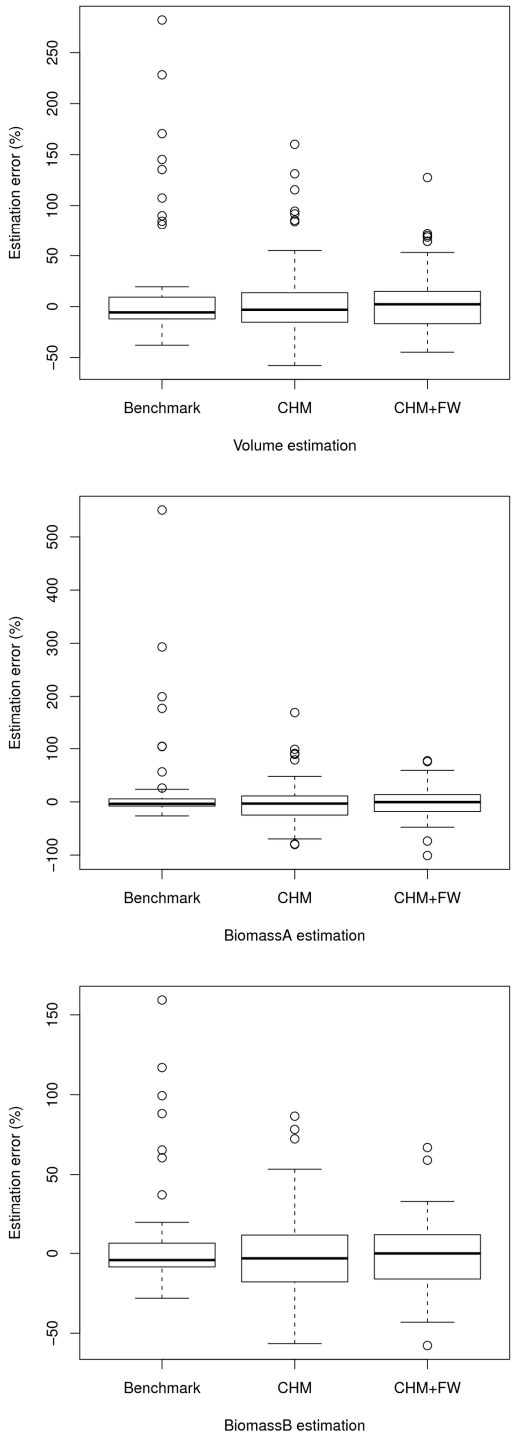



**Fig. 2**. Error distribution of volume (top), biomassA (middle) and biomassB (bottom) estimations, for benchmark, CHM and CHM+FW models.

provides a good description of the vertical structure of a tree [8] but it does not keep amplitude variability between trees (related to the crown dimension, i.e. signal counts falling inside a tree crown). At the opposite, a tree signal does not provide an accurate vertical profile distribution of tree components, but kept the signal amplitude comparable between trees (A broader tree crown will produce a higher signal amplitude).

### 3.3. Regression models

Among all the input metrics the most explanatory ones were first automatically selected using a stepwise regression algorithm minimizing the Akaike Information Criterion (AIC) [9]. Then, metrics selected in the three models (Volume, BiomassA and BiomassB) at the same time were retained. Using these metrics, final linear models were computed for CHM metrics on the one hand and with additional full-waveform metrics on the other hand. The adjusted determination coefficients (adj$R^2$) were calculated to compare model performances while taking into account of the difference in number of explanatory variables in the model. The Mean Absolute Percentage Error (MAPE) was also calculated in order to express the error of volume and biomass estimations in a generic percentage term.
Because the reference volume and biomass values were calculated with non-linear equations, we also computed, as a benchmark, linear regressions with reference field metrics, i.e. total height ($Ht_{ref}$) or DBH ($DBH_{ref}$) in order to assess the bias due to the linear model form.

## 4. RESULTS AND DISCUSSION

The stepwise algorithm kept both $Acrown_{CHM}$ and $Ht_{CHM}$ as CHM metrics for Volume, BiomassA and BiomassB estimates in the following model:

$$Estimate_{CHM} = a.Ht_{CHM} + b.\ Acrown_{CHM} + c \quad (5)$$

Regarding CHM+FW metrics, $Ht_{CHM}$, $Hcrown_{PROF}$, $Hmax_{PROF}$ and $I_{SIG}$ were selected for estimating Volume, BiomassA and biomassB in the following model:

$$Estimate_{CHMFW} = a.Ht_{CHM} + b.Hcrown_{PROF} + c.I_{SIG} + d \quad (6)$$

Adj$R^2$ and MAPE are summarized in Table 1. As expected, highest adj$R^2$ were obtained with the benchmark regressions. Using linear models instead of non-linear ones lead to a decrease in adj$R^2$ of 2 % to 5 % when timber volume and biomass are estimated from field measurements. Estimations performed with CHM+FW metrics are always more accurate than those performed with only CHM metrics leading to an increase in adj$R^2$ of 3% (Volume), 5% (BiomassA) and 4% (BiomassB). MAPE values are 3%, 7% and 5% better respectively with slightly less spread distribution errors (see

Fig. 2). This can be attributed to a better description of the tree shape as full-waveform metrics give information along the vegetation columns while the CHM metrics only refer to total tree height and crown area.
If the presented algorithm was applied on a mixed forest, further developments would have been necessary to identify tree species prior using a specific equation on each species. Another problem would have occurred in multilayered forests, where it would have been even more difficult to identify understorey trees, even if some studies were carried out to this purpose (e.g. [10]).

## 5. CONCLUSION

We demonstrated that using full-waveform lidar metrics in addition to CHM derived metrics improved timber volume and biomass estimates at tree level. However, further investigations are needed to further improve models. To that aim deriving other metrics from the lidar signal could be a clue. Improvement could also rely on the use of non linear models for volume and biomass estimates from lidar data; a (height x crown area) term might give better results. In order to conclude on the usefulness of full-waveform lidar signal in regard to discrete return data, a complementary study will be performed using metrics derived from discrete lidar point cloud, i.e. widely used height percentiles, in addition to CHM metrics.

## 6. REFERENCES

[1] A. Chauve, C. Vega, S. Durrieu, F. Bretar, T. Allouis, M. Pierrot-Deseilligny, W. Puech, "Advanced fullwaveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest", *International Journal of Remote Sensing*, vol. 30:19, pp. 5211-5228, 2009.

[2] Mallet, C. and Bretar, F., 2009. "Full-waveform topographic lidar: State-of-the-art", ISPRS *Journal of Photogrammetry and Remote Sensing*, vol. 64, pp. 1 – 16

[3] Gil, M. V. and Blanco, D. and Carballo, M. T. and Calvo, L. F., 2010. "Carbon stock estimates for forests in the Castilla y León region, Spain. A GIS based method for evaluating spatial distribution of residual biomass for bio-energy", *Biomass and Bioenergy*, vol. In Press, Corrected Proof, pp. –

[4] Jenkins, J. C. and Chojnacky, D. C. and Heath, L. S. and Birdsey, R. A., 2003. "National-scale biomass estimators for United States tree species", *Forest Science*, vol. 49, pp. 12-35

[5] Popescu, S. C., 2007. "Estimating biomass of individual pine trees using airborne lidar", *Biomass and Bioenergy*, vol. 31, pp. 646 – 655

[6] Véga C. and Durrieu S., Accepted. "Multi-level filtering segmentation to measure individual tree parameters based on Lidar data: application to a mountainous forest with heterogeneous stands". *International Journal of Applied Earth Observation and Geoinformation*

[7] Allouis, T., Durrieu, S., Cuesta, J., Chazette, P., Flamant, P. H., Couteron, P., 2010. "Assessment of tree and crown heights of a maritime pine forest at plot level using a fullwaveform UltraViolet Lidar prototype" , pp. 1382 -1385 *2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*,

[8] Allouis, T., Durrieu, S., Bailly, J., Chazette, P., Cuesta, J., Flamant, P., Couteron, P., In Review. "Potential of an ultraviolet, medium-footprint lidar prototype for retrieving tree heights and tree planting patterns", *ISPRS Journal of Photogrammetry and Remote Sensing*

[9] Akaike, H., 1980. "Likelihood and the Bayes procedure, Bayesian Statistics", Ed. Bernardo et. al., University Press, 1980

[10] Jaskierniak, D. and Lane, P. N. and Robinson, A. and Lucieer, A., 2011. "Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions", *Remote Sensing of Environment*, vol. 115, pp. 573 - 585

Table 1. Regression results. AdjR² = adjusted R², MAPE = Mean Absolute Percentage Error.

| Model | Metrics | Volume (Eq. 1) | | BiomassA (Eq. 2) | | BiomassB (Eq. 3) | |
|---|---|---|---|---|---|---|---|
| | | adjR² | MAPE (%) | adjR² | MAPE (%) | adjR² | MAPE (%) |
| Benchmark | $Ht_{ref}$ , $DBH_{ref}$ | 0.94 | 31 | 0.97 | 31 | 0.96 | 17 |
| CHM | $Acrown_{CHM}$, $Ht_{CHM}$ | 0.89 | 27 | 0.85 | 28 | 0.85 | 22 |
| CHM + FW | $Ht_{CHM}$, $Hcrown_{PROF}$, $I_{SIG}$ | 0.92 | 24 | 0.90 | 21 | 0.89 | 17 |

# Stem volume and above-ground biomass estimation of individual pine trees from lidar data: contribution of full-waveform signals

Tristan Allouis[a,d], *Student Member IEEE*, Sylvie Durrieu[a], Cédric Véga[a,b], Pierre Couteron[c]

[a] *Cemagref, UMR TETIS, 500, rue J.F. Breton BP 5095, 34196 Montpellier Cedex 5, France*
[b] *French Institute of Pondicherry, Geomatics and Applied Informatics Laboratory (LIAG), 11 Saint Louis Street, Pondicherry 605 001, India*
[c] *Institut de Recherche pour le Développement, UMR AMAP, Bd de la Lironde TA A51/PS2, 34398 Montpellier cedex 5, France*
[d] *L'Avion Jaune, 1 chemin du Fescau, 34980 Montferrier sur Lez France*

*Corresponding author: T. Allouis (e-mail: tristan.allouis@lavionjaune.fr)*

**Abstract**: The diameter at breast height (DBH) is the most extensively measured parameter in the field for estimating stem volume and aboveground biomass of individual trees. However, DBH can not be measured from airborne or spaceborne light detection and ranging (lidar) data. Consequently, volume and biomass must be estimated from lidar data using other tree metrics. The objective of this paper is to examine whether new full-waveform (FW) lidar data can improve volume and biomass estimation of individual pine trees, when compared to traditional lidar data. Sets of metrics are derived from traditional canopy height model (CHM-only metrics), from the vertical distribution of traditional discrete-return lidar data (CHM+DR metrics), and from new full-waveform lidar data (CHM+FW metrics). In each set, the most relevant and non-collinear metrics were selected using a combination of best subset and variance inflation factor methods, in order to produce predictive models of volume and biomass. CHM-only metrics (tree height and tree bounding volume (tree height x crown area) provided volume and biomass estimates of individual trees with an error of 19% and 30%, which is equivalent to previews studies. CHM+FW metrics did not improve stem volume estimates, but they significantly increased the accuracy of aboveground biomass estimates by 9% compared with the CHM-only model. The approach is limited to the delineation of individual trees. However, the results highlight the potential of full-waveform lidar data to improve aboveground biomass estimates through a better integration of branch and leaf biomass than with traditional lidar data.

## I. Introduction

Stem volume and aboveground biomass are two commonly used parameters for the assessment of forest productivity and carbon sequestration rates, and they are needed to define effective plans for sustainable forest management and climate change mitigation. While aboveground biomass has been identified as an essential climate variable, its estimation is usually performed by considering allometric relationships for which field-measured tree diameter at breast height (DBH) and total height are used as the main input variables. Forest inventories based on field sampling approaches can meet the Kyoto Protocol and REDD (Reducing Emissions from Deforestation in Developing nations) requirements, but spatially explicit field measurements at the forest stand scale are required in order to implement sustainable management practices and to perform detailed ecological studies. Therefore, improved assessment of how anthropogenic and environmental factors affect the distribution of aboveground biomass [1] has become crucial.

Remote sensing data with field control measurements provide effective solutions to assess spatially explicit inventories from individual tree to global scale and in inaccessible zones. Among the existing remote sensing techniques, lidar (light detection and ranging) delivers useful information to describe the vertical structure of vegetation [2], [3] with a lower sensitivity to signal saturation compared with radar instruments [4]. Because airborne and spaceborne lidar sense the forest from above, they do not provide direct access to the tree DBH. Consequently, other tree metrics must be derived from lidar data so that they can be used in allometric relationships for the estimation of volume and biomass.

Stand volume and aboveground biomass estimations are mainly based on metrics derived from the signal amplitude [5] when large footprint profiler sensors (laser beam diameter of 10 to 70 m on the ground) are used. The signal backscattered by the ground and the canopy can also be integrated. These integrals and their ratio may be used in addition to the signal amplitude, [6]. Inventories are performed at a higher spatial resolution using small-footprint lidar data (in tens of centimeters, with a measurement density varying from 0.5 to several per m² [8], [9]) acquired by airborne scanner sensors. At the plot level, volume and biomass are correlated with metrics derived from the vertical distribution of echoes (height percentiles), maximum height, mean height, and coefficient of variation of heights [10], [11]. Studies have also been performed at the individual tree level, allowing to directly link field measurements with lidar metrics. Some studies have estimated tree biomass using height percentiles (*e.g.*, [12], [13]). Popescu [14] assessed biomass using total height and crown diameter derived from a canopy height model (CHM), and Chen et al. [15] introduced the use of a "canopy geometric volume" metric derived from CHM for stem volume estimation. All these research efforts attest to the utility of measuring the primary structural characteristics of the vertical structure of trees including the crown dimensions and the total height.

The recent development of small-footprint full-waveform lidar sensors has provided new opportunities for vegetation studies [16]. Unlike traditional small-footprint discrete-return lidars that record only the most significant echoes, this new generation of sensors records the entire backscattered energy. New developments in signal processing lead to an increased number of extracted echoes compared with discrete-return systems [17], thus providing additional information such as echo shape or area under the waveform that might be used to refine vegetation structure measurements [18], [19].

In this paper, we examine whether full-waveform lidar metrics combined with usual lidar variables (*i.e.*, total tree height and tree crown diameter) can improve volume and biomass estimates at the

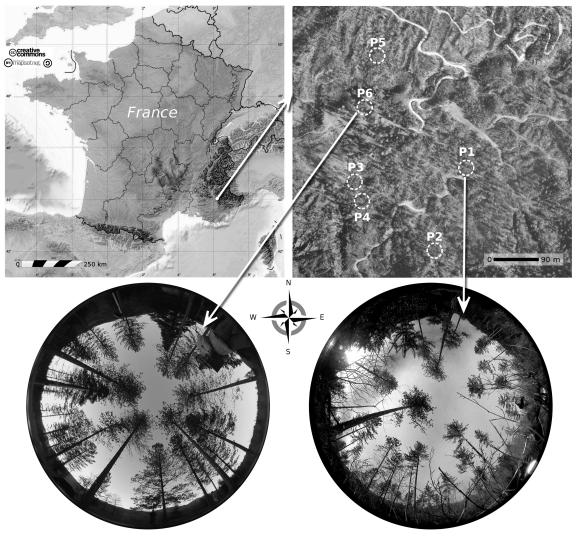
tree level. These estimates were compared with those performed using the usual variables alone or in combination with percentile metrics derived from discrete-return lidar data.

## II. Materials

### A. Study site

The study area is located in the Haute-Bléone forest (44.16°N, 6.32°W) in the southern French Alps (see Fig. 1), and close to the city of Digne-les-Bains. This 108 ha of mountainous area was afforested in the late nineteenth century with black pine (*Pinus nigra ssp. nigra* [Arn.]). Today, stand densities varies from 100 stems/ha to over 2,000 stems/ha. The soil on these black marl badlands is affected by intense erosion, but the afforestation was found to substantially reduce erosion by trapping eroded material [20]. Today, the stands are over-mature, and management plans must be developed to renew the forest. Due to particularly difficult access conditions, a lidar flight was performed to assess the potential of such data for forest mapping and management. This study site is particularly suited to the assessment of lidar efficiency in various terrain conditions because the mean slope is approximately 53% and can locally reach up to 100%.

Fig. 1. Study site and plot (white dot circles) location. Hemispherical photos show plots condition.




### B. Reference data

Field inventory data were collected during December 2007 on six circular plots with a 15 m radius. The plots are representative of several stand development stages. The mean plot heights ranged from 9 m to 22 m. Thanks to the homogeneous composition of the stand, we could perform our study using 61 reference trees [14], [21].

Each plot center was geo-located using a differential GPS and a total station. The following characteristics were measured for all trees having a diameter at breast height (DBH) greater than 7 cm: tree position (*i.e.*, X and Y coordinates calculated using distance and angle measurements from the plot center), tree DBH (measured using a tape), tree height and crown base height (measured using a Vertex III ultrasonic clinometer (Haglof Inc)). The plot characteristics are summarized in Table I. On two plots, the maximum extent of tree crowns was measured in two directions (*i.e.*, North-South and East-West) using a crown mirror and a tape.

The reference stem volumes were calculated for each tree from ground-measured DBH and total height using the following equation developed for *Pinus nigra* within France by the French Institute for Agricultural Research [22]. The equation was built using a destructive sampling of 1755 trees mainly located in the southeast of France, with DBH ranging from 4 cm to 30 cm and heights ranging from 10 m to 55 m.

$$V = 34111.14 + 0.020833846 \cdot H \cdot C^2 - 1486.2307 \cdot C + 2.2695012 \cdot C \cdot H + 15.664201 \cdot C^2 - 56.250923 \cdot H - 0.0061317691 \cdot H^2 \quad (1)$$

where V is the stem volume in $cm^3$, H is the total height in cm, and C is the trunk circumference at breast height in cm.

No allometric equation was available for aboveground biomass (AGB) estimation of black pine in France. Because our study site is within the Mediterranean climate zone, we used an equation developed for black pine in Spain, which is principally located in the Mediterranean climate zone. The model was established by Montero [23] on 253 *Pinus nigra* trees located in various Spanish regions, with DBH ranging from 7 cm to 70 cm.

$$AGB = 0.6073 \cdot DBH^2 - 5.0998 \cdot DBH - 23.729 \quad (2)$$

where AGB-level is in kg, and DBH is the trunk diameter at breast height in cm.

Reference volume and biomass were computed of each individual tree using the previous equations. Averaged values are given in Table I along with other mean tree characteristics.

TABLE I
Plot characteristics. Mean values of individual trees ± standard deviation are given

| Plot | Number of trees | DBH (cm) | Total Height (m) | Crown base height (m) | Mean terrain slope (%) | Stem Volume V ($dm^3$) | AGB (kg) |
|---|---|---|---|---|---|---|---|
| Plot 1 | 7 | 26.7 ± 1.2 | 17.7 ± 0.6 | 11.2 ± 1.0 | 11.0 | 489 ± 54 | 274 ± 33 |
| Plot 2 | 10 | 31.5 ± 3.7 | 19.1 ± 1.5 | 12.6 ± 1.2 | 22.5 | 746 ± 214 | 426 ± 127 |
| Plot 3 | 10 | 27.2 ± 4.1 | 18.3 ± 2.1 | 10.8 ± 1.8 | 42.5 | 548 ± 216 | 296 ± 126 |
| Plot 4 | 8 | 33.5 ± 4.9 | 22.4 ± 2.2 | 15.7 ± 2.3 | 20.0 | 976 ± 319 | 500 ± 172 |
| Plot 5 | 13 | 16.9 ± 3.7 | 9.0 ± 1.1 | 4.5 ± 0.8 | 20.5 | 115 ± 73 | 72 ± 68 |
| Plot 6 | 13 | 29.8 ± 6.2 | 16.7 ± 2.2 | 6.2 ± 0.8 | 7.5 | 616 ± 279 | 386 ± 195 |
| Total | 61 | 27.1 ± 7.1 | 16.6 ± 4.6 | 9.5 ± 4.1 | 20.7 | 553 ± 340 | 313 ± 194 |

### C. Lidar data

The data acquisition was performed in April 2007 using a RIEGL LMS-Q560 system. This sensor is a small-footprint full-waveform airborne laser scanner (see Wagner et al. [24] for system design and specifications) which was operated at a pulse rate of 111

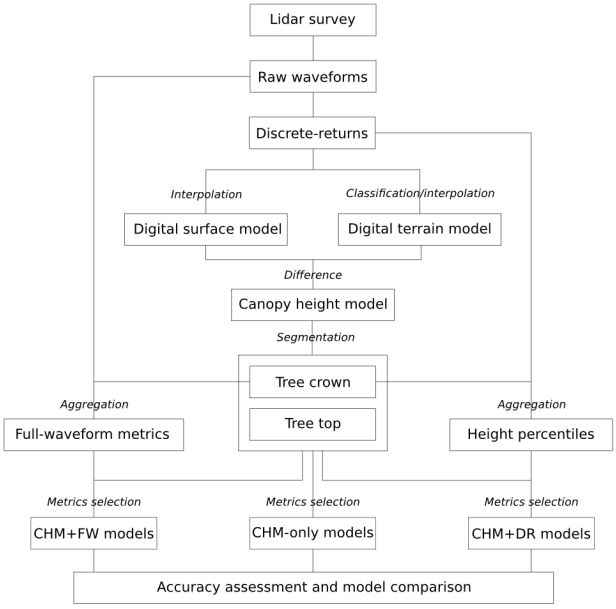
kHz with a scan angle of ±22.5°. The flight altitude was approximately 600 m above ground level, leading to a footprint diameter of approximately 0.25 m. The average measurement density was approximately 5 shots/m² (*i.e.*, 5 waveforms/m²).

The dataset was composed of the raw full-waveform signals (1D intensity profiles digitized at 1 GHz (15 cm)) and the discrete returns (3D point cloud) extracted from the raw signals by fitting sums of Gaussian functions [17].

## III. Methods

An overview of the method is provided in Fig. 2 and a summary of retrieved metrics is provided in Table II.

Fig. 2. Method overview.

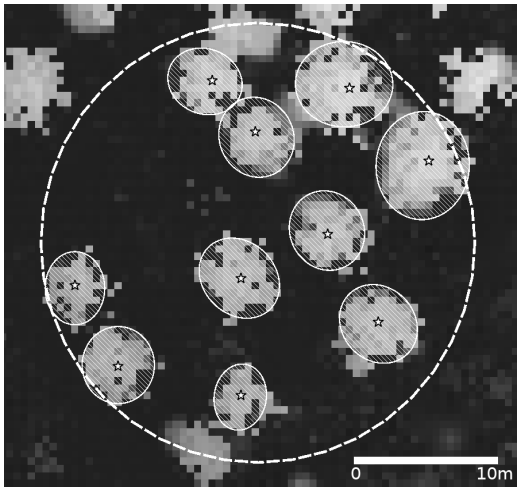



*A. Deriving metrics from the CHM*

A 0.5 m CHM resolution was obtained by subtracting the Digital Terrain Model (DTM) from the Digital Surface Model (DSM), both resulting from lidar point cloud processing. Each of the 0.5 m DSM cells were assigned the elevation of the highest echo within it, and the Inverse Distance Weighting (IDW) interpolation was used to compute empty cell values. Due to a lower density of ground points, the DTM was computed on a 1 m grid using the ordinary kriging interpolation of last echoes that actually reached the ground, and identified after an echo classification scheme described in Véga et al. [25].

Segmentation was performed on the CHM to extract and characterize individual tree crowns. The tree crown extraction was based on the identification of local maxima on a smoothed version of the CHM (using a Gaussian low-pass filter) and the subsequent modeling of elliptical crowns issued from crown dimension measurements in various directions (see Véga and Durrieu [26] for technical details) (Fig. 3).

Fig. 3. Identification of tree crowns (shade) and crown apex (star) on a 15 m radius plot (dashed).




For each crown, the tree total height ($Ht_{CHM}$) was estimated by selecting the value of the highest pixel of the CHM inside the modeled crown, and the crown projected area ($Acrown_{CHM}$) computed from the previously determined elliptical crowns. From these two metrics, we also derived the tree bounding volume ($TBV_{CHM}$), which is similar to the "canopy geometric volume" [15]. $TBV_{CHM}$ is the elliptic cylinder of height $Ht_{CHM}$ and base $Acrown_{CHM}$ ($TBV_{CHM} = Acrown_{CHM} \times Ht_{CHM}$). The accuracy of the metrics derived from the CHM was estimated using ground-truth measurements.

*B. Deriving metrics from the discrete-return lidar data (DR metrics)*

Height percentiles are related to the vertical distribution of lidar echoes reflected on the vegetation. The elevations of discrete-returns located inside each modeled tree crown were subtracted from the corresponding terrain elevation retrieved from the DTM. To remove the influence of undergrowth vegetation, only echoes higher than 3 m above the ground were considered. According to previous studies (*e.g.*, [12], [13]), 25th ($H25_{DR}$), 50th ($H50_{DR}$), 75th ($H75_{DR}$), 90th ($H90_{DR}$) height percentiles (or quantiles), mean height ($Hmean_{DR}$) and coefficient of variation ($CV_{DR}$) were retrieved from the distribution of echoes and were retained as discrete-return (DR) candidate variables.

*C. Deriving metrics from full-waveform signals and vegetation profiles (FW metrics)*

The raw small-footprint waveforms were georeferenced in a global coordinate system using the GPS position of the plane and the lidar scan angle. Waveforms that were fully located within modeled tree crowns were then summed to obtain a cumulative waveform per tree (Fig. 4). Five footprints (30 cm in diameter) per square meter were considered to provide the same representative measurements of the forest structure as five footprints of 56 cm diameter (providing a 100% coverage of the 30 m plots).

We also derived a vegetation profile from the cumulative waveform. The amplitude of the cumulative waveform was first corrected to account for the reflectance difference between ground and vegetation. According to an analysis of the signal dynamics (transmitted and received amplitude ratio) performed on the ground and vegetation single echoes, the average ground reflectance was 0.52

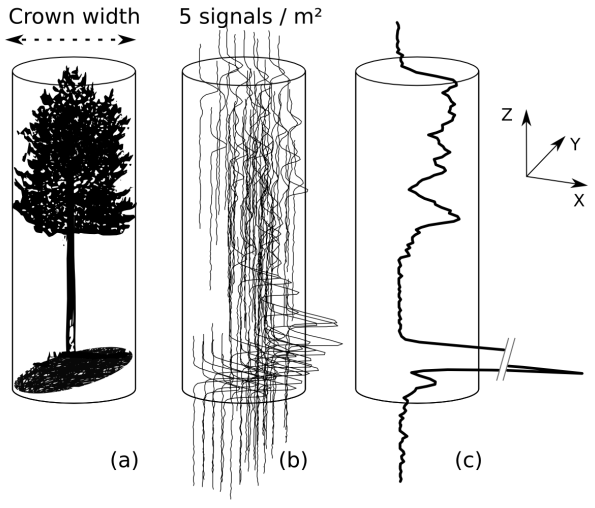
and the vegetation reflectance was 0.33 on our study site. Then, the cumulative waveform was corrected from the laser beam attenuation along its travel through the vegetation. To achieve this light attenuation correction we used a model previously adapted [30], [31] from the MacArthur and Horn equation [29]. This model transforms the cumulative waveform into a vegetation profile to provide an accurate description of the vertical structure of a tree:

$$D_{r_i} = \ln\left(\frac{\sum_{r_i}^{r_{max}} P(r)}{\sum_{r_{i+1}}^{r_{max}} P(r)}\right) \quad (3)$$

Where $D_{r_i}$ is the vegetation profile amplitude at range $r_i$ from the plane with $i \in \left\lfloor 1, r_{max-1} \right\rfloor$, and $P(r)$ is the recorded amplitude as a function of range. The correction is computed for each interval between two successive ranges from the atmosphere to signal extinction ($r_{max}$).

Cumulative waveforms and vegetation profiles provide complementary information to describe the structure of tree crowns. Indeed, vegetation profiles do not account for absolute signal amplitude, and large tree crowns (in area projected on the ground) cannot be distinguished from small ones. However, vegetation profiles provide an improved description of the vertical structure of a tree. Cumulative waveforms allow us to distinguish crown width because the signal amplitude tends to increase with the crown projected area. However, cumulative waveforms do not provide a normalized vertical distribution of the vegetation structure.

Fig. 4. Sum of full-waveform signals falling within crown boundaries (represented by the cylinder). a) scheme of a tree, b) signal selection and c) resulting cumulative waveform.

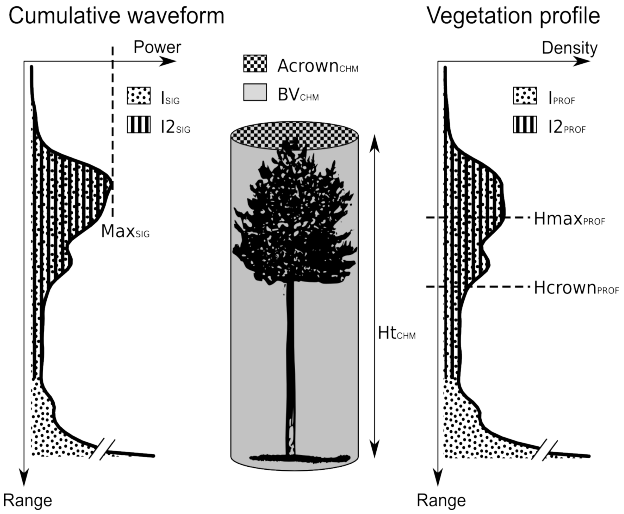



Several metrics were consequently derived from both cumulative waveforms and vegetation profiles. These metrics are summarized in Table II and illustrated in Fig. 5. $Max_{SIG}$ was determined as the maximum amplitude of each cumulative signal (except the ground component, *i.e.* higher than 3 m above the last lidar echo), which was consequently related to the number of signals located within tree boundaries. $Hmax_{PROF}$ was the height of the maximum profile amplitude (except the ground component) corresponding to the height where the crown is either the largest or the densest. The crown base height ($Hcrown_{PROF}$) corresponds to the greatest decrease in the profile, detected as the position of the minimum value of the profile derivative [31]. Signal integrals, corresponding to areas under the curve, were computed for both the cumulative waveform and profile, from a few meters above the tree to the ground ($I_{SIG}$ and $I_{PROF}$). Similarly, $I2_{SIG}$ and $I2_{PROF}$ were computed from a few meters above the tree to three meters above ground to remove the influence of undergrowth. $R_{SIG}$ and $R_{PROF}$ were then computed as ratios between $I_{SIG}$ and the energy reflected on the ground (area under the ground component curve) and the $I_{PROF}$ and profile ground component, respectively. The result is a crown porosity index.

TABLE II
SUMMARY OF THE METRICS DERIVED AT THE TREE SCALE USED IN THIS STUDY

| Metric ID | | Metric description |
|---|---|---|
| Canopy height model (CHM metrics) | $Ht_{CHM}$ | Total tree height extracted from the CHM |
| | $Acrown_{CHM}$ | Tree crown projected area extracted from the CHM |
| | $TBV_{CHM}$ | Tree bounding volume $TBV_{CHM} = Acrown_{CHM} \cdot Ht_{CHM}$ |
| Discrete-returns (DR metrics) | $H25_{DR}$ | 25th height percentile |
| | $H50_{DR}$ | 50th height percentile |
| | $H75_{DR}$ | 75th height percentile |
| | $H90_{DR}$ | 90th height percentile |
| | $Hmean_{DR}$ | Mean height of discrete-returns |
| | $CV_{DR}$ | Coefficient of variation |
| Cumulative signal (FW metrics) | $Max_{SIG}$ | Maximum signal amplitude (except ground component, *i.e.* higher than 3 m above the last lidar echo) |
| | $I_{SIG}$ | Signal integral |
| | $I2_{SIG}$ | Signal integral from the air to 3 m above-ground |
| | $R_{SIG}$ | Ratio between $I_{SIG}$ and the integral of the ground component |
| Vegetation profile (FW metrics) | $Hcrown_{PROF}$ | Crown base height |
| | $Hmax_{PROF}$ | Height of the maximum profile amplitude (except ground component) |
| | $I_{PROF}$ | Profile integral |
| | $I2_{PROF}$ | Profile integral from the air to 3 m above-ground |
| | $R_{PROF}$ | Ratio between $I_{PROF}$ and the integral of the ground component |

Fig. 5. The cumulative waveform (left), the corresponding vegetation profile after correction of laser attenuation (right), and the derived metrics.




*D. Building and comparing models for volume and biomass assessment*

Regression models were computed to estimate volume and biomass using metrics derived from the CHM alone (CHM-only) or combined either with discrete-return metrics (CHM+DR) or with full-waveform metrics (CHM+FW). The estimates were then compared with ground-truth values (Eqs. (1) and (2)). In accordance with some previous research, we developed linear regression models [10], [11].

*1) Metric selection:* Thanks to the relatively low number of candidate variables among the CHM-only, CHM+DR or CHM+FW metrics, predictive linear models were computed using an all-possible-subset regression. The objective is to find the linear combination of metrics that provides the best estimation of volume (1) or biomass (2) according the Bayesian information criterion (BIC) [32]:

$$BIC = k \cdot \ln(n) - 2 \cdot \ln(L) \quad (4)$$

where k is the number of parameters in the model (including the intercept), n is the number of observations, and L is the maximized value of the likelihood function.

By measuring the model goodness-of-fit while imposing a penalty for increasing the number of explanatory variables, the BIC addresses the problem of an artificially increased likelihood when a new variable is added in the model. The all-possible-subset regression was automatically performed thanks to the R software function called 'bestglm' [33] using the whole set of observations (61 trees). The set of variables in the model having the lowest BIC was retained as the "best subset".

While indicators related to the same tree inevitably have some inter-dependence , the collinearity between metrics can lead to model over-fitting, and this is liable to lead to weak predictions when using a new dataset. Consequently, collinearity in a model was verified for each of the all-possible-subset models by computing the Variance Inflation Factor (VIF). To achieve this, each metric was expressed as a linear combination of the other metrics in the subset, and the VIF of each metric was computed as:

$$VIF = \frac{1}{1 - R^2} \quad (5)$$

where $R^2$ is the coefficient of determination of the linear model developed for each metric of a given subset.

The set of variables having the lowest BIC and with all variables having a VIF < 10 [34] was retained as the "best independent subset".

*2) Building final models and assessing the estimates accuracy:* A cross-validation was performed to build and assess the accuracy of the final predictive models of volume and biomass using metrics selected in the "best subset" and in the "best independent subset". In compliance with the rules of the leave-one-out cross-validation method (LOOCV) [35], a single observation was retained from the data set as the validation data, and the remaining observations were used as training data. The process was repeated to use each observation once as the validation data, thus assessing the ability of the model to fit independent measurements.

The mean absolute percentage error (MAPE) was used to express in a generic percentage term the error of volume and biomass estimations:

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{A_i - E_i}{A_i} \right| \quad (7)$$

where $A_i$ is the actual value and $E_i$ is the estimated value for each step i of the LOOCV.

For each step of the LOOCV a single value of the adjusted coefficient of determination (adj$R^2$) was also calculated, in order to evaluate the model's goodness-of-fit while taking into account the differences in the number of explanatory variables in the model:

$$adjR^2 = 1 - \frac{\sum_{j=1}^{n} (y_j - \hat{y}_j)^2}{\sum_{j=1}^{n} (y_j - \bar{y})^2} \times \frac{n-1}{n-p-1} \quad (6)$$

where $y_j$ is the observed value, $\hat{y}_j$ is the modeled value, $p$ is the number of metrics in the model and $n$ is the number of ($y_j$, $\hat{y}_j$) pairs. The final adj$R^2$ value was obtained as the mean of each adj$R^2$ value.

Model coefficients were also averaged in order to produce final predictive models.

## IV. Results

*A. Accuracy of CHM metrics*

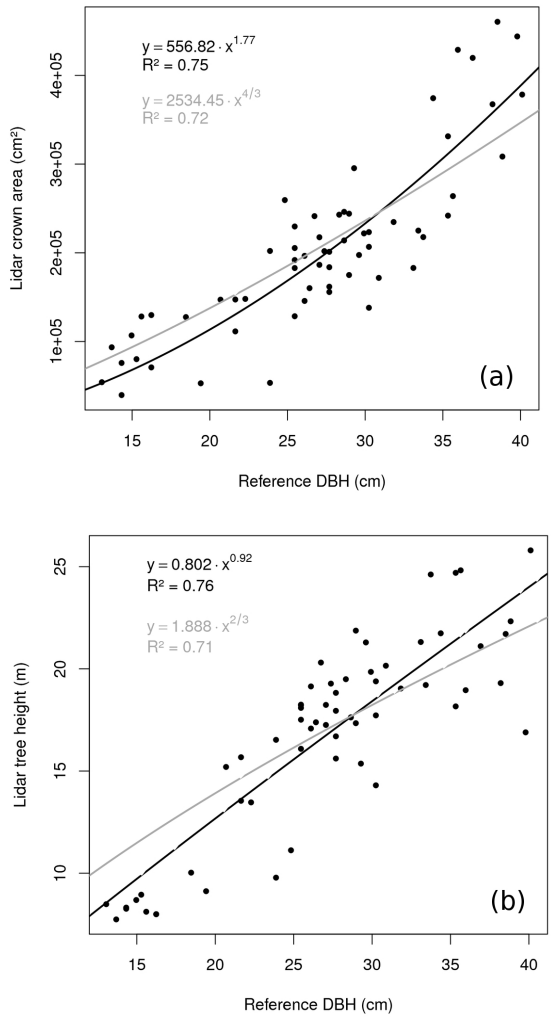
Tree heights were retrieved from the CHM with a bias of 0.09 m and a standard deviation of 0.57 m. The corresponding RMSE was also 0.57 m. The mean crown diameter was estimated with a 1.48 m RMSE on the two reference plots where the extents of the crowns were measured.

*B. Lidar metrics in allometric models*

Theoretically, the crown area was found by West et al. [27] to equate to $DBH^{4/3}$, while empirical data in a highly diverse neo-tropical forest yielded exponent values within the range of 1.33-1.39 [28], which is close to the theoretical value of 4/3. In our case, a 1.77 exponent was found instead of 1.33 (4/3) (Fig. 6) by fitting a power model on actual data. However, the strong relationship between crown dimension and the DBH was confirmed ($R^2$=0.75). The relationship between tree height and DBH was also verified using a power model (Fig. 6) ($R^2$=0.76), but the exponent value of 0.92 was not very close to

West et al.'s [27] theory (0.67) or to Muller-Landeau et al.'s [28] empirical data (0.59-0.60). Such relationships are probably a consistant property of forest trees, even if species- and context-specific features can be expected to modulate the exponent values. Consequently, it makes sense to consider tree height and crown area as analogous to height and DBH, and to use lidar-derived metrics to predict tree biomass from airborne or spaceborne platforms.

Fig. 6. Power regression models relating field-measured DBH and (a) crown area retrieved from the CHM and (b) tree height retrieved from the CHM. The gray curves were fitted using (a) the 4/3 exponent of West et al. (2009) and (b) the 2/3 exponent of the same study.




*C. Volume estimation models*

The models' characteristics are summarized in Table III.

*1) CHM-only metrics:* The best-subset model was of the following form:

$$CHM_{VOL} = -70388.402 + 19149.571 \cdot Ht_{CHM} - 26979.4560 \cdot Acrown_{CHM} + 1691.0874 \cdot TBV_{CHM} \quad (8)$$

This model, whose coefficients were averaged from the 61 models developed using the LOOCV method, had an average adj$R^2$ of 0.93, leading to an estimation error of 16%, but it presented a collinearity between $Acrown_{CHM}$ and $TBV_{CHM}$ (VIF of 19.7 and 31.4, respectively). The best independent subset, which was the second best-subset, retained $Ht_{CHM}$ and $TBV_{CHM}$:

$$CHM_{VOL} = -238055.3118 + 27713.7796 \cdot Ht_{CHM} + 884.1049 \cdot TBV_{CHM} \quad (9)$$

This latter model had the same adj$R^2$ (0.93), but it provided less accurate estimates (MAPE of 19%).

*2) CHM-DR metrics:* The best-subset model had the following form:

$$CHM+DR_{VOL} = -94676.239 - 17038.179 \cdot Acrown_{CHM} + 1699.641 \cdot TBV_{CHM} - 79999.818 \cdot H25_{DR} + 100349.532 \cdot Hmean_{DR} \quad (10)$$

This model had an adj$R^2$ of 0.94 and an estimation error of 18%. In addition to $Acrown_{CHM}$ and $TBV_{CHM}$, a strong collinearity was found between $H25_{DR}$ and $Hmean_{DR}$ (102.5 and 110.2, respectively). The best independent subset was the 19th best subset. This subset had a lower adj$R^2$ (0.93) but provided slightly more accurate estimates (MAPE of 16%), and it had the following form, where $Ht_{CHM}$ was replaced by $H75_{DR}$ compared with the CHM-only model (Eq. 11):

$$CHM+DR_{VOL} = -194855.1173 + 933.4998 \cdot TBV_{CHM} + 27288.4157 \cdot H75_{DR} \quad (11)$$

*3) CHM-FW metrics:* The best-subset model had the following form:

$$CHM+FW_{VOL} = -12703.9856 + 18728.4275 \cdot Ht_{CHM} - 26979.4560 \cdot Acrown_{CHM} + 1691.0874 \cdot TBV_{CHM} + 22.0595 \cdot I2_{SIG} \quad (12)$$

This model had an adj$R^2$ of 0.95 and provided estimates with an error of 17%, but it contained collinear variables ($Acrown_{CHM}$ and $TBV_{CHM}$) for the first 520 best subsets. The 521$^{st}$ subset was selected as the best independent subset, with the following form, and provided a lower adj$R^2$ (0.93) as well as a higher error (21%):

$$CHM+FW_{VOL} = -281729.6368 + 31047.4463 \cdot Ht_{CHM} + 646.2851 \cdot TBV_{CHM} + 9.6198 \cdot I2_{SIG} \quad (13)$$

*4) Comparison of the different models:* We first noticed that the tree bounding volume metric ($TBV_{CHM}$) was selected in every estimation model. Despite the similar adj$R^2$ in the different models (ranging from 0.93 and 0.95), the best estimates were made using either CHM-only metrics, if we consider the best subset of variables, or with CHM+DR metrics, if we consider the best independent subset. However, the additional DR metrics did not seem to significantly increase the accuracy of the stem volume estimations. Indeed, the CHM+DR estimates are better when using the best independent subset (16% vs. 19%) but worse when using the best subset (18% vs. 16%). Additional full-waveform lidar data led to slightly worse estimates compared with the CHM-only models.

*D. Biomass estimation models*

The models' characteristics are summarized in Table III.

*1) CHM-only metrics:* The best-subset model had the following form:

$$CHM_{BIOM} = -53.1488 + 7.9519 \cdot Ht_{CHM} + 0.6278 \cdot TBV_{CHM} \quad (14)$$

This model, whose coefficients were averaged from the 61 models developed using the LOOCV method, had an average adj$R^2$ of 0.87, leading to an estimation error of 30%. This model did not produce any collinearity between variables and this best subset was also the best independent subset.

*2) CHM-DR metrics:* The best-subset model had the following form:

$$CHM+DR_{BIOM} = -110.1881 + 0.5872 \cdot TBV_{CHM} - 57.4843 \cdot H25_{DR} + 67.7806 \cdot Hmean_{DR} \quad (15)$$

This model had an adjR² of 0.88 and an estimation error of 32%. A strong collinearity was observed between $H25_{DR}$ and $Hmean_{DR}$ (VIF of 96.8 and 109.4, respectively). The best independent subset had a slightly lower adjR² (0.87) but provided more accurate estimates (MAPE of 30%); it had the following form, where the collinear variables were replaced by $Ht_{CHM}$:

$$CHM+DR_{BIOM} = -53.1488 + 7.9519 \cdot Ht_{CHM} + 0.6278 \cdot TBV_{CHM} \quad (16)$$

*3) CHM-FW metrics:* The best-subset model had the following form:

$$CHM+FW_{BIOM} = -27.6083 + 8.8632 \cdot Ht_{CHM} - 9.3945 \cdot Acrown_{CHM} + 0.6208 \cdot TBV_{CHM} + 0.0193 \cdot I2_{SIG} \quad (17)$$

This model had an adjR² of 0.91, and it provided estimates with an error of 26%, but it contains collinear variables ($Acrown_{CHM}$ and $TBV_{CHM}$). The third best subset was selected as the best independent subset, with the following form, providing a slightly lower adjR² (0.90) but also a lower error rate (21%):

$$CHM+FW_{BIOM} = -121.2705 + 13.1535 \cdot Ht_{CHM} + 0.2569 \cdot TBV_{CHM} + 0.0150 \cdot I2_{SIG} \quad (18)$$

*4) Comparison of the different models:* We first noticed that $TBV_{CHM}$ was selected in every model and $Ht_{CHM}$ was not selected in only one model. The adjR² is clearly higher for the CHM+FW models (0.90 – 0.91) than for the CHM-only (0.87) and CHM+DR models (0.87 – 0.88). The DR metrics did not seem to increase the accuracy of biomass estimations. The most accurate estimates were performed with CHM+FW metrics using either best subsets or best independent subsets, mainly due to the use of $I2_{FW}$ metric. The decrease in MAPE was 9% compared with both CHM-only and CHM+DR models when using the best independent subset. The decrease was 4% and 6% compared with the CHM-only and CHM+DR models, respectively, when using the best subset.

TABLE III

RESULTS OF METRICS SELECTION AND MODEL ACCURACY ASSESSMENT USING METRICS FROM THE "BEST-SUBSET" (RANK = 1, *I.E.* WITH LOWEST BAYESIAN INFORMATION CRITERION) AND FROM THE "BEST INDEPENDENT SUBSET" (BEST-SUBSET WITH ALL VARIABLES HAVING A VARIANCE INFLATION FACTOR (VIF) < 10). ADJUSTED R² (ADJR²) AND MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) ARE AVERAGED RESULTS OF THE CROSS-VALIDATION METHOD (61 VALIDATIONS DATA PER MODEL).

| Estimates | | Metrics subset (VIF) | Rank of the metrics subset | adjR² | MAPE |
|---|---|---|---|---|---|
| Volume | CHM-only | $Ht_{CHM}$ (4.7)<br>$Acrown_{CHM}$ (19.7)<br>$TBV_{CHM}$ (31.4) | 1 | 0.93 | 16% |
| | | $Ht_{CHM}$ (2.9)<br>$TBV_{CHM}$ (2.9) | 2 | 0.93 | 19% |
| | CHM+DR | $Acrown_{CHM}$ (26.7)<br>$TVB_{CHM}$ (38.5)<br>$H25_{DR}$ (102.5)<br>$Hmean_{DR}$ (110.2) | 1 | 0.94 | 18% |
| | | $TBV_{CHM}$ (2.5)<br>$H75_{DR}$ (2.5) | 19 | 0.93 | 16% |
| | CHM+FW | $Ht_{CHM}$ (4.7)<br>$Acrown_{CHM}$ (27.1)<br>$TBV_{CHM}$ (31.7)<br>$I2_{SIG}$ (6.0) | 1 | 0.95 | 17% |
| | | $Ht_{CHM}$ (3.3)<br>$TBV_{CHM}$ (8.3)<br>$I2_{SIG}$ (4.3) | 521 | 0.93 | 21% |
| Biomass | CHM-only | $Ht_{CHM}$ (2.9)<br>$TBV_{CHM}$ (2.9) | 1 | 0.87 | 30% |
| | | $Ht_{CHM}$ (2.9)<br>$TBV_{CHM}$ (2.9) | 1 | 0.87 | 30% |
| | CHM+DR | $TBV_{CHM}$ (2.9)<br>$H25_{DR}$ (96.8)<br>$Hmean_{DR}$ (109.4) | 1 | 0.88 | 32% |
| | | $Ht_{CHM}$ (2.9)<br>$TBV_{CHM}$ (2.9) | 4 | 0.87 | 30% |
| | CHM+FW | $Ht_{CHM}$ (4.7)<br>$Acrown_{CHM}$ (27.1)<br>$TBV_{CHM}$ (31.7)<br>$I2_{SIG}$ (6.0) | 1 | 0.91 | 26% |
| | | $Ht_{CHM}$ (3.3)<br>$TBV_{CHM}$ (8.3)<br>$I2_{SIG}$ (4.3) | 3 | 0.90 | 21% |

## V. DISCUSSION

### *A. Usefulness of full-waveform lidar data for volume and biomass assessment at the tree level*

The accuracy of our estimates using CHM-only or CHM+DR metrics are in accordance with previews studies, having generated a 15% RMSE for volume estimation [13] and to a 33% RMSE for biomass estimation [14].

In this study, we tested whether the use of additional metrics related to the vertical structure of trees could improve volume and biomass estimates performed using CHM-only metrics. The tree bounding volume ($TBV_{CHM}$) was selected in all developed models, and therefore seems to be even more useful for predictionss of volume and biomass, which tends to confirm the findings of Chen et al. [15]. Additional metrics derived from discrete-return data did not improve the results. Additional full-waveform metrics did not have a positive influence on volume estimates, but they significantly improved the accuracy of biomass estimations. We can consider that stem volume is well described by crown dimension (related to DBH, see section IV.B), tree height metrics ($Ht_{CHM}$ or $H75_{DR}$) or both ($TBV_{CHM}$). But the biomass estimates benefit from using $I2_{SIG}$, which is related to the vertical structure of a tree and considerbranch and leaf biomass more effectively.

However, $I2_{SIG}$ was the only selected full-waveform metric. The crown height ($Hcrown_{PROF}$) was probably not considered as a robust estimator due to excessive variability among trees. An error in porosity ratios ($R_{SIG}$ and $R_{PROF}$) may have occurred due to the presence or absence of undergrowth, which modified the ground signal amplitude. Profile integrals ($I_{PROF}$ and $I2_{PROF}$) were not selected. This was probably due to the fact that they were only related to the vertical distribution of the vegetation and did not account for changes in crown dimensions (projected area and crown height). Signal integrals ($I2_{SIG}$), which gave information on both the vertical and horizontal (amplitude) distribution of the vegetation density proved to be a more valuable parameter for biomass estimation.

*B. Limits of the method regarding stand characteristics*

For this study, an automatic tree-to-tree comparison between reference and estimated data was performed. Consequently, we needed a perfect match between automatically segmented crowns and field-measured tree positions. For that reason, we only retained plots on which the algorithm presented in Véga and Durrieu [26] offered a 100% crown identification success rate, in plots of up to 500 stems/ha (the reference plots used in this paper had an average of 140 trees/hectare). Despite satisfactory average detection in higher density plots, crown omission in those plots increased due to the non-detection of dominated trees. It would have been even more difficult to identify understory trees in multilayered forests, even if some studies were carried out with this purpose in mind using lidar point distributions when very dense point clouds are available (e.g., Jaskierniak et al. [36]).

In complex or dense stands, where the identification of individual tree crowns is not possible, the estimation of biomass directly at the plot level could provide analternative solution. In this case, biomass and volume estimations can be performed based on the vertical distribution of lidar discrete returns (see [11]), or from full-waveform signals summed at the plot level for simulating a large-footprint signal (which has been successfully used to predict biomass; see [37]). However, the impact of topography on the signal profile is an issue that must be addressedbefore deriving metrics from vegetation profiles to assess volume and biomass. To that aim, an analysis of a DTM derived from last echoes to assess the local slope could be considered prior to summing full-waveform signals.

## VI. Conclusion

Our study confirmed the usefulness of tree metrics derived from a CHM for stem volume and aboveground biomass predictions based onlidar data sensed from above the forest. After building estimation models from CHM-only metrics, CHM combined with discrete return metrics (CHM+DR) and CHM combined with full-waveform metrics (CHM+FW), we concluded that the tree bounding volume, obtained by multiplying the crown area by tree height, is one of the most useful metrics for volume and biomass estimations. We also observed that the use of additional FW metrics did not seem to improve volume estimation but led to a significant improvement of biomass estimation compared with models developed using CHM-only or CHM+DR metrics. The use of FW metrics probably allowed for a better accounting of the part of the aboveground biomass made up of branches and leaves.

### Acknowledgments

This work is part of the ExFOLIO project and was realized thanks to the financial support of the CNES (Centre National d'Études Spatiales) and the French Languedoc-Roussillon region. We would also like to thank GIS Draix for providing the full-waveform lidar data and for helping with the ground truth surveys. We are grateful to INSU for its support to GIS Draix through the ORE program, and we address our special thanks to Laurent Albrech, from UMR TETIS-Cemagref, for his help with the collection of field data.

### Bibliography

[1] T. Urquiza-Haas, P. M. Dolman and C. A. Peres, "Regional scale variation in forest structure and biomass in the Yucatan Peninsula, Mexico: Effects of forest disturbance", *Forest Ecol. Manag.,* vol. 247, no. 1-3, pp. 80-90, 2007.

[2] M. A. Wulder, C. W. Bater, N. C. Coops, T. Hilker and J. C. White, "The role of LiDAR in sustainable forest management", *Forest. Chron.,* vol. 84, no. 6, pp. 807-826, 2008.

[3] M. van Leeuwen and M. Nieuwenhuis, "Retrieval of forest structural parameters using LiDAR remote sensing", *Eur. J. Forest Res.,* vol. 129, pp. 749-770, 2010.

[4] M. L. Imhoff, "Radar backscatter and biomass saturation: ramifications for global biomass inventory", *IEEE Trans. Geosci. Remote Sensing,* vol. 33, no. 2, pp. 511-518, 1995.

[5] L. I. Duncanson, K. O. Niemann and M. A. Wulder, "Integration of GLAS and Landsat TM data for aboveground biomass estimation", *Can. J. Remote Sens.,* vol. 36, no. 2, pp. 129-141, 2010.

[6] J. E. Means, S. A. Acker, D. J. Harding, J. B. Blair, M. A. Lefsky, W. B. Cohen, M. E. Harmon and W. A. McKee, "Use of large-footprint scanning airborne Lidar to estimate forest stand characteristics in the western cascades of Oregon", *Remote Sens. Environ.,* vol. 67, no. 3, pp. 298-308, 1999.

[7] Q. Chen, "Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry", *Remote Sens. Environ.,* vol. 114, no. 7, pp. 1610-1627, 2010.

[8] M. Maltamo, J. Hyyppä and J. Malinen, "A comparative study of the use of laser scanner data and field measurements in the prediction of crown height in boreal forests", *Scand. J. Forest Res.,* vol. 21, no. 3, pp. 231-238, 2006.

[9] S. Magnussen, E. Næsset and T. Gobakken, "Reliability of LiDAR derived predictors of forest inventory attributes: A case study with Norway spruce", *Remote Sens. Environ.,* vol. 114, no. 4, pp. 700-712, 2010.

[10] J. E. Means, S. A. Acker, B. J. Fitt, M. Renslow, L. Emerson and C. J. Hendrix, "Predicting forest stand characteristics with airborne scanning lidar", *Photogramm. Eng. Rem. S.,* vol. 66, no. 11, pp. 1367-1371, 2000.

[11] Y. Li, H. Andersen and R. McGaughey, "A Comparison of statistical methods for estimating forest biomass from light detection and ranging data", *Western J. App. Forest.,* vol. 23, no. 4, pp. 223-231, 2008.

[12] Z. J. Bortolot and R. H. Wynne, "Estimating forest biomass using small footprint LiDAR data: An individual tree-based approach that incorporates training data", *ISPRS J. Photogramm.*, vol. 59, no. 6, pp. 342-360, 2005.

[13] S. Tonolli, M. Dalponte, L. Vescovo, M. Rodeghiero, L. Bruzzone and D. Gianelle, "Mapping and modeling forest tree volume using forest inventory and airborne laser scanning", *Eur. J. Forest Res.,* vol. 130, no. 4, pp. 569-577, 2010.

[14] S. C. Popescu, "Estimating biomass of individual pine trees using airborne lidar", *Biomass Bioenerg.,* vol. 31, no. 9, pp. 646-655, 2007.

[15] Q. Chen, P. Gong, D. Baldocchi and Y. Q. Tian, “Estimating basal area and stem volume for individual trees from lidar data”, *Photogramm. Eng. Rem. S.,* vol. 73, no. 12, pp. 1355-1365, 2007.

[16] C. Mallet and F. Bretar, “Full-waveform topographic lidar: State-of-the-art”, *ISPRS J. Photogramm.,* vol. 64, no. 1, pp. 1-16, 2009.

[17] A. Chauve, C. Vega, S. Durrieu, F. Bretar, T. Allouis, M. Pierrot-Deseilligny and W. Puech, “Advanced fullwaveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest”, *Int. J. Remote Sensing,* vol. 30:19, pp. 5211–5228, 2009.

[18] J. Heinzel and B. Koch, “Exploring full-waveform LiDAR parameters for tree species classification”, *Int. J. App. Earth Observation and Geoinformation,* vol. 13, no. 1, pp. 152 - 160, 2011.

[19] J. D. Muss, D. J. Mladenoff and P. A. Townsend, “A pseudo-waveform technique to assess forest structure using discrete lidar data”, *Remote Sens. Environ.,* vol. 115, no. 3, pp. 824 - 835, 2011.

[20] F. Rey, “Influence of vegetation distribution on sediment yield in forested marly gullies”, *Catena,* vol. 50, no. 2-4, pp. 549-562, 2003.

[21] D. Kwak, W. Lee, H. Cho, S. Lee, Y. Son, M. Kafatos and S. Kim, “Estimating stem volume and biomass of Pinus koraiensis using LiDAR data”, *J. Plant Res.,* vol. 123, no. 4, pp. 421-432, 2010.

[22] J. Toth J. and M. Turrel, “La productivité du Pin noir d'Autriche dans le Sud-Est de la France”. *Revue forestière française*, vol. 2, pp. 111-121, 1983.

[23] G. Montero. Cuantificacion de la biomasa forestal aerea y radical de distintas especies arboreas. “*Montes y energıas renovables. Ponencias y Comunicaciones Santiago de Compostela*”, pp. 115-131, 2004.

[24] W. Wagner, A. Ullrich, V. Ducic, T. Melzer and N. Studnicka, “Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner”, *ISPRS J. Photogramm.,* vol. 60, no. 2, pp. 100-112, 2006.

[25] C. Véga, S. Durrieu and T. Allouis, “A Lidar filtering algorithm for terrain modeling in complex forested environments*”, Comput. Geosci.,* accepted, 2011.

[26] C. Véga and S. Durrieu, “Multi-level filtering segmentation to measure individual tree parameters based on Lidar data: Application to a mountainous forest with heterogeneous stands”, *Int. J. Applied Earth Observation and Geoinformation,* vol. 13, no. 4, pp. 646 - 656, 2011.

[27] G. B. West, B. J. Enquist and J. H. Brown, “A general quantitative theory of forest structure and dynamics”, *Proceedings of the National Academy of Sciences of the United States of America,* vol. 106, no. 17, pp. 7040-7045, 2009.

[28] H. C. Muller-Landau, R. S. Condit, J. Chave, S. C. Thomas, S. A. Bohlman, S. Bunyavejchewin, S. Davies, R. Foster, S. Gunatilleke, N. Gunatilleke, K. E. Harms, T. Hart, S. P. Hubbell, A. Itoh, A. R. Kassim, J. V. LaFrankie, H. S. Lee, E. Losos, J. .. Makana, T. Ohkubo, R. Sukumar, I. .. Sun, M. N. Nur Supardi, S. Tan, J. Thompson, R. Valencia, G. V. Muñoz, C. Wills, T. Yamakura, G. Chuyong, H. S. Dattaraja, S. Esufali, P. Hall, C. Hernandez, D. Kenfack, S. Kiratiprayoon, H. S. Suresh, D. Thomas, M. I. Vallejo and P. Ashton, “Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests”, *Ecol. Lett.,* vol. 9, no. 5, pp. 575-588, 2006.

[29] R. H. MacArthur and H. S. Horn, “Foliage profile by vertical measurements”, *Ecology,* vol. 50, no. 5, pp. 802-804, 1969.

[30] M. A. Lefsky, D. Harding, W. B. Cohen, G. Parker and H. H. Shugart, “Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA”, *Remote Sens. Environ.,* vol. 67, no. 1, pp. 83-98, 1999.

[31] T. Allouis, S. Durrieu, J.-S. Bailly, P. Chazette, J. Cuesta, C. Véga, P. Flamant and P. Couteron, “Potential of an ultraviolet, medium-footprint lidar prototype for retrieving tree heights and tree planting patterns”, *ISPRS J. Photogramm.,* In press, 2011.

[32] Neter, Wasserman and Kutner, 1985. Applied Linear Statistical Models, 2nd ed, Irwin

[33] A. McLeod and C. Xu, Package ‘bestglm’, version 0.31. http://cran.r-project.org/web/packages/bestglm/, 2010.

[34] R. M. O'Brien, “A caution regarding rules of thumb for variance inflation factors”, *Qual. Quant.,* vol. 41, no. 5, pp. 673-690, 2007.

[35] R. R. Picard and R. D. Cook, “Cross-validation of regression models”, *Journal of the American Statistical Association,* vol. 79, no. 387, pp. 575-583, 1984.

[36] D. Jaskierniak, P. N. Lane, A. Robinson and A. Lucieer, “Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions”, *Remote Sens. Environ.,* vol. 115, no. 2, pp. 573 - 585, 2011.

[37] M. A. Lefsky, W. B. Cohen, D. J. Harding, G. G. Parker, S. A. Acker and S. T. Gower, “Lidar remote sensing of above-ground biomass in three biomes”, *Global Ecol. Biogeogr.,* vol. 11, no. 5, pp. 393-399, 2002.

# Chapitre 2

# Expérimentation d'un capteur lidar à moyennes empreintes

La seconde partie de ce document présente une expérimentation d'un capteur lidar profileur à moyennes empreintes ($\simeq$ 2,4 m), enregistrant le signal complet. Un tel système, dont l'encombrement permet son embarquement à bord d'un ULM, facilite et diminue les coûts de déploiement par rapport aux traditionnels systèmes scanners. L'utilisation d'une taille empreinte intermédiaire fut envisagée pour faciliter la prise en compte de la rétrodiffusion du signal par l'ensemble de la colonne de végétation, tout en fournissant des données de meilleure résolution que les celles systèmes à larges empreintes (10 – 70 m). Une validation de ce prototype était cependant nécessaire afin de situer ses performances par rapport à celles des autres systèmes utilisés en forêt.

Le premier objectif de cette étude était donc d'évaluer la précision du prototype pour la mesure de paramètres forestiers à l'échelle de la parcelle : hauteurs moyennes d'arbres et de houppiers, hauteur dominante, motif de plantation, espacement des arbres. Le second objectif visait à connaitre le nombre de tirs minimum nécessaires pour mesurer correctement la parcelle, afin réfléchir à la stratégie d'échantillonnage à adopter à l'échelle d'un massif forestier.
La validation de l'utilisation d'un laser ultraviolet pour l'étude de la forêt réprésentait aussi un enjeu, puisqu'une telle longueur d'onde n'avait jusqu'alors pas été testée à partir d'un vecteur volant. Cette longueur d'onde étant utilisée pour l'étude de paramètres atmosphériques (*Chazette et al., 2007*), la mesure conjointe de paramètres forestiers permet d'envisager le développement de capteurs lidar bi-fonctions, particulièrement utiles pour l'étude des échanges entre l'atmosphère et la végétation.

Les données sont différentes de celles fournies par les scanners ou les systèmes satellitaires à plus larges empreintes. Des développements méthodologiques ont donc été nécessaires pour extraire les paramètres forestiers à différentes échelles : de la placette à la parcelle. À l'instar de la premi partie de la thèse, les méthodes de traitement développées ont été inspirées par celles utilisées pour traiter les signaux à larges empreintes (*Means et al., 1999; Lefsky et al., 2005*).

# ASSESSMENT OF TREE AND CROWN HEIGHTS OF A MARITIME PINE FOREST AT PLOT LEVEL USING A FULLWAVEFORM ULTRAVIOLET LIDAR PROTOTYPE

*Tristan Allouis* [a,*], Student Member, IEEE, *Sylvie Durrieu* [a], *Juan Cuesta* [b], *Patrick Chazette* [b], *Pierre H. Flamant* [b], *Pierre Couteron* [c]

[a] Cemagref / AgroParisTech, UMR TETIS, Montpellier, France
[b] Institut Pierre Simon Laplace, LMD / LSCE, Palaiseau / Gif sur Yvette, France
[c] Institut de Recherche pour le Développement, UMR AMAP, Montpellier, France
* Corresponding author. E-mail address: tristan.allouis@teledetection.fr

**ABSTRACT**

This study aims to determine the potential of a new lidar prototype with an ultraviolet laser and a medium footprint for retrieving forest parameters. The lidar is embedded on an ultra-light aircraft. The choice of the Landes forest in southwestern France as study area was made regarding the flat topography of the area and the stand height consistency. We chose three plots from different stands (different height characteristics) and compared the lidar derived metrics to field measurements. To derive metrics from lidar data, we summed the lidar waveforms within a plot and calculated derived reflectance profiles to correct the lidar signal from the occlusion effect. We then retrieve plot mean total and mean crown base heights measurement from reflectance profiles. We obtain a good consistency of the lidar measurements compared to field measurements, even if we noticed the existence of a 5 to 10% bias probably linked to the lidar sampling strategy.

***Index Terms***— lidar, forestry, signal analysis, sampling methods, aerosols.

## 1. INTRODUCTION

Lidar (light detection and ranging) is an active technique based on the emission and reception of laser pulses. It enables surface relief measurement, as passive techniques do, but also provides additional information on the vegetation structure inside and below the canopy. It is commonly used in the forestry research field to characterize tree or stand heights [1], total volume and spatial organization of vegetation material [2]. These data can be used as inventory data for forestry management and as model input data for ecological applications. The lidar system parameters (laser beam divergence, laser wavelength, measurement repetition rate, waveform digitization frequency) and flight parameters (altitude, speed) have an impact on the recorded backscattered signals and on the way the forest is sampled (laser footprint size, distance between footprints). These parameters must therefore be chosen related to the scale of observation, the needed precision and the amount of data to be managed.

The aim of this study is to determine the potential of a new lidar prototype with an ultraviolet (UV) laser and a medium footprint (2.4 m diameter) [3] to provide forest structure information at plot level.

## 2. MATERIAL

### 2.1. Lidar system

The lidar prototype is a profiler based on the LAUVA (Airborne UltraViolet Aerosol Lidar) initially developed by the Commissariat à l'Énergie Atomique (French Atomic Energy Commission) and the Centre National de la Recherche Scientifique (National Center for Scientific Research) for atmospheric applications [4]. The system was modified for canopy measurements. The laser beam footprint has a 2.4 m diameter for a 300 m flight altitude. Such footprint size is in between conventional small footprint topographic lidars and large footprint experimental systems, providing respectively 15 cm to 1 m and 10 m to 70 m footprint diameters [5] [6]. The UV wavelength was chosen for its ability to characterize the atmosphere, while providing adequate eye safety. The lidar emits high energy pulses (17 mJ versus < 0.2 mJ for commercial near infrared topographic lidars) to obviate the

difference in vegetation response regarding the wavelength. Actually, reflectance and transmittance of vegetation in the UV wavelength are lower than in the near infrared wavelength traditionally used in earth surface observation lidars. The fullwaveform backscattered signal is recorded with a 100 MHz sampling rate witch provides a vertical measurement precision of 1.5 m.

The lidar was embedded on an ultra-light aircraft (ULA) for rapid deployment and flexibility of flight plans. The ULA flies at about 30 m/s, and the lidar operates at 20 Hz during one second every two seconds. Considering that the laser footprint diameter is 2.4 m, we get sequences of continuous measurements along 30 m profiles every 30 m.

### 2.2. Study site

The experiment took place on the Landes forest (Landes of Gascony) in southwestern France (44° 10' N, 1° 12' W). It is the largest Maritime pine (*Pinus pinaster*) forest in Europe. With a planted area of about a million hectares it is a major economic pole in France with 30,000 jobs related to forestry and wood transformation.

The lidar flew on three different stands of about 1.2 $km^2$ each in the Mimizan commune: a young plantation, a middle-aged plantation and a mature naturally regenerated stand. For each stand, reference field data were acquired in measuring both the tree total height and the crown base height of each tree within a square plot of 30x30 m (Table 1).

## 3. METHOD

Lidar measurements were geolocated using both the position and the orientation of the system, measured by a differential GPS (DGPS) and an electronic flight instrument system (EFIS) respectively. The EFIS provides angles measurements with an accuracy of 0.5° and one EFIS measurement and five DGPS measurements are available for one 20 lidar shots sequence. Consequently, the (X, Y, Z) geolocation accuracy of a lidar spot is estimated to be about 5 m.

This low geolocating accuracy is not compatible with a study at tree level. Therefore, we chose to compare lidar and field measurements at plot level by summing the 20 successive backscattered signals, for a sequence of footprints located in each plot. This level of study also corresponds to traditional forest inventory field measurements level. In addition, the probability to obtain a ground echo within the waveform sum is increased. Each waveform sum was then transformed into reflectance profile according to an adaptation of the MacArthur and Horn equation [7]. This adapted equation transforms the backscattered lidar signal into the amount of cover per height class, while taking account of lidar signal attenuation inside the vegetation:

$$D_{r1,r2} = \ln\{P_{r1}/P_{r2}\}$$

Where $D_{r1,r2}$ is the intensity of reflectance between $range_1$ and $range_2$ from the plane with $range_1 < range_2$, $P_{r1}$ is the cumulative power between $range_1$ and $range_{max}$, $P_{r2}$ is the cumulative power between $range_2$ and $range_{max}$. The intensity of reflectance is calculated for each interval between two sucessive ranges from $range_{min}$ to $range_{max}$ (figure 1). As far as we did not take into account the difference of reflectance between ground and vegetation, we called our profile reflectance profile instead of vegetation profile.

Finally, the location of the last reflectance profile peak is considered to correspond with the ground range ($R_{ground}$). The first reflectance profile point that exceeds the noise level is considered to correspond with the top of canopy range ($R_{top}$). The point situated before the strongest decrease of reflectance profile and between the last canopy increase and the ground peak is considered to correspond with the crown base range ($R_{crown}$) (figure 1). The distance between $R_{ground}$ and $R_{top}$ is assumed to be the plot mean total height. The distance between $R_{ground}$ and $R_{crown}$ is assumed to be the plot crown base mean height. These two heights are then compared to the reference data. To that aim, we calculated the mean total height and the mean crown base height for each plot.

## 4. RESULTS

Figure 2 shows lidar signals summed at plot level, derived reflectance profiles and plots average trees for the three different plots. We can see a good agreement between the average tree, the lidar signal variations and reflectance profile distribution.

Except for the crown base height of plot c, all the lidar derived heights are consistent with those calculated from field measurements (Table 2) considering that the lidar derived heights are given with a precision of 1.5 m (the waveform digitization frequency) and considering the standard deviations of field tree heights.

## 5. DISCUSSION AND CONCLUSION

We chose to sum waveforms of contiguous laser spots in order to produce 30 m long and 2.4 m wide footprints and perform a study at plot level. However, the reference plots are 30x30 m square. Consequently, the information held in a summed lidar waveform characterizes only a part of the plot. The increase in measurement errors, from young stand plot to mature stand plot, is probably linked to the lidar sampling strategy with regard to stand structure variability.

Actually, trees in the young stand plot are plentiful (127 trees), planted in regular lines. On the contrary, the naturally regenerated stand plot held fewer trees (15 trees) with random position. A larger number of trees are then measured by the lidar in the young stand plot than in the mature stand plot. Consequently, measurements are more robust on the young stand plot. These results highlight the importance of the lidar sampling strategy in order to retrieve consistent heights measurements.

In depth analysis of these data will be performed in the future. To study the impact of the sampling strategy, the reflectance profiles derived from spot sizes of 30x2.4 m (presented in this study) and 30x4 m (obtained by flying the lidar at an altitude of 500 m) will be compared. We will also seek to extract the tree cover from the plot reflectance profiles. In addition, we will consider other statistics than the mean to compare field measurements with lidar derived metrics. Actually, the mean was suitable for this study because of tree height distribution homogeneity, but will not be suitable for multi-layered stands.

In spite of the previously discussed sampling bias, the results demonstrate the potential of this lidar prototype for tree heights estimation. This is promising for emergency inventory applications that can be facilitated by the rapid deployment and flexibility of ultra-light aircraft (ULA) in the case of storm or fire events. The obtained results will also be useful for designing systems for a future spaceborne lidar mission dedicated to forest measurements [8].

For future experiments, the EFIS accuracy will be improved to enable direct comparison between recorded waveform and field data (stand characteristics and topography). As the initial function of the LAUVA system was to measure atmospheric aerosols [9], the conception of a bi-function lidar for studying the forest response to atmospheric pollution is possible. The addition of a channel to study chlorophyll fluorescence is also considered for tree species recognition.

## 6. REFERENCES

[1] A. Chauve, C. Vega, S. Durrieu, F. Bretar, T. Allouis, M. Pierrot-Deseilligny, W. Puech, "Advanced fullwaveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest", *International Journal of Remote Sensing,* vol. 30:19, pp. 5211-5228, 2009.

[2] M. A. Lefsky, W. B. Cohen, S. A. Acker, G. G. Parker, T. A. Spies, D. Harding, "Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests", *Remote Sensing of Environment*, vol. 70, pp. 339-361, 1999.

[3] J. Cuesta, P. Chazette, J. Sanak, T. Allouis, S. Durrieu, P. Genau, C. Flamant, P. H. Flamant, "New airborne lidar observes forest canopies", *SPIE Newsroom*, 28 September 2009.

[4] P. Chazette, J. Sanak, F. Dulac, "New approach for aerosol profiling with a lidar onboard an ultralight aircraft: application to the African monsoon", *Environmental Science & Technology*, vol. 41, pp. 8335-8341, 2007.

[5] C. Mallet, F. Bretar, "Full-waveform topographic lidar: State-of-the-art", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, pp. 1-16, 2009.

[6] D. J. Harding, C. C. Carabajal, "ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure", *Geophysical Research Letters*, vol. 32, pp. 1-4, 2005.

[7] R. H. MacArthur, H. S. Horn, "Foliage Profile by Vertical Measurements", *Ecology*, vol. 50, pp. 802-804, 1969.

[8] P. H. Flamant, "Atmospheric and meteorological Lidar: from pioneers to space applications", *Comptes Rendus Physique*, vol. 6:8, pp. 864-875, 2005.

[9] J.-C. Raut, P. Chazette, "Assessment of vertically-resolved PM10 from mobile lidar observations", *Atmospheric Chemistry and Physics*, vol. 9:21, pp. 8617-8638, 2009.

| | Mean total height ± standard deviation (m) | Mean crown base height ± standard deviation (m) | Number of trees |
|---|---|---|---|
| Plot a (young) | 9.5 ± 0.9 | 4.5 ± 0.8 | 127 |
| Plot b (intermediate) | 15.9 ± 1.1 | 10.3 ± 0.6 | 56 |
| Plot c (mature) | 21.9 ± 0.9 | 15.6 ± 1.1 | 15 |

**Table 1**. Tree plots characteristics.

| | Lidar total plot height (m) | Total height error (lidar – field) | Lidar crown base plot height (m) | Crown base height error (lidar – field) |
|---|---|---|---|---|
| Plot a (young) | 10.5 | 1 | 4.5 | 0 |
| Plot b (intermediate) | 16.5 | 0.6 | 9 | -1.3 |
| Plot c (mature) | 21 | -0.9 | 13.5 | -2.5 |

**Table 2**. Tree heights extracted from reflectance profiles and their comparison with mean heights calculated from field measurements.

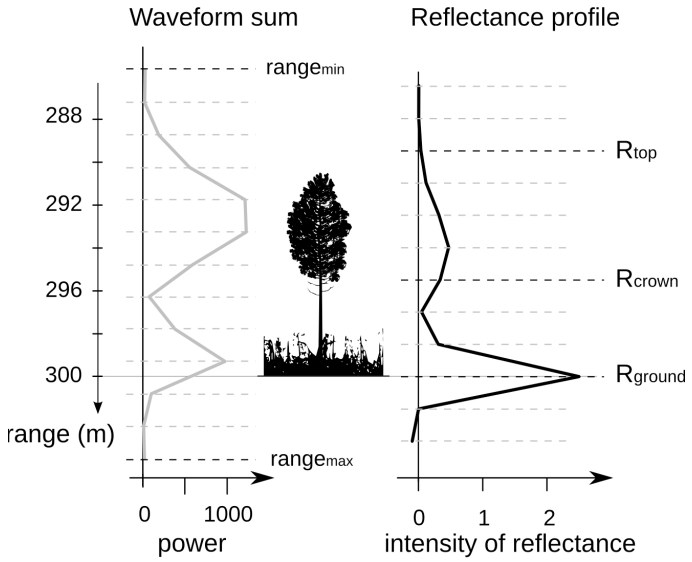



**Figure 1**. A tree, the lidar derived reflectance profile, and the positions of $range_{min}$, $range_{max}$, $R_{top}$, $R_{crown}$ and $R_{ground}$.

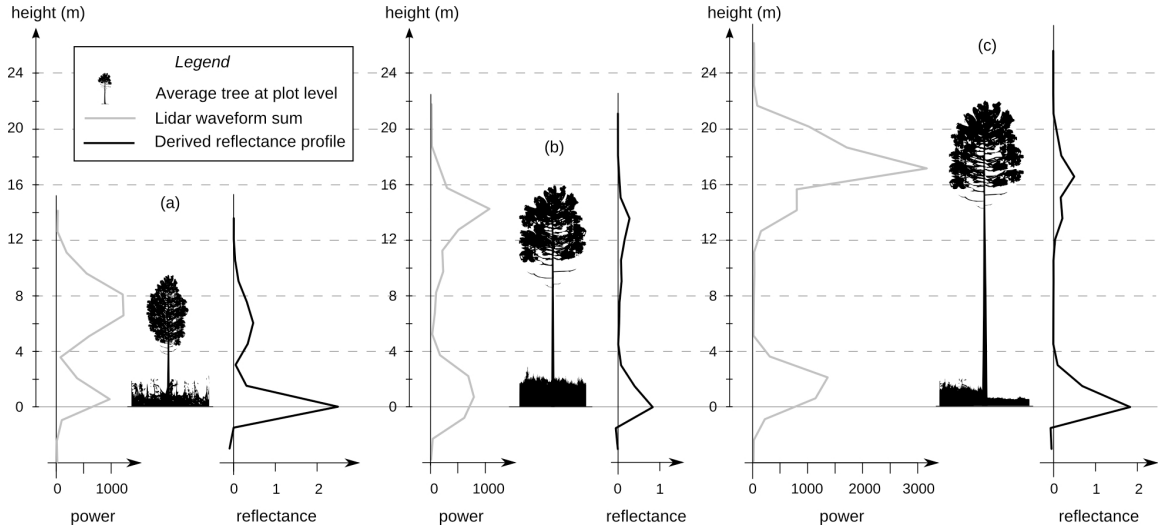



**Figure 2**. Lidar waveform sums (grey curves), derived reflectance profiles (black curves) and average trees at plot level derived from field measurements (black), for the three plots. The average trees are related to total mean height, base crown mean height, mean crown diameter and undergrowth distribution.

Contents lists available at SciVerse ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

# Potential of an ultraviolet, medium-footprint lidar prototype for retrieving forest structure

Tristan Allouis [a,b,*], Sylvie Durrieu [a], Patrick Chazette [c], Jean-Stéphane Bailly [b], Juan Cuesta [d], Cédric Véga [a], Pierre Flamant [d], Pierre Couteron [e]

[a] Cemagref, UMR TETIS, 500 rue Jean-François Breton, 34093 Montpellier Cedex 5, France
[b] AgroParisTech, UMR TETIS, 500 rue Jean-François Breton, 34093 Montpellier Cedex 5, France
[c] IPSL, Laboratoire des Sciences du Climat et l'Environnement, Orme des Merisiers, 91191 Gif-sur-Yvette Cedex, France
[d] École Polytechnique, Laboratoire de Météorologie Dynamique, 91128 Palaiseau Cedex, France
[e] Institut de Recherche pour le Développement, UMR AMAP, Bd de la Lironde TA A51/PS2, 34398 Montpellier Cedex 5, France

ARTICLE INFO

*Article history:*
Available online xxxx

*Keywords:*
Ultra light aircraft
Signal processing
Semi-variogram
Forest structure
Laser attenuation

ABSTRACT

The aim of the paper is to carry on methodological development for retrieving forest parameters from medium-footprint lidar signals and for assessing the performance of different sampling strategies.

The 2.4 m footprint lidar prototype (a profiler instrument using an ultraviolet laser) was flown above two different maritime pine stands: a young plantation (10 years old) and a mature, semi-natural stand (55 years old), both in the Landes forest, France.

The vertical distribution of lidar measurements was studied for retrieving forest height parameters (mean total height, mean crown height and top height). The processing algorithm was based on an aggregation of successive signals followed by the correction of the signal attenuation along the travel through the vegetation. The performance of different sampling strategies was assessed by comparing the results for the full dataset (several fight lines over the stands) and for only a data subset (one flight line).

In addition, the horizontal distribution of height measurements was studied for identifying the planting pattern of the stands and assessing the tree spacing of the semi-natural parcel, using geostatistics.

We obtained a sub-metric estimation error (lidar – reference) of 0.2 m on the mean total height in the young stand (−0.7 m in the mature stand), a bias of −0.3 m (−0.3 m) on the mean crown height measurement and of 0.6 m (−1.0 m) on the top height. The planting pattern was also successfully identified, and the distance between trees was assessed in agreement with ground measurements.

Having demonstrated its ability to assess forest structure, even with a unique flight line, the lidar prototype seems to be a valuable sensor for performing fast forest inventory at regional scale. In addition, this sensor opens the way to the development of bi-functional lidar for both atmosphere and vegetation remote sensing.



## 1. Introduction

Achieving a precise inventory of the world's forests has become a priority ever since the role that forests could play in climate regulation has been brought to light. Improved estimations of forest aboveground biomass and $CO_2$ emissions from deforestation are needed to reduce uncertainties on the terrestrial carbon pool (ESA, 2008). Currently, the Forest Resource Assessment (FRA) of the United Nations Food and Agricultural Organization (FAO) is the main source of information on forest resources at the international level. However, FRA is not spatially explicit. Data are compiled at the national level and are reported through a single value of forest growing stocks, while spatially distributed estimations of growing stocks are required by vegetation models. To fill this gap, remote sensing data supported by ground observations are considered the key to effective monitoring (DeFries et al., 2007). Indeed, remote sensing techniques can provide forest structure information on large areas through photogrammetry (e.g., Soenen et al., 2010), lidar (light detection and ranging) (e.g., Wynne, 2006), or both (e.g., Véga and St-Onge, 2008; Packalen et al., 2009).

Lidar remote sensing not only provides surface elevation measurements as photogrammetric techniques do, but it can also provide information about vegetation structure inside and below the canopy. Even though the ground detection is still challenging in tropical or dense forests (Drake et al., 2002; Clark et al., 2004), this advantage explains why many studies have focused on the use of

* Corresponding author at: Cemagref, UMR TETIS, 500 rue Jean-François Breton, 34093 Montpellier Cedex 5, France. Tel.: +33 4 67 54 87 19; fax: +33 4 67 54 87 00.
*E-mail address:* tristan.allouis@teledetection.fr (T. Allouis).

0924-2716/$ - see front matter © 2011 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2011.08.007

lidar data to characterize forest environments. Tree or stand heights were successfully retrieved and the total volume and spatial organization of vegetation material were determined (e.g., Lefsky et al., 1999; Chen et al., 2007; Popescu, 2007). Due to their accuracy, forest parameters derived from lidar data can be used for forest inventory and as model inputs for ecological applications (e.g., Waser et al., 2008; Silva-Santos et al., 2010). However, because lidar system parameters (i.e., laser beam divergence, laser wavelength, measurement repetition rate, waveform digitization frequency) along with the flight parameters (altitude, speed) have an impact on the recorded backscattered signals and the way forests are sampled (laser footprint size, distance between footprints) (Næsset, 2009), they must be chosen in relation to the scale of observation, the accuracy needed and the amount of data to be managed.

Two types of lidar systems are currently used for forest application: small- and large-footprint systems. The data from each system are processed using different techniques for retrieving forest parameters. Because of its commercial availability, small-footprint (15–30 cm), airborne lidar systems are commonly used (e.g., Mallet and Bretar, 2009). For these systems, the range measurements between the instrument and the targets intercepted by the laser beam are extracted from multi-pulse or full-waveform backscatter signals. The accurate three-dimensional ($X$, $Y$, $Z$) point cloud, which is derived from the range measurements, the scan angle and both plane attitude and position, is then classically processed to create a digital terrain model (DTM) by selecting ground points among the last echoes. Then, the forest parameters are retrieved by either analyzing a Canopy Height Model (CHM) obtained by subtracting the DTM from the first echoes (e.g., Means et al., 2000; Chauve et al., 2009) or analyzing the height distribution of the points inside the point cloud at tree or plot levels (Kato et al., 2009). In contrast, forest parameters can be directly computed from large-footprint lidar waveforms due to the presence of a ground return inside a 10–70 m spot. From such large-footprint airborne or satellite instruments, the stand height and total above-ground and foliage biomass can be estimated (e.g., Means et al., 1999; Lefsky et al., 2005). However, large footprint data do not allow operating at an individual tree level, and the accuracy of tree height estimation suffers from the strong influence of the terrain slope. Indeed, the slope produces a mixture of vegetation and terrain echoes (Chen, 2010).

The majority of lidar systems used for forest applications use infrared or green wavelengths. However, the UV wavelength is suitable to detect the presence and the concentration of atmospheric pollutants (aerosols) (Chazette et al., 2007). Validating the capacity for a UV lidar to measure 3D forest structure would then open the way to the development of bi-functional lidars able to study the interactions between vegetation and aerosols, which is particularly interesting at the interface between forest and urban areas. To that purpose, a lidar prototype was designed by the Commissariat à l'Énergie Atomique (CEA) and the Centre National de la Recherche Scientifique (CNRS) based on a UV lidar sensor initially developed for atmospheric applications (Chazette et al., 2007; Raut and Chazette, 2009). The first experiments of this prototype provided measurements with a 2.4 m footprint (due to the laser beam divergence and a flight altitude of 300 m) in order to obtain a backscattered signal integrating information from a larger area (from sub-tree to tree level) than is observed with small-footprint lidar systems. In many forest types, an increase in the footprint size would increase the probability of each laser pulse to intercept both tree tops and ground. But the drawback is a higher sensitivity of the measurement to the local topography that would limit the use of large-footprint lidar on uneven or sloped terrain. Consequently, a compromise must be reached and an intermediate footprint size could be a solution. However, new methods for processing such lidar data must be developed, taking into account that a medium footprint signal does not systematically contain a ground return, and can therefore not be processed exactly the same way as large footprint signals. On the other hand, medium footprint backscatter signals are too complex for being processed like small footprint waveforms. Indeed, the assumption made by Hofton et al. (2000) ("returning laser pulse is composed of a series of potentially-overlapping reflections similar in shape to the impulse response"), would require a too large amount of signal components.

The main goal of this study is to carry on the development of methods for retrieving forest parameters that are commonly used in forestry and to confirm the potential of our medium-footprint system to measure forest structure. Cuesta et al. (2010) presented the first results showing the potential of the data acquired by this medium-footprint system for retrieving forest structural information on canopy top, tree crown base and undergrowth heights, through the direct process of lidar waveforms.

In the current paper, an extensive analysis of both vertical and horizontal measurement distributions is performed to estimate usual forest parameters from a medium footprint lidar system. Forest heights (mean total height, mean crown height and top height) are estimated by correcting the lidar signal from its attenuation through the vegetation. Planting pattern and tree spacing are retrieved by analyzing the spatial distribution of height measurements using geostatistical methods. We also evaluate the performance of different sampling strategies for assessing forest parameters from a profiler system. This was done by comparing the results obtained using the full dataset or using data subsets.

## 2. Material

### 2.1. Study area

The experiment took place in September 2008 at the Landes forest (Landes de Gascogne) in the southwest of France (44°10′N, 1°12′W) (Fig. 1). With a planted area close to a million hectares, it is the largest maritime pine (*Pinus pinaster*) forest in Europe and is a major economic pillar in France, with 30,000 jobs related to forestry and wood transformation. This area was also chosen because of its flat topography. Both the simplicity of the stand structure and the absence of ground slope were expected to make the process of obtaining the lidar data and interpreting the results easier.

For this study, we selected two stands with distinct characteristics: a young plantation (10 years old, 8 ha) with a regular structure and high tree density along lines and a semi-natural (also called natural regeneration (Williston and Balmer, 1974)) mature stand (55 years old, 10 ha) with a more irregular spatial distribution of trees.

### 2.2. Ultraviolet, medium-footprint lidar prototype and dataset

The lidar prototype is a Nd:YAG profiling sensor based on the LAUVA (Airborne UltraViolet Aerosol Lidar), which was initially developed by the Commissariat à l'Énergie Atomique (CEA) and the Centre National de la Recherche Scientifique (CNRS) to monitor aerosol dispersion in the low and middle troposphere (Chazette et al., 2007).

The beam divergence of the original system was set to 4 mrad, providing a 2.4 m footprint at a 300 m flight altitude. This footprint diameter was between the small- and large-footprint diameters conventionally used. The UV wavelength (355 nm) was retained for its ability to characterize atmospheric aerosols (Raut and Chazette, 2009) while providing adequate eye safety, since the UV radiation is absorbed by the eye cornea and the crystalline before

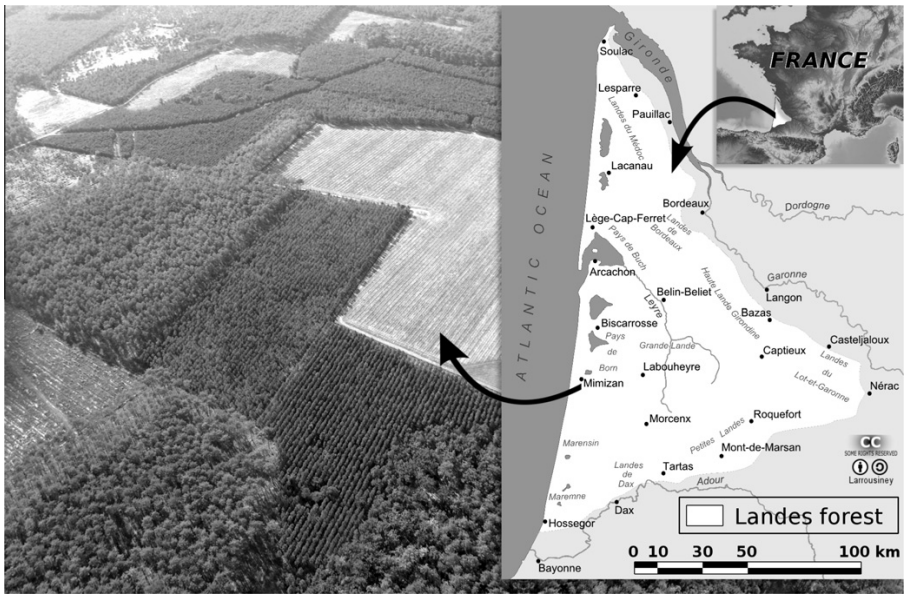



**Fig. 1.** Location of the study area in the Landes forest in southwestern France.

reaching the retina, even for high-energy pulses (according to the international standard IEC 60825-1; http://www.leosphere.com/). The laser emits higher energy pulses than commercial, near-infrared topographic lidars (16 mJ versus <0.2 mJ) to offset the difference in vegetation response regarding the wavelength. The reflectance and transmittance of vegetation are approximately 10 times lower (Stam, 2008) in the UV wavelength than in the near-infrared wavelength traditionally used in Earth surface observation lidars. The emitted pulse duration of 5 ns, jointly with the full-waveform receiver sampling rate of 100 MHz, lead to a vertical measurement precision close to 1.5 m. The lidar operates at 20 Hz 1 s for every 2 s, and the data are written during the remaining second (Cuesta et al., 2010).

The lidar was embedded on an ultra-light aircraft (ULA) to allow rapid deployment and flexibility of flight plans (see Chazette et al., 2010; Cuesta et al., 2010). Several flight lines were acquired on the study area in order to provide spatially explicit measurements on the area. The 30 m/s ULA flight speed, the lidar operating frequency and the footprint size yielded continuous measurements along 30 m profiles every 30 m (Fig. 2). Due to slight ULA flight speed fluctuations, the length of the profiles varied slightly. The ULA position and attitude were given by a differential GPS (DGPS) and an electronic flight information system (EFIS). The DGPS yielded five measurements per second (5 Hz) of the ($X$, $Y$, $Z$) plane position with a centimeter-level precision, while the EFIS provided one angle measurement for the first shoot of each profile of 20 successive shoots (0.5 Hz) with a 0.5° accuracy. Consequently, the theoretical absolute error of the lidar profile position was about 4.5 m, without taking into account the error in angles and position values due to the interpolation from the EFIS and DGPS measurements. However, if we estimate that the ULA movement was uniform during 1 s, successive shoots within a 20 shoot series were positioned relatively to each others with a higher accuracy (<1 m).

The lidar data falling inside the two stands selected for this study were then extracted. We obtained 94 groups of 20 shoots (1880 shoots) in the young stand and 83 groups of 20 shoots (1660 shoots) in the mature stand. A subset of these data was also extracted, in order to assess the ability of a unique flight line to provide tree measurements that are representative of a whole stand. This subset contains eight groups of 20 shoots (160 shoots) in each stand (Fig. 2).

### 2.3. Field reference data

Field data were collected in April 2009 on two 30 × 30 m plots in the young stand (planted in regular lines, plots y1, y2) and three 30 × 30 m plots for the mature stand (semi-natural, plots m1, m2, m3) (Fig. 2). The total height, crown base height, and crown diameter were measured for individual trees. The top heights and the distance between trees were also estimated. The field data are summarized in Table 1.

All stems were located (i.e., $X$, $Y$, $Z$ coordinates) for each plot, and their diameter at breast height (1.30 m DBH) was measured using a tape.

The total and crown base heights were measured for all trees on the semi-natural plots using a Vertex III ultrasonic clinometer (Haglöf, Sweden). For planted plots, we only measured the total and crown base heights for one out of three trees for height homogeneity and because of the large number of trees.

The crown diameters were measured on the ground using a tape after positioning the maximum extend of the tree crown in two directions using a densiometer. This was done for one out of three trees along the cardinal directions (north–south and east–west) in the semi-natural plots, and for one out of six trees perpendicularly and along tree lines in the planted plots.

The top height of a stand has various slightly different definitions in forestry (Nakai et al., 2010), and is often defined as the mean total height of the 100 largest (i.e., with the largest DBH) trees per hectare. According to the conventional method for top height estimation, the 9 largest trees in a 0.09 ha plot must be selected in the field (Garcia and Batho, 2005). Because the lidar cannot directly measure DBH, the reference top height of a stand is defined in this paper as the mean total height of the 9 tallest trees in each plot.

The reference tree spacing was simply computed as the mean of the nearest neighbor distance of each tree in the plot. Because the

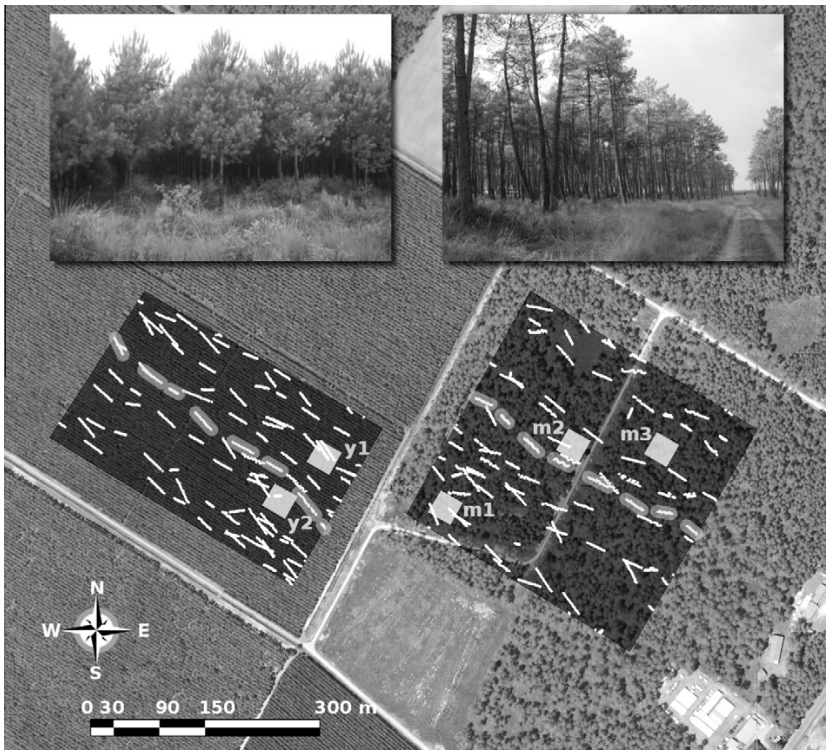



**Fig. 2.** The study area. The white dots represent the lidar shoots falling inside the young stand (west) and the mature stand (east). The enlighten dots (gray) represent lidar shoots of a unique flight line. The light-gray squares represent the locations of the terrain plots (young plots: y1 and y2, mature plots: m1, m2 and m3).

**Table 1**
Plot characteristics. Direction A is along plantation lines for young plots and north for mature plots. Direction B is between lines for young plots and east for mature plots.

| | Plot y1 (young) | Plot y2 (young) | Plot m1 (mature) | Plot m2 (mature) | Plot m3 (mature) |
|---|---|---|---|---|---|
| Number of trees | 127<br>229 | 102 | 25<br>57 | 15 | 17 |
| Mean total height ± standard deviation (m) | 9.5 ± 0.9<br>9.3 ± 0.9 | 9.1 ± 0.8 | 20 ± 1.6<br>21.6 ± 2 | 21.9 ± 0.9 | 23 ± 2.3 |
| Mean crown base height ± standard deviation (m) | 4.5 ± 0.8<br>4.9 ± 0.8 | 5.3 ± 0.5 | 14 ± 1.5<br>15.3 ± 1.6 | 15.6 ± 1.1 | 16.3 ± 1.3 |
| Crown dimensions in direction A (m) | 2.7<br>2.8 | 2.8 | 6.6<br>7.0 | 5.5 | 8.6 |
| Crown dimensions in direction B (m) | 3.4<br>3.3 | 3.1 | 6.0<br>6.1 | 5.1 | 7.2 |
| Top height ± standard deviation (m) | 10.1 ± 0.6<br>9.9 ± 0.5 | 9.6 ± 0.4 | 22.3 ± 1.1<br>23.2 ± 1.5 | 22.5 ± 0.6 | 24.8 ± 1.5 |
| Tree spacing (m) | NA | | 5.1 | | |

plantations had different tree spacing along the lines and between the lines, the tree spacing trees was not used in this study for plantations.

## 3. Method

The backscattered signal of our system was not composed of a succession of narrow and separated echoes, nor did it systematically contain both tree and ground returns in the same waveform, which is essential for the direct extraction of tree heights from the individual waveform. Consequently, we developed specific methods to process the medium-footprint lidar waveforms based on a three-phase approach. The first phase (described in Section 3.1) consisted of aggregating the 20 successive waveforms corresponding to a pulse emission sequence of 1 s. A logarithmic function was then used to correct the waveform sum from the laser attenuation inside the canopy and to retrieve ground position, mean total and mean crown height on a 30 m-long transect (Fig. 3). This methodology reduces ambiguity in the detection of ground and tree structural parameters due to the foliage occlusion effect, which was noted in Cuesta et al. (2010). In the second phase (described in Section 3.2), individual waveforms were corrected and individual total heights were retrieved even for waveforms without a ground return, by considering the ground position estimated in phase one. This second phase detects tree heights at a finer scale to compute stand top heights. In the third phase (described in Section 3.3), geostatistics were used to identify the planting pattern of the stands and estimate tree spacing from individual height measurements. The results were then compared to field data to quantify the accuracy of the lidar-derived parameters, considering that field plot

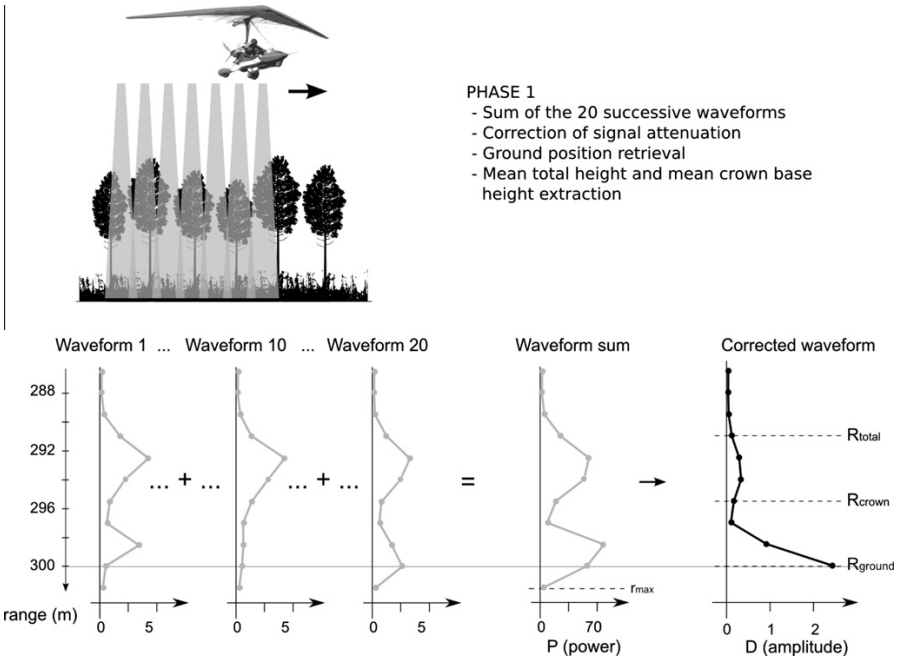



**Fig. 3.** Overview of the algorithm's first phase, for the extraction of mean total height and mean crown base height (phase 1).

measurements are statistically representative of the whole stands. The three phases were applied on the whole dataset, and phase 1 and phase 2 were also applied on the one flight line subset.

### 3.1. Waveform aggregation and signal attenuation correction

To retrieve the total tree height, we needed to have both a return from the canopy top and a return from the ground in a lidar waveform. Because the individual waveforms, which corresponded to a signal backscattered by a 4.5 $m^2$ circular area, did not systematically include a ground return, we chose to aggregate the 20 successive shoots. This approach allowed us to increase the probability of obtaining a ground return in the waveform sum corresponding to a signal backscattered in an area of about 30 m × 2.4 m (72 $m^2$). This method also increased the signal-to-noise ratio within the square root of the amount of summed waveforms, i.e., $\sqrt{20} \approx 4.5$ (Measures, 1984), because the noise can be approximated as a random process.

Before summing, each of the 20 successive waveforms was first georeferenced, and their altitudes were synchronized using the plane attitude information (DGPS + IMU). The amplitudes of the 20 successive waveforms were then summed. To prepare the correction of the laser attenuation, an estimation of the last echo location ($r_{max}$) was performed by selecting the last point of the waveform sum exceeding twice the maximum noise level. This maximum noise level corresponded to the maximum waveform amplitude during the laser pulse propagation through the atmosphere. Although it may increase the probability of a non-detection, this threshold level was set using a generate and test method to decrease the risk of false alarm (the probability of which can be considered equal to 0).

The waveform sums were then corrected to account for the attenuation of the laser beam intensity along its passage through the canopy. To this aim, we adapted the MacArthur and Horn equation used for computing foliage profiles (MacArthur and Horn, 1969). A similar approach was used by Lefsky et al. (1999) to compute canopy height profiles. Our adapted equation corrected the signal attenuation inside the vegetation using a logarithmic function:

$$D_{r_i} = \ln\left(\frac{\sum_{r_i}^{r_{max}} P(r)}{\sum_{r_{i+1}}^{r_{max}} P(r)}\right) \quad (1)$$

where $D_{r_i}$ is the amplitude of the corrected waveform at range $r_i$ from the plane with $i \in [1, r_{max-1}]$, and $P(r)$ is the recorded power as a function of range. The correction is calculated for each interval between two successive ranges from the atmosphere to $r_{max}$ (Fig. 3).

Contrary to the method for processing ground measurements proposed by MacArthur and Horn (1969), we performed the correction calculation from sky to ground to deal with the specificity of airborne lidar measurements. As a result, our corrected waveform contained a component of energy reflected from the ground. Because we did not take into account the difference in reflectance between ground and vegetation, we could not call the corrected waveforms "foliage profiles". However, this correction allowed us to obtain an amplitude profile close to a real foliage profile and to have an enhanced ground peak, making the determination of the ground location easier and more accurate (Fig. 3).

#### 3.1.1. Mean total height and mean crown height assessment

The last peak location of the corrected waveform was considered to correspond to the ground range ($R_{ground}$) (Fig. 3). The first point of the corrected waveform that exceeded the noise threshold was considered to correspond to the top of canopy range ($R_{total}$). This noise threshold was set as twice the noise level measured on the corrected waveform in the atmosphere using a generate and test method to support non-detection instead of false alarm. The point located after the strongest decrease in the corrected waveform and a 3-m minimum above $R_{ground}$ (to avoid undergrowth detection) was considered to correspond to the crown base range ($R_{crown}$). The distance between $R_{ground}$ and $R_{total}$ was assumed to be the mean total height of the trees located inside the 20 lidar footprints used for computing the corrected waveform. The distance between $R_{ground}$ and $R_{crown}$ was assumed to be the mean

crown base height. We calculated the mean total height and the mean crown base height of all 20 shoot sequences included in the studied stands. Heights were also calculated for shoot sequences of a unique flight line. The resulting values were analyzed against the field-based measurements.

### 3.2. Re-processing of individual waveforms

The first phase was designed to provide robust measurements of the ground position and mean tree height. However, finer-scale measurements were necessary to obtain information about the spatial variability of the stands' structural characteristics.

Hence, we corrected the lidar signal as presented in Section 3.1 but now for each individual waveform instead of the waveform sum. The process was performed from the atmosphere to the $r_{max}$ determined in the first phase for each group of 20 shoots. For each individual corrected waveform, we extracted the total height range $R_{\text{IndivTop}}$ but kept the robust measurement of $R_{\text{ground}}$ extracted in the first phase. The individual total heights were then calculated as the difference between $R_{\text{IndivTop}}$ and $R_{\text{ground}}$.

#### 3.2.1. Top height assessment

To assess the lidar top height, the tallest individual height measurement in each series of 20 shoots was first selected. To keep the ratio of 100 trees measured per hectare specified in the definition of the dominant height in forestry, we selected approximately 72% of the tallest individual measurements while considering a theoretical measurement area of 0.072 ha for the 20 shoot lidar footprint ($30\,\text{m} \times 2.4\,\text{m}$). Top height was calculated for both the entire dataset and for shoot sequences of a unique flight line. The resulting values were compared to the reference top height measurements.

### 3.3. Semi-variogram computation from individual measurements

Individual measurements of total height provided additional precision for studying the spatial variation of forest stands. In this section, we present geostatistical-based methods for assessing planting patterns and tree spacing. The following methods were applied only on the whole dataset.

#### 3.3.1. Semi-variograms and their use in remote sensing

High spatial resolution sensors allow the textures of observed surfaces to be identified. For targets having a periodic pattern, the best way to identify texture is to use the Fourier transform (e.g., Delenne et al., 2008) or wavelet transform (e.g., Ranchin et al., 2001). However, these techniques can only be applied to data resulting from regularly spaced measurements, such as measurements provided by image sensors. Because our lidar measurements were irregularly distributed in space, the Fourier or wavelet transforms could not be performed. Consequently, we used variograms, which can be applied to regular or irregular data as well as continuous or discontinuous data.

In spatial statistics, the variogram, or semi-variogram, is a function describing the spatial correlation in observations measured at sample locations. The variogram displays the variance within groups of observations plotted as a function of distance between the observations. For all observations $z_i$ at locations $s_i, \ldots, s_k$ with $i = 1, \ldots, k$, the omnidirectional variogram $\hat{\gamma}(d)$ is defined as (Cressie, 1993):

$$\hat{\gamma}(d) = \frac{1}{2|N(d)|} \sum_{(i,j)\in N(d)} (z_i - z_j)^2 \tag{2}$$

where $d$ is the distance between observations, $N(d)$ denotes the set of distinct pairs of observations $(i,j)$ such that $|s_i - s_j| = d \pm \Delta d$ ($\Delta d$ is a tolerance distance) and $|N(d)|$ is the number of distinct pairs in the set.

The variogram can also be computed with pairs of observations satisfying a directional constraint (directional variogram), such as

$$\hat{\gamma}(d,\theta) = \frac{1}{2|N(d,\theta)|} \sum_{(i,j)\in N(d,\theta)} (z_i - z_j)^2 \tag{3}$$

where $N(d,\theta)$ denotes the set of distinct pairs of observations $(i,j)$ such that $|s_i - s_j| = d \pm \Delta d$ and $(\vec{s}_i \cdot \vec{s}_j) = \theta \pm \Delta\theta$. $|N(d,\theta)|$ is the number of distinct pairs in the set.

A variogram plotted from observations is called an empirical variogram, and it can be approximated by a model function to estimate the following three parameters: nugget, sill and range. The nugget represents the semi-variogram discontinuity at the origin (which shows the spatially random component), the sill is the maximum variance when the distance tends toward infinity and the range is the distance at which the variogram reaches the sill.

Semi-variograms have been used in remote sensing for textural information assessment, but they were applied to optical images (St-Onge and Cavayas, 1997). Consequently, the presence of a planting pattern was indirectly studied though sun-induced shadows on the image. In contrast, lidar data give pure elevation information that can be directly used to study the spatial structure of a stand. A recent study used variograms to initialize the size of a search window used for locating the apex of trees (Tesfamichael et al., 2009). In our case, the variograms were computed with a 2.4 m binning to make them consistent with the laser footprint size. Because of the footprint size and the sampling pattern of the lidar measurements, it was not possible to precisely assess the stand structure. However, we assumed that the identification of a directional structure tendency in a stand and the mean tree spacing could be derived from semi-variograms.

#### 3.3.2. Planting pattern assessment

Tree position and the spatial distribution of tree heights can present various patterns according to the origin of the stand, its age and the management practices. In mono-specific even-aged stands, two main patterns are commonly found according to the origin of the stand: planted or semi-natural. Unlike a semi-natural stand, a plantation can present a strong directional pattern. This pattern is studied in this paper through the computation of directional variograms.

Considering all pairs of total height measurements, the empirical variograms were then computed for six directions from the north and clockwise (0°, 30°, 60°, 90°, 120°, 150°) with a tolerance angle of 15°. This range of angles was chosen to explore anisotropy with a better accuracy than using the four cardinal directions alone, while keeping a sufficient number of observation pairs (>80) in each variogram sample. The existence of a directional anisotropy was studied in each stand through visual interpretation of the directional variogram shapes. A difference in shape according to different directions was expected for stands planted in lines due to the distance between trees along the line being shorter than the inter-line distance.

#### 3.3.3. Tree spacing assessment

When no clear anisotropy was revealed by the previous analysis with empirical directional variograms, the omnidirectional variogram was computed and modeled.

A canopy variogram is traditionally approximated with widely accepted circular or spherical models (Curran, 1988; St-Onge and Cavayas, 1997; Tesfamichael et al., 2009) with ranges that provide a good estimation of the mean spacing between objects (see Wackernagel, 1995, p. 45). However, we assumed that the spatial distribution of trees in semi-natural stands could be modeled using a

**Table 2**
Tree (total and crown) heights extracted from corrected waveform sums and their comparison with mean heights calculated from field measurements. *n* is number of measurements (corrected waveform sum) for each stand.

| | Young plantation | | Mature stand | |
|---|---|---|---|---|
| | Whole dataset ($n = 94$) | One flight line ($n = 8$) | Whole dataset ($n = 83$) | One flight line ($n = 8$) |
| Mean of lidar total heights ± standard deviation (m) | 9.5 ± 1.1 | 9.6 ± 1.6 | 20.9 ± 1.6 | 21 ± 1.3 |
| Mean total height bias (lidar – field) (m) | 0.2 | 0.3 | −0.7 | −0.6 |
| Mean of lidar crown base heights ± standard deviation (m) | 4.6 ± 1.2 | 4.7 ± 1.5 | 15.0 ± 1.9 | 15.2 ± 1.5 |
| Mean crown base height bias (lidar – field) (m) | −0.3 | −0.2 | −0.3 | −0.1 |

**Table 3**
Top heights extracted from corrected waveforms and their comparison with those calculated from field measurements. n represents the number of measurements used to compute the top height.

| | Young plantation | | Mature stand | |
|---|---|---|---|---|
| | Whole dataset (n = 59) | One flight line (n = 6) | Whole dataset (n = 68) | One flight line (n = 6) |
| Lidar top height ± standard deviation (m) | 10.5 ± 1.2 | 10.5 ± 1.3 | 22.2 ± 1.1 | 22 ± 1.2 |
| Top height bias (lidar – field) (m) | 0.6 | 0.6 | – 1.0 | – 1.2 |

random process called the "dead leaves model" (Lantuejoul, 2002, p. 175). This model is "constructed from hard spheres of constant diameter" (Gille, 2002) that represent tree crowns. This model was also demonstrated to be suitable for describing the spatial correlation in a pine stand and was used by Boone and Bullock (2008) to characterize inter-tree competition. Because Gille (2002) showed that the outcome of this random process is a variogram that can be approximated by a Matern model, we used this model to approximate the empirical variogram on the semi-natural stand. The Matern model generalizes exponential models with asymptotic sill. For such models, the "practical range" (which is equivalent to the range parameter of fixed sill models) corresponds to the distance where the semi variance reaches 95% of the sill. It equals three times the range parameter given in the Eq. (4) (Wackernagel, 1995, p. 41). Consequently, the practical range of the Matern model is used to assess the mean spacing between objects. The Matern model equation is (Ribeiro and Diggle, 2010):

$$C(d) = \frac{1}{\Gamma(v) \times 2^{v-1}} \times \left(\frac{d}{\varphi}\right)^{v} \times K_v\left(\frac{d}{\varphi}\right) \quad (4)$$

where $\varphi$ is the range parameter, $v$ is the smoothness parameter, $K_v$ is the modified Bessel function of the third kind of order $v$ and $\Gamma$ is the gamma function. The function is valid for $\varphi > 0$ and $v > 0$.

The Matern model was then automatically fitted to the empirical variogram to adjust its parameters (nugget, sill and range) using the R software with the geoR package (Ribeiro and Diggle, 2010). This software provides a non-linear least squares adjustment algorithm weighted by the number of pairs in each distance class. Afterward, the practical range of the fitted model was computed as an estimation of the mean spacing between trees.

## 4. Results

### 4.1. Mean total height and mean crown height assessment

Table 2 shows a quantitative assessment of the mean total and mean crown heights. There is a slight overestimation of the mean total height in the young stand and a slight underestimation in the mature stand. Nevertheless, the absolute values of these errors are lower than the standard deviation of the reference tree heights (0.9 m in the young stand and 2 m in the mature stand; see Table 1).

The error on the mean crown height presents a slight underestimation, but its absolute value is also lower than the standard deviation of the reference crown heights (0.8 m in the young stand and 1.6 m in the mature stand; see Table 1).

The values computed for the entire stand and for one flight line are consistent and differ from only 10–20 cm.

### 4.2. Top height assessment

About 72% of the tallest individual measurements represent 59 selected spots for the young stand and 68 for the mature stand, when using the whole dataset. When considering only one flight line, six measurements were used to compute top height.

The top height was estimated with a 1 m underestimation within the mature stand (Table 3) but this assessment was not significantly different from the reference measurement because the absolute error was lower than the standard deviation of the reference measurements. In contrast, the top height derived from the lidar data in the young plantation was overestimated by 0.60 m, which is 0.1 m above the standard deviation of the reference measurements.

The values computed for the entire stand and for one flight line are consistent and differ from only 0–20 cm.

### 4.3. Planting pattern assessment

Directional variograms are shown in Figs. 4 and 5 for the young and mature stands, respectively. Fig. 4 shows a significant nugget anisotropy (Zimmerman, 1993). A directional anisotropy can be seen for a distance between measurements ranging between 0 and 4 m, through two distinct clusters of semi-variance values for the first distance class (1.2 m). One cluster is composed of 0–60° directions without any spatial correlation (stationary or decreasing semi-variance along distance), while the other contains directions from 90° to 150° with little spatial correlation (increasing semi-variance) for distances up to 4 m.

Because the directions of the plantation lines were close to 120° (see Fig. 2), the first cluster refers to the perpendicular direction of the lines, while the second cluster refers to the same direction as the plantation lines. The presence of clusters can be explained by a stronger variance of lidar-based vegetation heights in the perpendicular direction of the plantation lines than along the flight lines. In the perpendicular direction, the height measurements alternatively reached the tree tops and the ground/low vegetation due to gaps between lines. More homogeneous measurements were made along the lines due to the absence of gaps between tree crowns.

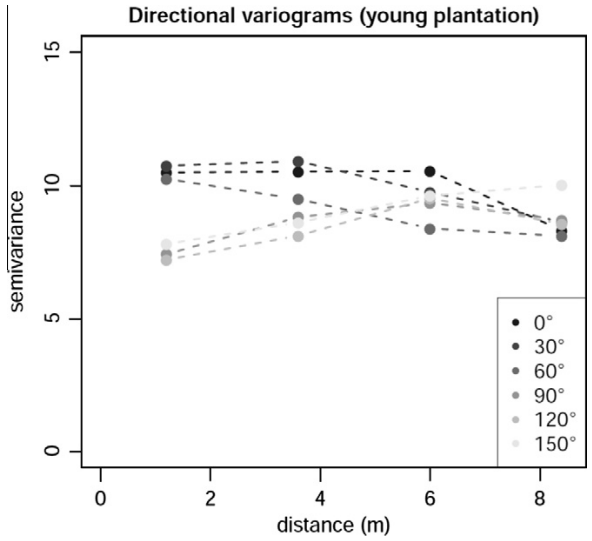



**Fig. 4.** Directional variograms for six directions in the young plantation.

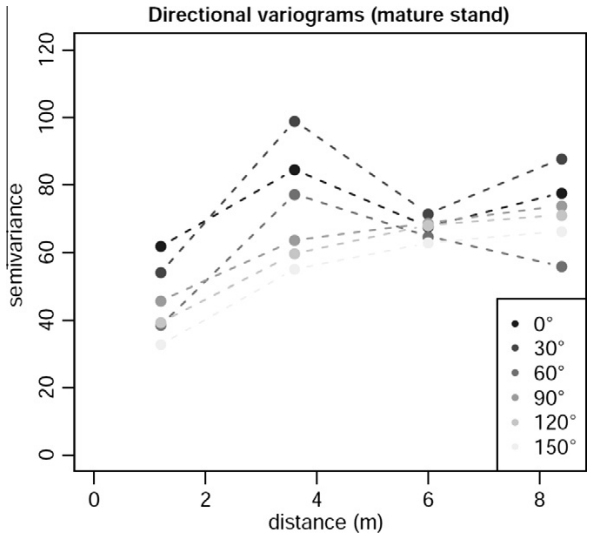



**Fig. 5.** Directional variograms for six directions in the mature semi-natural stand.

In contrast, Fig. 5 does not show semi-variance clusters in specific directions. Thus, the mature stand did not show a specific planting pattern, and it was consequently assumed not to be planted in lines. However, the gradient of variance depending on variogram direction shows a geometric anisotropy between 0 and 4 m. This observation can be explained by the crown shape anisotropy. Actually, the mean crown dimensions are 0.9 m greater in the north direction than the east direction, as shown by the reference data in Table 1, which typically produces a geometric anisotropy.

### *4.4. Tree spacing assessment*

Fig. 6 shows the results of the Matern model fitted on the empirical variogram for the mature stand. The "practical range" of the variogram gave an estimation of the mean spacing between trees of 5.1 m ($3 \times 1.7$), which has exactly the same value as the estimation derived from the field measurements.

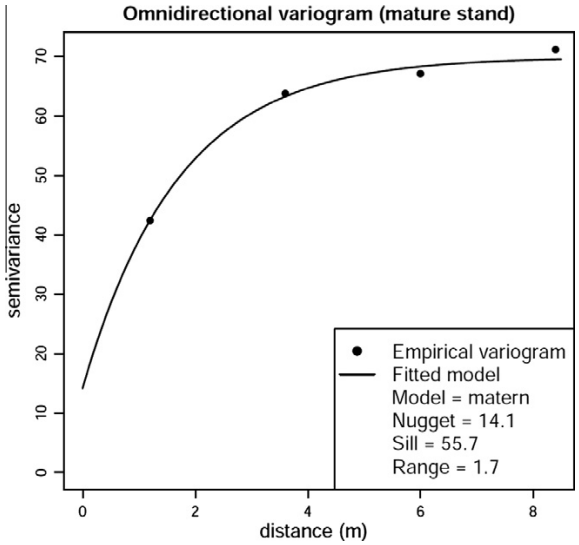



**Fig. 6.** Omnidirectional variogram for the mature, semi-natural stand.

## 5. Discussion

Methods presented in this paper were successfully tested on even-aged, monospecific stands. If inaccurate crown base height detection is liable to occur in multi-layered forests, retrieval of top height aims to work on most forests. The only requirement is the forest to be transparent enough for the lidar signal to hit the ground and come back. But the method presented here to extract heights was designed to avoid problems linked to non-systematic ground detection. Moreover, the number of lidar shoots to aggregate can be tuned according to the probability of a shoot to hit the ground. A denser forest will need a larger number of shoots to be summed for retrieving local ground position.

### *5.1. Global positioning and terrain influence*

As explained in the Section 2.2, the global positioning accuracy of laser profiles (series of 20 lasers shoots) was around 4.5 m, with a relative accuracy lower than 1 m between spots of a same series. This is a clear restriction for computing high resolution digital terrain models. However, methods proposed in this paper for measuring trees and planting patterns were designed to get rid of this inaccurate positioning.

To this aim, tree heights were measured relatively to the ground position detected inside an aggregated waveform, whereas for small-footprint lidar data heights are extracted after an absolute positioning of each echo. This relative measurement of tree height allowed us to accurately measure trees.

However, height measurements could not be accurately positioned in a global referential. Consequently, a tree to tree comparison with reference measurement could not be performed as it is often done in studies using small-footprint data. Instead, the comparison was statistically performed at stand level, considering that the reference plots were representative of the entire stands. Despite the low absolute positioning accuracy, we were able to recognize the planting pattern of stands and estimate the distance between trees. The reason is that variograms were computed between 0 and 8 m. At these distances, the majority of measurements pairs came from the same series of shoots and they consequently benefit from the sub metric relative positioning accuracy.

Retrieving canopy top and ground position from the same waveform seems similar to direct methods used for large-footprint

data (Chen, 2010), but it is here much less sensitive to terrain slope influence on ground positioning, thanks to a smaller footprint size. Considering a 2.4 m footprint containing a ground echo, the inaccuracy of ground positioning will lead to a 2.4 m error in a 100% terrain slope. In the case of a waveform aggregation the error will rise, according to a bigger footprint size, but re-processing individual waveforms (which is not possible with large-footprint data) for refining the ground positioning is conceivable and would lead to improve height estimates.

### 5.2. Phase 1 and 2: mean total height, mean crown height, and top height assessment

The mean total height assessment did not seem to notably benefit from the correction of the signal attenuation compared to Cuesta et al. (2010). However, the accuracy estimation of the mean crown height was increased by about 1.7 m compared to the previous study. This result demonstrates the importance of correcting the signal from its attenuation when retrieving structural parameters inside the canopy.

The lidar measurement was expected to underestimate the tree height due to the combination of two factors. First, the resolution $r$ of the waveform digitizer (1.5 m) should lead to a mean underestimation of the tree height between $r/2$ and $r$ (0.75–1.5 m), which is partially compensated for by the large pulse width (5 ns, i.e., 1.5 m) of the current system. Second, the tree heights were underestimated due to a reduced probability of the laser to exactly reach the tree tops (e.g., Chauve et al., 2009). Indeed, the tree tops are highly punctuate regarding the stand area, especially in mature stands, and the probability of detecting them is inversely proportional to the number of trees. The results found were in agreement with these statements, as we observed an underestimation between 0.1 and 0.7 m for almost retrieved heights. Nevertheless, the total height estimation for the young plantation was overestimated. A total height underestimation of tall trees jointly with an overestimation of young trees was already reported by Brandtberg et al. (2003). The explanation they gave was that field measured tree heights were affected by error introduced by the field personnel. But we had no difficulties to locate top of trees from the ground in our young plantation. A possible explanation lies in the consequence of the vertical sampling of the backscatter signal on height measurements. However this overestimation is lower than the standard deviation of the terrain measurements indicating a good accuracy of estimates.

The accuracy of the method used for estimating top height should depend on stand characteristics. Stands with fewer trees would normally require a greater number of measurements for computing an accurate top height. The less accurate estimation of top height in the young plantation could possibly be linked to a bias in the reference data. Only one out of every three trees was measured in the young stand plots. A detailed inspection of the reference data collected in plots y1 and plot y2 showed that the total height of the 9 trees having the largest diameters at breast height in each plot were not measured. Consequently, the reference top height was not accurately set in the young plantation, resulting in a gap with the lidar-retrieved top height. The underestimation of the top height in the mature stand is certainly linked to the causes of underestimation already discussed above. However, the reliability of the top height assessment in the mature stand was demonstrated, as no error exists in reference data, and the bias was lower than the standard deviation of the terrain measurements.

The slight difference between measurements computed for the entire dataset and for only one flight line shows that a 2.4 m wide non-continuous transect across an homogeneous stand allows to provide reliable mean tree height, mean crown height and top height estimates. This result is highly encouraging for providing forest inventories on large area with low costs. If parcel or of stand-type maps are available, a well designed flightplan with at least one flight line intersecting each forest unit would provide mean and top heights at least on even-aged stands. Other experiments are required to check if this result is still valid on multilayered forests, but we can imagine that one flight line over randomly positioned trees will give a good description of the entire stand.

### 5.3. Phase 3: planting pattern and tree spacing assessment

Contrary to height retrieval methods, the following analysis using variograms needs an explicit lidar sampling of the forest stand to work. At least, it would require two perpendicular flight lines to compute directional variograms. Several flight lines in two perpendicular directions would provide an even bigger amount of measurement pairs per class of distance for computing significant variograms. In this study, flight lines where performed in only one direction. Consequently, we got a higher number of measurement pairs in the direction of the flight lines. However, the ULA attitude provided not always straight measurements, allowing then to get a sufficient number of pairs in every direction. We checked that each variogram sample was computed with a minimum of 80 distinct pairs.

In Fig. 5, we can notice a weak behavior in semi-variance between 6 and 8 m for directions 0° and 30° (a decrease followed by an increase in semi-variance) resulting from some measurement variations that were not totally captured at small distances. Soil physical properties may produce such a cyclic phenomenon in variograms also known as "hole-effect" (Journel and Huijbregts, 1978). Soil samples have to be taken to validate this hypothesis, but the aim of the directional variogram analysis was only to assess the presence of an anisotropy based on the nugget parameter.

Associating a confidence value to the detection of a directional anisotropy would have been useful, but it was not possible to perform due to the difficulty in simulating a non-Gaussian distribution of height for such a test. The way to calculate a confidence value is to verify whether an empirical directional variogram is included in the confidence envelope built from a simulated dataset satisfying an isotropy hypothesis (Lantuejoul, 2002). The simulated dataset also needs to fit the model previously adjusted on the empirical omnidirectional variogram, and this methodology requires a Gaussian distribution of the studied variable. Because the distribution of our lidar data was non-Gaussian, as it was not composed of individual trees heights but of maximum heights measured within footprints, such a test needs further methodological development. Such developments, at present time, should rely on transforming non-Gaussian to Gaussian fields or indicator fields (Emery, 2002).

The variogram model is a continuous function defined in $R^+$. Although variogram estimation starts at distance 1.2 m (Fig. 6) due to the minimum spacing between two lidar measurements, the function was plotted from distance 0 m where it always have variance 0. The variogram function values between 0 and 1.2 m result from model choice and parameters fit. They mainly reveal the nugget parameter with an uncertainty due to the lack of data at these distances. However, the method presented in this paper allows estimating the mean spacing between trees through the practical range parameter. The uncertainty between 0 and 1.2 m did not affect the practical range parameter and the tree spacing was accurately estimated on the semi-natural stand. In the young stand, the mean spacing between trees in a line and between lines may differ, and they could not be extracted from the unique omnidirectional variogram model. Nevertheless, it may be possible to model directional variograms to retrieve tree spacing in specific directions. In this study, however, the georeferencing accuracy was insufficient to finely estimate the metrics using modeled

directional variograms on different series of 20 shoots. A shoot was well georeferenced relative to the others inside the same group of 20 shoots, but as explained in Section 2.2, the absolute accuracy (between two series of shoots) was about 4.5 m. Such accuracy was insufficient to extract metrics from directional variograms in the perpendicular direction of the flight lines, but we demonstrated the possibility of identifying tendencies in the spatial distribution of trees, such as the identification of the planting pattern.

Despite that the variogram based methods cannot be used to process a one flight line profiler dataset, it could be of great interest to process small-footprint lidar data. It would allow retrieving mean inter-tree distance, and therefore stand density, through a direct processing of the first return lidar point cloud after correcting heights from ground elevation, instead of searching for local maxima on previously interpolated raster Digital Canopy Models.

### 5.4. Spaceborne UV lidar

Studying vegetation and aerosols at the same time was not possible in this experiment because the sensor was too close to the forest. Aerosols detection would have required the study of too small signal variations compared to those produced by the laser backscattering on the canopy. But a simultaneous detection of both forest and aerosols would be easier from satellite platforms due to a lower signal dynamics caused by the atmospheric transmission coefficient.

Beside Rayleigh scattering caused by atmosphere gases, clouds will play a major role in laser attenuation. Despite that semi-transparent cirrus clouds will not have a drastic impact on laser move forward, dense cumulus-type clouds will produce signal extinction. But the problem is not limited to ultraviolet wavelength and also occurs in both visible and near-infrared wavelenghts (Chazette et al., 2001). However, the UV laser would be less sensitive to multiple backscattering in the vegetation, due to a smaller leaf reflectance in the UV compared to NIR (Grant et al., 2003), and would consequently increase the accuracy of ground positioning under the canopy compared to existing spaceborne lidars.

## 6. Conclusion

This paper presents methods developed for retrieving forest parameters from a new, eye-safe medium-footprint lidar sensor initially designed for atmospheric applications using an ultraviolet laser. The advantage of a medium-footprint size is that it provides higher resolution data than large footprint systems while increasing the probability of reaching tree tops more effectively than small-footprint lidar data. A first evaluation of the system and of its potential for forest parameter retrieval took place in one of the major planted forests in Europe, the Landes forest in France. Both the simple stand structure and the flat topography were expected to make the data processing and interpretation of results easier.

The methods presented in this paper were developed to retrieve mean total height, mean crown height, top height, planting pattern and tree spacing according to the specificity of medium-footprint lidar data. Despite the absolute positioning uncertainties due to the lack of ULA accurate attitude data, the mean total height, mean crown height and top height were estimated with an absolute error generally lower than the standard deviation of the reference measurements. In addition, the obtained results regarding crown height estimation were better than in the previous study (Cuesta et al., 2010), demonstrating the importance of correcting the signal from its attenuation for assessing forest structure inside the canopy. A method using geostatistics for recognizing the planting pattern was also performed successfully, and the tree spacing was accurately estimated for the semi-natural stand. Although the methods presented in this paper were developed for medium-footprint lidar data, they can be also used for processing small- or large-footprint lidar data. In particular, the correction of laser attenuation inside vegetation is applicable to any kind of lidar waveform, while the geostatistically based processes might give even more precise results on a small-footprint lidar dataset which is normally of better absolute precision than the one available here.

Consequently, the LAUVA prototype demonstrated its ability to assess forest structure. The lightness of this lidar sensor allows the use of an ultra-light aircraft, which can be more rapidly deployed than other airborne missions. The cost is also lower, particularly for surveys of small areas, which are not optimal with other airborne platforms. In addition, we demonstrated that measuring canopy height along a single 2.4 m wide transect across a stand is sufficient to provide reliable usual inventory parameters, which can be of great interest to perform larger scale inventories at a minimal cost (saving flight time, data storage and process). It is therefore a valuable sensor for performing emergency forest inventory. However, we hope to increase the georeferencing accuracy using a finer IMU in future developments. It would allow us to accurately position each laser spot in a global referential. When a dense and spatially explicit sampling strategy is chosen this would make possible the computation of high resolution digital terrain models, and would allow further investigations on variogram methods to refine the estimation of forest parameters. The waveform digitization frequency will also be increased, providing additional accuracy for tree height measurements. As LAUVA is a profiler system, the sampling strategy could be rethought on the basis of the results obtained in this study to optimize the sampling strategy and increase the speed of data acquisition. One option would be to perform continuous measurements with only one transect per parcel. In such a configuration, the processing methods will also need to be rethought and will probably be closer to existing process for spaceborne lidar. The success in forest structure estimation also gives confidence for the adaptation of this system for a spaceborne mission (Flamant, 2005) with a smaller footprint than existing systems. To that aim, additional experiments will be performed at different flight altitudes, for obtaining various footprint sizes up to 10–15 m. Hence, the impact of the footprint dimension on measurement accuracy of forest structure will be assessed. The main advantage of such a system would be to limit the terrain slope influence on the tree height measurements compared to ICESat system with a 70 m footprint diameter. However, this system will require further methodological work to process waveforms with footprints smaller than 72 $m^2$. Finally, as the initial function of the LAUVA system was to measure atmospheric aerosols (Raut and Chazette, 2009), the first results obtained on forests confirm that the conception of a bi-functional lidar for studying forest responses to atmospheric pollution seems also feasible.

## Acknowledgments

This work is part of the ExFOLIO project and was realized thanks to the financial support of the CNES (Centre National d'Études Spatiales), the CEA, the CNRS (PNTS funding), and the French Languedoc-Roussillon region. The authors also thank the Mimizan town hall representatives (Mrs. Dehureau-Foidart and Mr. Bertrand) for facilitating the mission. We also thank Laurent Albrech, from UMR TETIS-Cemagref, for developing the georeferencing algorithm.

## References

Boone, E.L., Bullock, B.P., 2008. Spatial correlation matrix selection using Bayesian model averaging to characterize inter-tree competition in loblolly pine trees. Journal of Applied Statistics 35 (9), 967–977.

Brandtberg, T., Warner, T.A., Landenberger, R.E., McGraw, J.B., 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density lidar data from the eastern deciduous forest in North America. Remote Sensing of Environment 85 (3), 290–303.

Chauve, A., Vega, C., Durrieu, S., Bretar, S., Allouis, T., Pierrot-Deseilligny, M., Puech, W., 2009. Advanced fullwaveform lidar data echo detection: assessing quality of derived terrain and tree height models in an alpine coniferous forest. International Journal of Remote Sensing 30 (19), 5211–5228.

Chazette, P., Pelon, J., Mégie, G., 2001. Determination by spaceborne backscatter lidar of the structural parameters of atmospheric scattering layers. Applied Optics 40 (21), 3428–3440.

Chazette, P., Sanak, J., Dulac, F., 2007. New approach for aerosol profiling with a lidar onboard an ultralight aircraft: application to the African monsoon. Environmental Science & Technology 41 (24), 8335–8341.

Chazette, P., Raut, J.-C., Dulac, F., Berthier, S., Kim, S-W., Royer, P., Sanak, J., Loaëc, S., Grigaut-Desbrosses, H., 2010. Simultaneous observations of lower tropospheric continental aerosols with a ground-based an airborne and the spaceborne CALIOP lidar systems. Journal of Geophysical Research 115 (16), D00H31.

Chen, Q., Gong, P., Baldocchi, D., Tian, Y.Q., 2007. Estimating basal area and stem volume for individual trees from lidar data. Photogrammetric Engineering & Remote Sensing 73 (12), 1355–1365.

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 114 (7), 1610–1627.

Clark, M.L., Clark, D.B., Roberts, D.A., 2004. Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. Remote Sensing of Environment 91 (1), 68–89.

Cressie, N., 1993. Statistics for Spatial Data, Wiley Interscience.

Cuesta, J., Chazette, P., Allouis, T., Flamant, P.H., Durrieu, S., Sanak, J., Genau, P., Guyon, D., Loustau, D., Flamant, C., 2010. Observing the forest canopy with a new ultra-violet compact airborne lidar. Sensors 10 (8), 7386–7403.

Curran, P.J., 1988. The semivariogram in remote sensing: an introduction. Remote Sensing of Environment 24 (3), 493–507.

DeFries, R., Achard, F., Brown, S., Herold, M., Murdiyarso, D., Schlamadinger, B., de Souza Jr., C., 2007. Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. Environmental Science & Policy 10 (4), 385–394.

Delenne, C., Durrieu, S., Rabatel, G., Deshayes, M., Bailly, J.S., Lelong, C., Couteron, P., 2008. Textural approaches for vineyard detection and characterization using very high spatial resolution remote sensing data. International Journal of Remote Sensing 29 (4), 1153–1167.

Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G.k., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F., Prince, S., 2002. Estimation of tropical forest structural characteristics, using large-footprint lidar. Remote Sensing of Environment 79 (2–3), 305–319.

Emery, X., 2002. Conditional simulation of nongaussian random functions. Mathematical Geology 34 (1), 79–100.

ESA, 2008. Biomass, mission assessment report, SP-1313/2. Available from: http://esamultimedia.esa.int/docs/SP1313-2_BIOMASS.pdf (accessed 4.07.11).

Flamant, P.H., 2005. Atmospheric and meteorological lidar: from pioneers to space applications. Comptes Rendus Physique 6 (8), 864–875.

Garcia, O., Batho, A., 2005. Top height estimation in lodgepole pine sample plots. Western Journal of Applied Forestry 20 (1), 64–68.

Gille, W., 2002. The set covariance of a dead leaves model. Advances in Applied Probability 34 (1), 11–20.

Grant, R.H., Heisler, G.M., Gao, W., Jenks, M., 2003. Ultraviolet leaf reflectance of common urban trees and the prediction of reflectance from leaf surface characteristics. Agricultural and Forest Meteorology 120 (1–4), 127–139.

Hofton, M.A., Minster, J.B., Blair, J.B., 2000. Decomposition of laser altimeter waveforms, IEEE Transactions on Geoscience and Remote Sensing 38 (4 II), 1989–1996.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, New York.

Kato, A., Moskal, L.M., Schiess, P., Swanson, M.E., Calhoun, D., Stuetzle, W., 2009. Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. Remote Sensing of Environment 113 (6), 1148–1162.

Lantuejoul, C., 2002. Geostatistical Simulation: Models and Algorithms. Springer Verlag, Berlin, 256 p.

Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., Harding, D., 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. Remote Sensing of Environment 70 (3), 339–361.

Lefsky, M.A., Harding, D.J., Keller, M., Cohen, W.B., Carabajal, C.C., Del Bom Espirito-Santo, F., Hunter, M.O., de Oliveira Jr., R., 2005. Estimates of forest canopy height and aboveground biomass using ICESAT. Geophysical Research Letters 32 (22), 1–4.

MacArthur, R.H., Horn, H.S., 1969. Foliage profile by vertical measurements. Ecology 50 (5), 802–804.

Mallet, C., Bretar, F., 2009. Full-waveform topographic lidar: state-of-the-art. ISPRS Journal of Photogrammetry and Remote Sensing 64 (1), 1–16.

Means, J.E., Acker, S.A., Harding, D.J., Blair, J.B., Lefsky, M.A., Cohen, W.B., Harmon, M.E., McKee, W.A., 1999. Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western cascades of Oregon. Remote Sensing of Environment 67 (3), 298–308.

Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., Hendrix, C.J., 2000. Predicting forest stand characteristics with airborne scanning lidar. Photogrammetric Engineering & Remote Sensing 66 (11), 1367–1371.

Measures, R.M., 1984. Laser Remote Sensing: Fundamentals and Applications. John Wiley & Sons, Inc., New York.

Næsset, E., 2009. Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. Remote Sensing of Environment 113 (1), 148–159.

Nakai, T., Sumida, A., Kodama, Y., Hara, T., Ohta, T., 2010. A comparison between various definitions of forest stand height and aerodynamic canopy height. Agricultural and Forest Meteorology 150 (9), 1225–1233.

Packalen, P., Suvanto, A., Maltamo, M., 2009. A two stage method to estimate species-specific growing stock. Photogrammetric Engineering & Remote Sensing 75 (12), 1451–1460.

Popescu, S.C., 2007. Estimating biomass of individual pine trees using airborne lidar. Biomass and Bioenergy 31 (9), 646–655.

Ranchin, T., Naert, B., Albuisson, M., Boyer, G., Astrand, P., 2001. An automatic method for vine detection in airborne imagery using wavelet transform and multiresolution analysis. Photogrammetric Engineering & Remote Sensing 67 (1), 91–98.

Raut, J.-C., Chazette, P., 2009. Assessment of vertically-resolved PM10 from mobile lidar observations. Atmospheric Chemistry and Physics 9 (21), 8617–8638.

Ribeiro, Jr., P.J., Diggle, P.J., 2010. Package geoR version 1.6–27.

Silva-Santos, P., Valentim, H., Luís, A., Queirós, L., Travassos, P., Cabral, J.A., 2010. Stochastic dynamic methodology (StDM) to simulate the effects of fire on vegetation and bird communities in Pinus pinaster stands. Ecological Indicators 10 (2), 206–211.

Soenen, S.A., Peddle, D.R., Hall, R.J., Coburn, C.A., Hall, F.G., 2010. Estimating aboveground forest biomass from canopy reflectance model inversion in mountainous terrain. Remote Sensing of Environment 114 (7), 1325–1337.

St-Onge, B.A., Cavayas, F., 1997. Automated forest structure mapping from high resolution imagery based on directional semivariogram estimates. Remote Sensing of Environment 61 (1), 82–95.

Stam, D.M., 2008. Spectropolarimetric signatures of earth-like extrasolar planets. Astronomy and Astrophysics 482 (3), 989–1007.

Tesfamichael, S., Ahmed, F., van Aardt, J., Blakeway, F., 2009. A semi-variogram approach for estimating stems per hectare in Eucalyptus grandis plantations using discrete-return lidar height data. Forest Ecology and Management 258 (7), 1188–1199.

Véga, C., St-Onge, B., 2008. Height growth reconstruction of a boreal forest canopy over a period of 58 years using a combination of photogrammetric and lidar models. Remote Sensing of Environment 112 (4), 1784–1794.

Wackernagel, H., 1995. Multivariate Geostatistics. Springer Verlag, Berlin.

Waser, L., Baltsavias, E., Ecker, K., Eisenbeiss, H., Feldmeyer-Christe, E., Ginzler, C., Küchler, M., Zhang, L., 2008. Assessing changes of forest area and shrub encroachment in a mire ecosystem using digital surface models and CIR aerial images. Remote Sensing of Environment 112 (5), 1956–1968.

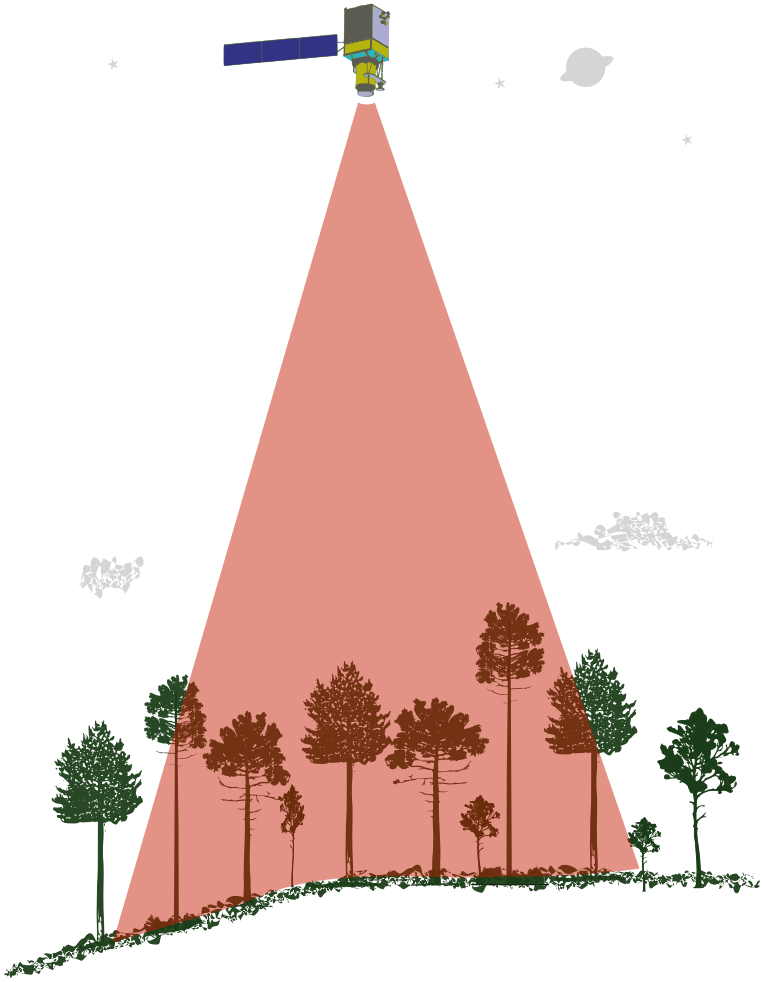
Williston, H.L., Balmer, W.E., 1974. Managing for Natural Regeneration. USDA Forest Service State and Private Forestry. Forest Management Bulletin. 6pp.

Wynne, R.H., 2006. Lidar remote sensing of forest resources at the scale of management. Photogrammetric Engineering & Remote Sensing 72 (12), 1310–1314.

Zimmerman, D.L., 1993. Another look at anisotropy in geostatistics. Mathematical Geology 25 (4), 453–470.

# Chapitre 3

# Simulation et traitement de signaux lidars à larges empreintes



Les capteurs présentés dans les deux parties précédentes sont adaptés à l'étude de la végétation depuis l'échelle de l'arbre jusqu'à celle d'un massif forestier, sur des surfaces d'acquisition pouvant aller jusqu'à quelques milliers de kilomètres carrés. Pour des raisons de coût, une couverture complète à l'échelle d'un pays reste cependant difficilement envisageable à partir d'un vecteur aéroporté. Entre 2003 et 2006, le satellite ICESat (Ice, Cloud,and land Elevation Satellite) a quant à lui fourni des données lidar que les forestiers ont utilisés pour étudier la structure de la végétation à l'échelle régionale (Boudreau et al., 2008). Bien qu'elle reste encore invalidée, une carte des hauteurs d'arbres à l'échelle globale a aussi été produite en combinant les données d'ICESat avec des images optiques (Lefsky, 2010).

Le besoin d'estimation de la ressource forestière à l'échelle mondiale pousse aujourd'hui les agences spatiales à envisager de nouvelles missions satellitaires partiellement ou totalement dédiées à l'étude de la végétation. Dans ce contexte, l'Irstea à répondu en 2010 à l'appel à proposition de l'Agence Spatiale Européenne (ESA) pour sa mission Earth Explorer 8 en proposant le satellite LEAF (Lidar for Earth And Forests) (*Durrieu, 2010*). La troisième partie de ce document présente les travaux réalisés dans le cadre de LEAF, à la demande du Centre National d'Études Spatiales (CNES). Ils portent sur la simulation de signaux lidar à larges empreintes par l'agrégation de données expérimentales à petites empreintes. L'objectif de cette étude était de développer une méthode permettant, à terme, de simuler la dynamique du signal dans différents contextes forestiers. La finalité de cette étude est d'apporter des informations permettant de dimensionner l'intensité d'émission du capteur LEAF. Une énergie trop faible ne permettrait pas de pénétrer les forêts les plus denses, alors qu'une énérgie trop forte ferait saturer la mesure en zone peu dense.

Bien que les performances d'ICESat aient été démontrées sur des terrains plats, une des principales limitations de ce système était la taille de son empreinte (70 m). À cette résolution, le topographie a une forte influence sur la mesure des hauteurs d'arbres. En effet, l'imprécision de positionnement du sol dans une empreinte de 70 m sur un terrain en pente de 10° peut mener à des erreurs de plus de 12 m sur la mesure des hauteurs d'arbres. À l'instar de LEAF, les nouvelles missions satellitaires tendent à proposer des tailles d'empreintes plus faibles, bien qu'une empreinte de 30 m mène toujours à des erreurs pouvant atteindre 5 m. Dans ce contexte, des développements méthodologiques en traitement du signal, dont les résultats sont présentés dans cette troisième partie, se sont montrés incontournables.

# Simulation de signaux lidars à larges empreintes à partir de données lidars topographiques

Tristan Allouis, Sylvie Durrieu

## 1. Introduction

Il existe au niveau mondial une grande diversité de types de forêts (boréales, tempérées, méditerranéennes, tropicales, etc.) présentant des caractéristiques très différentes de structure, en termes de hauteur, de densité et de stratification du couvert.

Le lidar a montré son potentiel pour mesurer certains paramètres de structure de la forêt, mais pour accéder au paramètre clé qu'est la hauteur, il est nécessaire que le système enregistre à la fois des informations sur la position du sommet de la végétation et sur la position du sol.

Dans les forêts hautes et denses, l'énergie du laser va être fortement atténuée et la quantité d'énergie qui va atteindre le sol risque d'être insuffisante. Un dimensionnement approprié du système est donc requis pour qu'une quantité suffisante de photons puisse atteindre le sol et être détectée par le capteur, tout en répondant aux normes de sécurité oculaire.

Parmi les paramètres système ajustables, on trouve la puissance d'émission, la dimension du télescope ainsi que le diamètre de l'empreinte au sol.

Dans le cadre des études de phase 0 (identification des besoins, analyse de la mission) pilotées par le CNES, de premières simulations ont été proposées par EADS Astrium (*European Aeronautic Defence and Space Company*) pour évaluer la performance de mesures lidars depuis l'espace sur des forêts denses. Pour ces simulations, Astrium utilise un modèle de forêt en trois couches : une couche supérieure à l'interface de l'atmosphère et de la végétation, une couche de couvert végétal et une couche de sol. L'énergie rétrodiffusée et reçue par le capteur est donnée par :

$$P_{reçue} = P_{émise} \times T_{atm}^2 \times R_{canop} \times T_{canop} \times T_{veg}^2 \times R_{sol} \qquad (1)$$

où : $P_{émise}$ est la puissance émise par le lidar, $T_{atm}$ est la transmission atmosphérique, $R_{canop}$ est la réflectance de la couche supérieure à l'interface de l'atmosphère et de la végétation, $T_{canop}$ est la transmittance de la couche supérieure, $T_{veg}$ est la transmittance de la couche de couvert végétal et $R_{sol}$ est la réflectance de la couche de sol. La présence des termes au carré est due à l'aller retour effectué par le signal entre le capteur et le sol.

Ce modèle présente l'avantage d'être générique et applicable à différents types de forêts mais des valeurs réalistes pour les différents paramètres restent à estimer, en particulier pour les forêts tropicales qui présentent les conditions les plus défavorables à la pénétration du signal lidar. Aussi, le modèle trois couches ne permet pas de bien décrire l'interaction entre le signal et la canopée, cette dernière n'étant pas une couche horizontale régulière et continue.

Une étape importante dans la spécification des paramètres techniques d'un lidar satellitaire dédié à la cartographie de forêts est donc d'affiner les estimations de la dynamique du signal dans différents contextes forestiers.

L'objectif de cette étude est de proposer une méthode de simulation de signaux à large empreinte à partir de données expérimentales acquises avec des capteurs aéroportés à petite empreinte dites « classiques ». Ces données aéroportées sont couramment acquises dans le cadre de projets de recherches, et il sera donc possible de simuler des signaux à large empreinte dans différents contextes forestiers.

## 2. Matériel

*2.1 Site d'étude*

Le site d'étude est la forêt de Haute-Bléone dans les Alpes de Haute-Provence (44.16°N, 6.32°W). C'est une forêt méditerranéenne de 108 hectares qui fut afforestée au XIX$^{ème}$ siècle avec du Pin noir d'Autriche (Pinus nigra ssp. nigra [Arn.]) dont la densité varie de 100 à plus de 2000 arbres par hectare. Cette forêt de montagne permet de fixer le sol composé de marnes noires particulièrement friables, et donc de lutter contre l'intense érosion.

*2.2 Données lidar*

Une acquisition de données lidar a été effectuée en avril 2007 avec un capteur aéroporté RIEGL LMS-Q560 enregistrant le signal retour complet. La hauteur de vol de 600 m et la divergence du faisceau laser ont produit une taille d'empreinte au sol de l'ordre de 30 cm. La vitesse de vol et la vitesse de répétition des mesures (111kHz) ont mené à un taux de cinq mesures par mètre carré. Le système utilise un laser proche-infrarouge de longueur d'onde 1.5 μm.

Le jeu de données comporte les signaux bruts (émis et reçus) numérisés à 1 GHz (15 cm), ainsi que le nuage de points en 3D (coordonnées X, Y et Z) issus du traitement de ces signaux par l'ajustement de n fonctions gaussiennes qui minimisent la somme des résidus entre le signal et le modèle gaussien (Chauve et al., 2009). Chaque gaussienne représente alors le pic d'énergie réfléchi par une cible et permet de détecter sa position. Ces données ponctuelles sont appelées échos discrets ou données multiéchos.

## 3. Méthode

L'objectif de l'étude est de simuler des signaux complets à large empreinte, par l'agrégation de données multiéchos à petite empreinte. Les données fournies par un prestataire de service peuvent contenir différents niveaux de détails sur les caractéristiques des échos. C'est pourquoi plusieurs modèles d'agrégation sont présentés dans ce document. Ces modèles reposent sur l'hypothèse selon laquelle le signal retour est la convolution de la distribution verticale des cibles par l'impulsion émise. Cette dernière est caractérisée par une

amplitude (puissance de l'impulsion émise) et une durée (largeur d'impulsion).

## *3.1. Traitements préliminaires*

### *3.1.1. Calcul du modèle numérique de terrain*

Certaines des méthodes de simulation présentées dans ce document sont en partie basées sur l'utilisation d'un modèle numérique de terrain (MNT) afin de distinguer les échos ayant été réfléchis sur le sol de ceux ayant été réfléchis sur la végétation. Deux MNT différents ont été utilisés dans cette étude et leur impact sur la précision des simulations a été étudié (voir section 3.7).

Lorsque qu'une classification des échos lidar peut être réalisée, il est possible de produire un MNT à haute résolution. Dans cette étude, le premier MNT est issu de la classification des échos selon une méthode présentée par Véga et al. (2011).

Le second MNT (appelé MNT imprécis) a été calculé en découpant la surface occupée par l'empreinte lidar en neuf cellules et en prenant l'écho d'altitude minimale dans chaque cellule. Les neuf points ont ensuite été interpolés linéairement sur l'ensemble de l'empreinte. Cette méthode se veut générique et applicable à tout jeu de données lidar. Sa simplicité peut cependant mener à des résultats imprécis, en particulier sur des zones caractérisées par une topographie accidentée.

### *3.1.2. Détermination des réflectances du sol et de la végétation*

Certaines des méthodes présentées dans ce document nécessitent de connaître les réflectances du sol et de la végétation à la longueur d'onde du laser. Pour cela, nous avons recherché les échos uniques, rencontrés lorsque l'impulsion émise a été réfléchie par une seule cible. Grâce à un MNT les échos uniques ont d'abord été identifiés comme ayant été réfléchis sur le sol ou sur la végétation (seuil de 1 m au-dessus du MNT). La réflectance a ensuite été calculée comme le ratio entre l'énergie reçue et l'énergie émise, elles-mêmes calculées comme les intégrales des signaux complets correspondants.

## *3.2. Calcul des signaux de référence*

Les signaux de référence pour cette étude ont été calculés par l'agrégation de signaux bruts à petite empreinte sur une surface de 30 m de diamètre. Pour le jeu de données utilisé dans cette étude, les cinq empreintes (30 cm de diamètre) par mètre carré ont été supposées fournir des données suffisamment représentatives de la structure forestière, comparé à cinq empreintes de 56 cm de diamètre (correspondant à une couverture complète de la zone étudiée).

L'influence de l'atmosphère n'a pas été prise en compte dans ces calculs, mais elle ne devrait apporter qu'un bruit supplémentaire pouvant être ajouté ultérieurement. Les signaux individuels ont été pondérés de manière à respecter la distribution gaussienne de l'énergie émise dans l'empreinte agrégée. Les intensités les plus fortes se trouvent au centre de l'empreinte et décroissent en s'en éloignant (voir Figure 1). La forme des signaux ainsi agrégés est donc supposée s'approcher de celle de vrais signaux à large empreinte. La fréquence d'échantillonnage du signal agrégé est conservée à 1 GHz (15 cm).

## *3.3. Cas X,Y,Z, Intensité*

Les signaux simulés ont été calculés par l'agrégation de données multiéchos sur une surface de 30 m de diamètre. Dans le cas le plus favorable, l'intensité de chaque écho est fournie en plus de sa position. Le signal complet est alors simulé en pondérant l'histogramme de la distribution verticale des échos par l'intensité des échos. Une

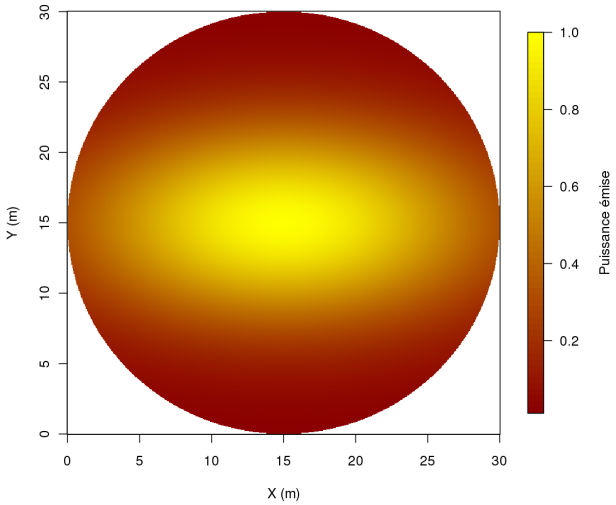



Figure 1. Distribution de l'énergie émise dans une empreinte de 30 m de diamètre.

pondération additionnelle a été calculée en fonction de la localisation de chaque échos par rapport au centre de la surface d'agrégation (30 m de diamètre), afin de simuler la distribution gaussienne de l'énergie dans l'empreinte (voir Figure 1). L'histogramme est calculé avec une résolution égale au pas de numérisation du signal complet. Il est ensuite convolué à une courbe gaussienne qui représente l'impulsion émise, avec un écart-type de 1.7 (largeur à mi-hauteur = 4 ns) et une amplitude de 1, puisque l'information d'intensité est déjà prise en compte dans l'histogramme pondéré.

## *3.4. Cas X,Y,Z, Temps*

Un simple histogramme de la distribution verticale des échos ne prends pas en compte l'atténuation du laser le long de son trajet dans la végétation. En effet, l'amplitude des échos situés plus profondément dans la forêt est généralement moins importante que celle des échos de surface. Afin de simuler des signaux sans disposer de l'amplitude des échos, il est donc nécessaire d'estimer l'atténuation du signal.

Dans le cas où l'horaire d'émission de l'impulsion émise est fourni pour chaque écho, il est possible d'identifier le nombre d'échos par impulsion émise, et donc leur ordre de réflexion en les classant selon l'altitude. Les échos réfléchis à une altitude supérieure à 3 m au-dessus du MNT ont été considérés comme des échos réfléchis sur la végétation, et les autres comme des échos réfléchis sur le sol. Ce seuil de 3 mètres est ajustable, mais doit être suffisamment élevé pour éviter les erreurs dans le cas où le MNT utilisé serait imprécis.

L'atténuation de l'amplitude des échos a été estimée en fonction de leur ordre de réflexion par les formules suivantes :

$$PVEG_n = R_{VEG} \cdot (1 - R_{VEG})^{2\cdot(n-1)} \qquad (2)$$

où $PVEG_n$ est la pondération associée à l'écho de rang n réfléchi sur la végétation, $R_{VEG}$ est la réflectance de la végétation à la longueur d'onde du laser, $(1 - R_{VEG})$ représentant la transmittance de la végétation. Le facteur 2 du terme en exposant se réfère à l'atténuation de l'onde lumineuse dans la végétation à l'aller et au retour, pour les échos de rang n.

L'atténuation des échos réfléchis par le sol vers le capteur est donnée par :

$$PSOL_n = R_{SOL} \cdot (1 - R_{VEG})^{2\cdot(n-1)} \qquad (3)$$

où $PSOL_n$ est la pondération associée à l'écho de rang n réfléchi sur le sol et $R_{SOL}$ est la réflectance du sol à la longueur d'onde du laser.
L'histogramme de la distribution verticale des échos est ensuite pondéré par $PVEG_n$ pour les échos réfléchis sur la végétation et par $PSOL_n$ pour les échos réfléchis sur le sol.

Une pondération additionnelle a été calculée en fonction de la localisation de chaque échos par rapport au centre de la surface d'agrégation (30 m de diamètre), afin de simuler la distribution gaussienne de l'énergie dans l'empreinte (voir Figure 1). L'histogramme pondéré est enfin convolué par l'impulsion émise, de forme gaussienne avec une amplitude et une durée correspondant à l'impulsion émise par le capteur.

### *3.5. Cas X,Y,Z*

Dans le cas le plus défavorable, où seule la position de chaque écho est fournie, il n'est plus possible d'estimer l'atténuation du signal selon son ordre de réflexion. L'atténuation est alors considérée ici comme progressive suivant la pénétration du laser dans la végétation, afin que les échos ayant pénétré davantage aient une intensité plus faible.

Le taux de pénétration du laser dans la végétation est donné par le taux d'échos réfléchis sur le sol ($T_s$ = nombre d'échos sol / nombre total d'échos). Le taux d'atténuation correspondant est rapporté à un taux moyen par mètre donné par : $T_n = \frac{1-T_s}{Z_{max}-Z_{min}}$ . Enfin, le pseudo champ intensité donne la pondération à appliquer à histogramme de la distribution verticale des échos (Pn=1 pour Zn=Zmax *i.e.* pas d'atténuation sur la canopée ; Pn=Ts pour Zn=Zmin *i.e.* atténuation maximale sur le sol) :

$$P_n = 1 - \frac{1-T_s}{Z_{max}-Z_{min}} \cdot (Z_{max} - Z_n) \qquad (4)$$

Ce taux d'atténuation est calculé pour chaque écho en fonction de son altitude par rapport au MNT. Il est ensuite pondéré par la réflectance de la végétation s'il s'agit d'un écho réfléchi sur la végétation et par la réflectance du sol s'il s'agit d'un écho réfléchi sur le sol. La distinction entre ces deux types d'échos étant réalisée grâce à au MNT. Une pondération additionnelle a été calculée en fonction de la localisation de chaque échos par rapport au centre de la surface d'agrégation (30 m de diamètre), afin de simuler la distribution gaussienne de l'énergie dans l'empreinte (voir Figure 1). L'histogramme pondéré est enfin convolué par l'impulsion émise, de forme gaussienne avec une amplitude et une durée correspondant à l'impulsion émise par le capteur.

### *3.6. Performance des différents modèles*

Les signaux simulés pour les cas 3.2, 3.3 et 3.4 sont comparés aux signaux de référence calculés en 3.1. Nous avons choisi d'exprimer la cohérence des simulations en fonction de deux critères.
Le premier est la corrélation entre les deux courbes, donnée par le coefficient de corrélation de Bravais-Pearson :

$$\rho = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{N} (y_i - \bar{y})^2}} \qquad (5)$$

où x est la série de valeurs de longueur N du signal simulé et y celle du signal de référence.
Cet indicateur permet d'évaluer la concordance des variations entre les deux courbes, mais ne tient pas compte de la différence d'amplitude.

Pour évaluer la différence entre les amplitudes des deux courbes, nous avons choisi de calculer le rapport entre la moyenne des erreurs entre les deux signaux et la moyenne des valeurs du signal de référence :

$$\mu = \frac{\sum_{i=1}^{N} x_i - y_i}{\sum_{i=1}^{N} y_i} \qquad (6)$$

où x est la série de valeurs de longueur N du signal simulé est y celle du signal de référence.

Les résultats sont présentés sur quatre placettes forestières, sélectionnées pour la diversité de leurs caractéristiques de densité et de hauteurs d'arbres : densité élevée avec de grands arbres (P31), densité élevée avec de petits arbres (P12), faible densité avec de grands arbres (P166), faible densité avec de petits arbres (P257).

### *3.7. Sensibilité des modèles aux paramètres d'entrée*

#### *3.7.1. Influence de la qualité du MNT sur l'estimation du signal*

Les méthodes de correction 3.3 et 3.4 sont en partie basées sur l'utilisation d'un modèle numérique de terrain. Nous avons évalué l'impact de la précision du MNT sur la qualité des signaux simulés.

Dans les forêts tropicales, où il est attendu que le taux d'échos sol soit faible (de l'ordre de 2%), les méthodes de classification existantes risquent de donner des résultats imprécis. Pour cette raison, nous avons évalué la précision des estimations de signaux en utilisant un MNT volontairement imprécis dont la méthode de calcul est présentée dans la section 3.1.1. Les résultats ont été comparés à ceux obtenus en utilisant un MNT à haute résolution.

Dans le modèle XYZ, l'utilisation du MNT permet de distinguer les échos sol des échos végétation (et donc de fixer leur réflectance respectives), mais il permet aussi de calculer le taux d'échos sols, lui-même utilisé dans le calcul de la pondération (voir équation 4). Afin d'étudier précisément la sensibilité du modèle X,Y,Z à une erreur sur le taux de points sol, nous avons effectué des simulations en utilisant des taux de points sols volontairement erronés. Ces simulations ont été effectuées en utilisant le MNT à haute résolution.

#### *3.7.2. Sensibilité des simulations à la précision des estimations de réflectances*

Les méthodes de correction 3.3 et 3.4 sont en partie basées sur des valeurs de réflectances du sol et de la végétation à la longueur d'onde du laser. Ces valeurs ont été approximées selon une méthode présentée dans la section 3.1.2, mais comportent potentiellement des imprécisions.

Nous avons donc étudié l'impact de valeurs de réflectances imprécises sur la qualité des simulations de signaux. Pour cela, nous avons tracé la variation des indicateurs de qualité des estimations ($\mu$ et $\rho$, voir section 3.6) en fonction de valeurs volontairement imprécises de la réflectance du sol d'une part et de la réflectance de la végétation d'autre part.

Les courbes présentées ont été obtenues en utilisant le modèle X,Y,Z avec le MNT à haute résolution.

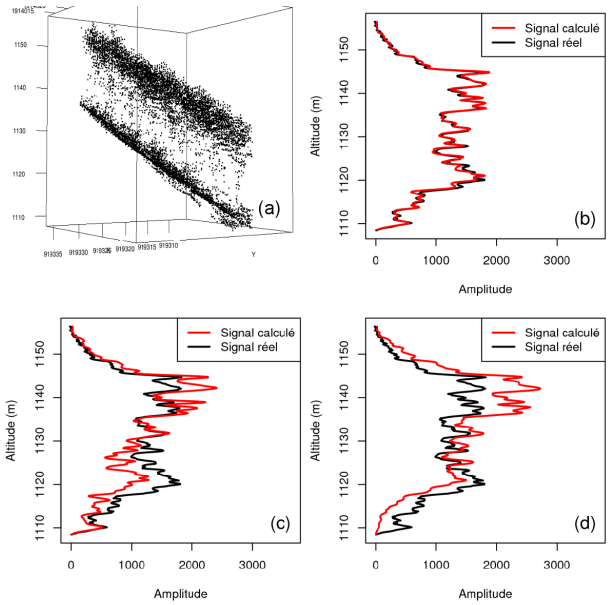



Figure 2. Simulation des signaux pour la placette 31 (grands arbres, forte densité). Les signaux de référence apparaissent en rouge et les signaux simulés en noir. (a) Vue de la distribution des échos en trois dimensions. (b) Modèle X,Y,Z,Intensité (voir section 3.3). (c) Modèle X,Y,Z,Temps (voir section 3.4). (d) Modèle X,Y,Z (voir section 3.5). Les signaux (c) et (d) ont été simulés en utilisant le MNT à haute résolution.

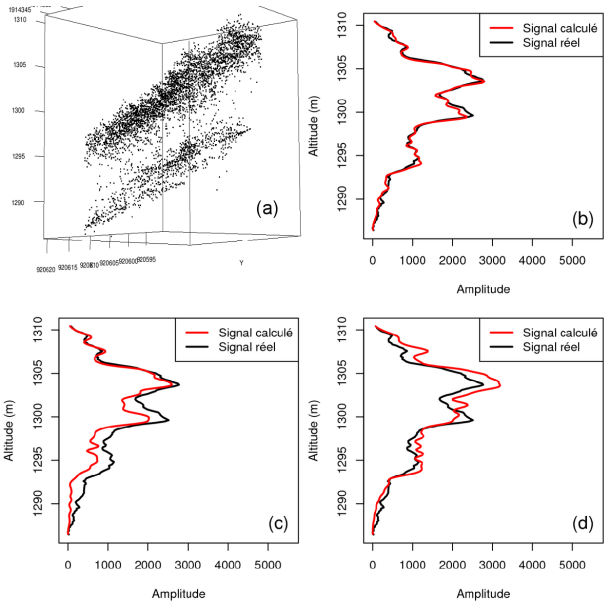



Figure 4. Simulation des signaux pour la placette 12 (petits arbres, forte densité). Les signaux de référence apparaissent en rouge et les signaux simulés en noir. (a) Vue de la distribution des échos en trois dimensions. (b) Modèle X,Y,Z,Intensité (voir section 3.3). (c) Modèle X,Y,Z,Temps (voir section 3.4). (d) Modèle X,Y,Z (voir section 3.5). Les signaux (c) et (d) ont été simulés en utilisant le MNT à haute résolution.

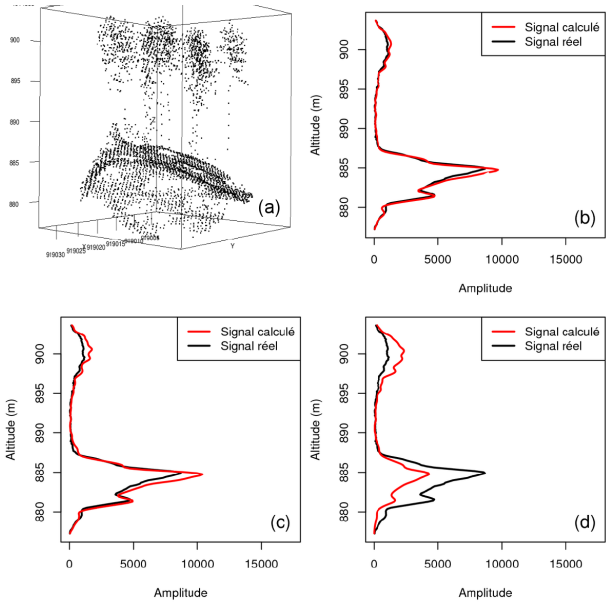



Figure 3. Simulation des signaux pour la placette 166 (grand arbres, faible densité). Les signaux de référence apparaissent en rouge et les signaux simulés en noir. (a) Vue de la distribution des échos en trois dimensions. (b) Modèle X,Y,Z,Intensité (voir section 3.3). (c) Modèle X,Y,Z,Temps (voir section 3.4). (d) Modèle X,Y,Z (voir section 3.5). Les signaux (c) et (d) ont été simulés en utilisant le MNT à haute résolution.

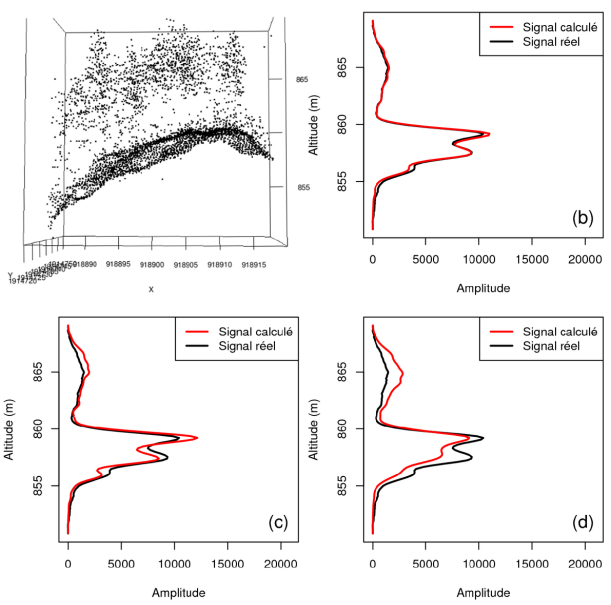



Figure 5. Simulation des signaux pour la placette 257 (petits arbres, faible densité). Les signaux de référence apparaissent en rouge et les signaux simulés en noir. (a) Vue de la distribution des échos en trois dimensions. (b) Modèle X,Y,Z,Intensité (voir section 3.3). (c) Modèle X,Y,Z,Temps (voir section 3.4). (d) Modèle X,Y,Z (voir section 3.5). Les signaux (c) et (d) ont été simulés en utilisant le MNT à haute résolution.

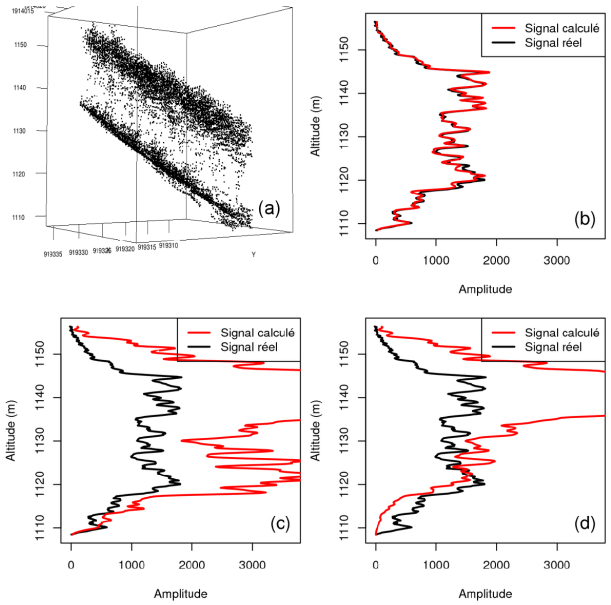



Figure 6. Simulation des signaux pour la placette 31 (grands arbres, forte densité). Les signaux de référence apparaissent en rouge et les signaux simulés en noir. (a) Vue de la distribution des échos en trois dimensions. (b) Modèle X,Y,Z,Intensité (voir section 3.3). (c) Modèle X,Y,Z,Temps (voir section 3.4). (d) Modèle X,Y,Z (voir section 3.5). Les signaux (c) et (d) ont été simulés en utilisant le MNT imprécis.

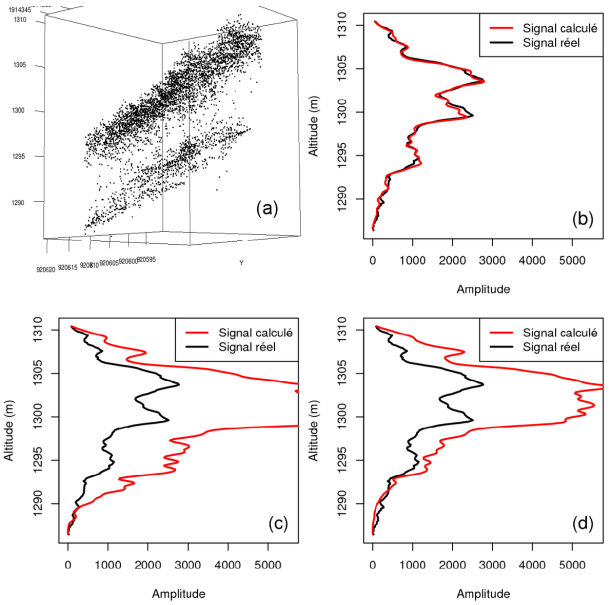



Figure 8. Simulation des signaux pour la placette 12 (petits arbres, forte densité). Les signaux de référence apparaissent en rouge et les signaux simulés en noir. (a) Vue de la distribution des échos en trois dimensions. (b) Modèle X,Y,Z,Intensité (voir section 3.3). (c) Modèle X,Y,Z,Temps (voir section 3.4). (d) Modèle X,Y,Z (voir section 3.5). Les signaux (c) et (d) ont été simulés en utilisant le MNT imprécis.

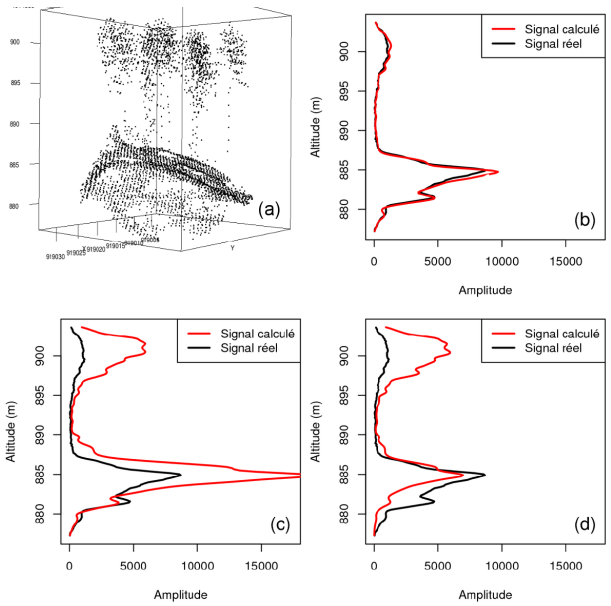



Figure 7. Simulation des signaux pour la placette 166 (grand arbres, faible densité). Les signaux de référence apparaissent en rouge et les signaux simulés en noir. (a) Vue de la distribution des échos en trois dimensions. (b) Modèle X,Y,Z,Intensité (voir section 3.3). (c) Modèle X,Y,Z,Temps (voir section 3.4). (d) Modèle X,Y,Z (voir section 3.5). Les signaux (c) et (d) ont été simulés en utilisant le MNT imprécis.

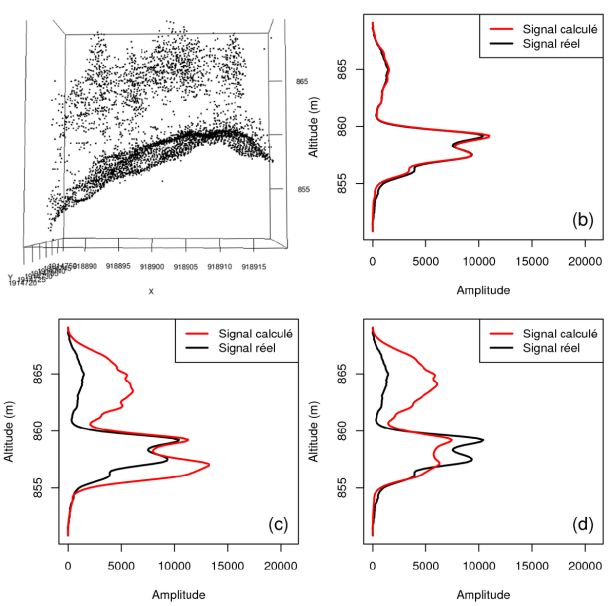



Figure 9. Simulation des signaux pour la placette 257 (petits arbres, faible densité). Les signaux de référence apparaissent en rouge et les signaux simulés en noir. (a) Vue de la distribution des échos en trois dimensions. (b) Modèle X,Y,Z,Intensité (voir section 3.3). (c) Modèle X,Y,Z,Temps (voir section 3.4). (d) Modèle X,Y,Z (voir section 3.5). Les signaux (c) et (d) ont été simulés en utilisant le MNT imprécis.

## 4. Résultats et discussion

### *4.1. Traitements préliminaires*

#### *4.1.1. MNT*

Le MNT à haute résolution a été calculé avec une erreur (RMSE) de 57 cm sous le couvert forestier (Véga et al., 2011).

Le second MNT utilisé dans cette étude (appelé MNT imprécis) a été calculé avec une RMSE de 1.50 m.

#### *4.1.2. Détermination des réflectance du sol et de la végétation*

Pour le sol, une réflectance moyenne de 0.45 a été calculée avec un écart-type de 0.09, sur 22112 valeurs. La réflectance moyenne de la végétation est de 0.24 avec un écart-type de 0.05, calculée sur 40832 valeurs.

### *4.2. Signaux simulés*

#### *4.2.1. Performance des différents modèles*

Les meilleurs résultats ont été obtenus en utilisant le MNT à haute résolution. Les figures 2, 3, 4 et 5 présentent les résultats des simulations pour les trois modèles développés sur quatre placettes de caractéristiques différentes. Les indicateurs de correspondance entre les signaux simulés et les signaux de référence sont donnés dans le Tableau 1.

On observe que les signaux simulés avec le modèle X,Y,Z,Intensité en utilisant le MNT à haute résolution présentent les plus fortes ressemblances avec les signaux de références. La performance de ce modèle est confortée par des corrélations supérieures à 0.99 et des différences d'amplitudes comprises entre -1% et 6%.

Le modèle X,Y,Z,Temps est un peu moins performant que le précédent, avec toutefois des valeurs de corrélations supérieures à 0.87 et des amplitudes parfois surestimées et parfois sous-estimées, à 12% près en moyenne.

Tableau 1. Évaluation de la performance des différents modèles, sur quatre empreintes : P31 (grands arbres, forte densité), P12 (petits arbres, forte densité), P166 (grands arbres, faible densité), P257 (petits arbres, faible densité). Le calcul des différents MNT est décrit dans la section 3.1.1. ρ est le coefficient de corrélation et μ le rapport d'amplitude.

| | | MNT haute résolution | | MNT imprécis | |
|---|---|---|---|---|---|
| | | ρ | μ | ρ | μ |
| *Cas X,Y,Z, Intensité* | P31 | 0.99 | -0.01 | 0.99 | -0.01 |
| | P12 | 0.99 | -0.01 | 0.99 | -0.01 |
| | P166 | 0.99 | 0.06 | 0.99 | 0.06 |
| | P257 | 1.00 | 0.00 | 1.00 | 0.00 |
| *Cas X,Y,Z, Temps* | P31 | 0.87 | -0.07 | 0.78 | 2.06 |
| | P12 | 0.97 | -0.22 | 0.93 | 1.71 |
| | P166 | 0.98 | 0.17 | 0.89 | 1.30 |
| | P257 | 0.98 | 0.02 | 0.83 | 1 |
| *Cas X,Y,Z* | P31 | 0.87 | 0.13 | 0.64 | 1.12 |
| | P12 | 0.97 | 0.14 | 0.96 | 1.08 |
| | P166 | 0.83 | -0.30 | 0.90 | 0.34 |
| | P257 | 0.95 | -0.03 | 0.71 | 0.4 |

Le modèle X,Y,Z comporte les plus fortes erreurs. Les valeurs de corrélations sont supérieures à 0.83, et la différence d'amplitude des signaux simulés par rapport aux signaux de référence est de 15% en moyenne. Ce modèle ne disposant que des informations minimales sur la position des échos, il est donc difficile d'estimer précisément l'amplitude de ceux-ci.

#### *4.2.1. Influence de la qualité du MNT*

Les indicateurs de performances des différents modèles en utilisant le MNT imprécis sont donnés dans le Tableau 1. Les figures 6, 7, 8 et 9 présentent les signaux correspondants.

Le MNT n'entre pas en compte dans le modèle X,Y,Z,Intensité.

Pour le modèle X,Y,Z,Temps, on observe une perte de corrélation de 9% en moyenne par rapport celle des signaux simulés en utilisant le MNT à haute résolution. Une forte sur-estimation de l'amplitude apparaît (150% en moyenne).

Malgré des corrélations inférieures à celles obtenues avec le modèle X,Y,Z,Temps (0.80 contre 0.85 en moyenne en utilisant un MNT imprécis), le modèle X,Y,Z fournit des signaux dont les amplitudes sont plus proches de celles des signaux de référence (74% contre 150% en moyenne pour le modèle X,Y,Z,Temps). Il semble donc plus judicieux d'utiliser le modèle X,Y,Z dans le cas où le MNT utilisé est imprécis.

La Figure 10 montre l'influence d'une imprécision d'estimation du taux d'échos sol sur la qualité des simulations. Le rapport d'amplitude varie de -4% à 25% pour des erreurs sur l'estimation du taux de points sol comprises entre -100% et +100%. De la même manière, le coefficient de corrélation entre les signaux varie de 0.78 à 0.86. Bien que la sensibilité du modèle X,Y,Z soit sensible aux erreurs d'estimation du taux de point sol, les variations observée sont moins forte que dans le cas d'erreurs sur l'estimation de valeurs de réflectances du sol et de la végétation (voir section suivante).

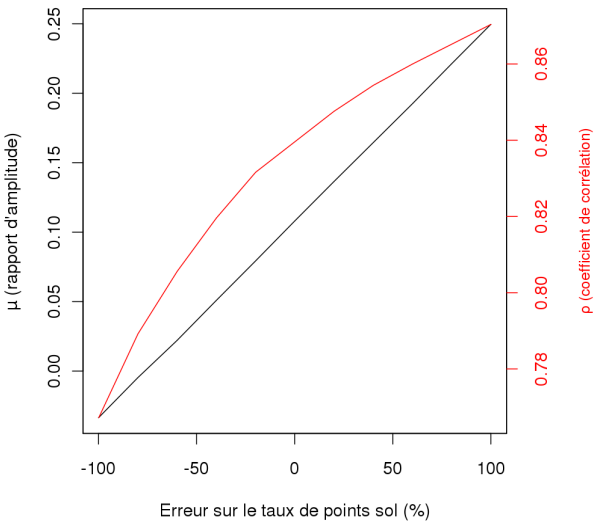



Figure 10. Sensibilité des simulations à la précision des estimations de la réflectance du sol. Les courbes représentent les moyennes des valeurs calculées sur 25 empreintes.

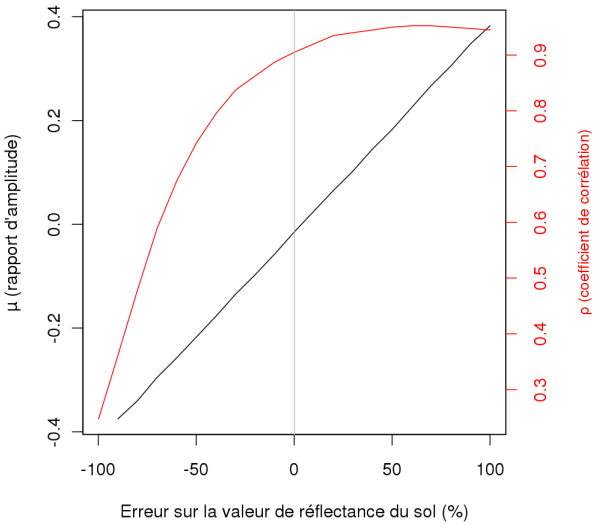



Figure 11. Sensibilité des simulations à la précision des estimations de la réflectance du sol. Les courbes représentent les moyennes des valeurs calculées sur 25 empreintes. La réflectance de référence est de 0.45 pour le sol.

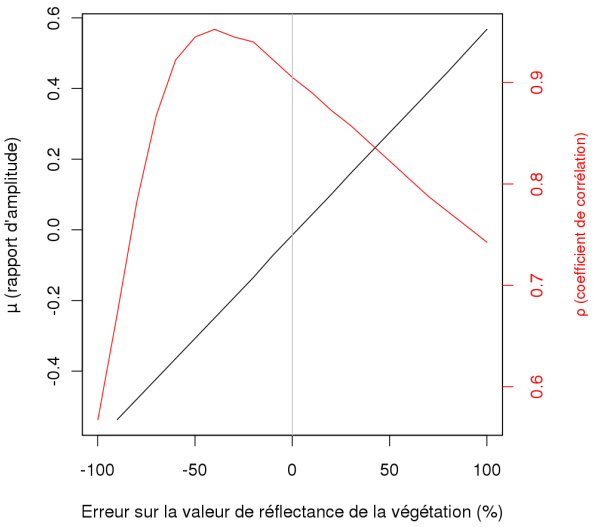



Figure 12. Sensibilité des simulations à la précision des estimations de la réflectance de la végétation. Les courbes représentent les moyennes des valeurs calculées sur 25 empreintes. La réflectance de référence est de 0.24 pour la végétation.

### *4.2. Sensibilité aux estimations de réflectances*

Les résultats des simulations effectuées avec des valeurs imprécises de la réflectance de la végétation et du sol sont présentés respectivement dans les figures 11 et 12. Elles ont été calculées avec le modèle X,Y,Z en utilisant le MNT à haute résolution. Les valeurs représentées sont les moyennes respectives des coefficients de corrélation (ρ) et des rapports d'amplitude (μ) calculés sur 25 empreintes différentes, pour différents niveaux d'erreur sur la valeur de réflectance.

On observe que le rapport d'amplitude de nul pour des erreurs nulles sur les valeurs de réflectance de sol (figures 11) et de la végétation (figure 12). L'amplitude des signaux simulés est sur-estimée lorsque les valeurs de réflectance sont sur-estimées, et inversement. Il apparaît donc que les valeurs de réflectances déterminées dans la section 4.1.2 sont cohérentes (0.24 pour la végétation et 0.45 pour le sol).

On observe cependant que les valeurs du coefficient de corrélation atteignent leur maximum pour une sur-estimation de la réflectance du sol de 50% (figure 11) et pour une sous-estimation de la réflectance de la végétation de 50% (figure 12). Ces résultats sont à nuancer par le fait que, dans les deux cas, des coefficients de corrélation supérieurs à 0.90 sont obtenus avec les valeurs correctes de réflectances.

## 5. Conclusion

L'objectif de ces travaux était de proposer des méthodes de simulation de signaux à larges empreintes, permettant de par la suite d'estimer la dynamique du signal dans différents contextes forestiers. L'approche retenue était de simuler ces signaux à partir de données réelles multi-échos (issues de capteurs lidars topographiques) contenant différents niveaux d'informations.

L'étude s'est focalisée sur les interactions entre le laser et la végétation, l'influence de l'atmosphère sur le signal pouvant être prise en compte a posteriori.

Deux des trois modèles développés requièrent l'utilisation d'un modèle numérique de terrain. Les résultats obtenus avec un MNT à haute résolution ont montré démontré la validité des modèles, puisque des corrélations supérieures à 0.83 (dans le pire des cas) ont été obtenues. Cependant, plus les informations sur les échos étaient succinctes, plus il a été difficile d'estimer leur amplitude avec précision.

Les méthodes présentées dans ce document visent à être améliorées. Elles sont destinées à être appliquées sur différents jeux de données avec pour objectif final d'apporter des informations permettant de dimensionner correctement l'énergie à émettre depuis un satellite afin d'obtenir des signaux exploitables sur différents contextes forestiers.

## Bibliographie

A. Chauve, C. Vega, S. Durrieu, F. Bretar, T. Allouis, M. Pierrot-Deseilligny and W. Puech, “Advanced fullwaveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest”, Int. J. Remote Sens., vol. 30:19, pp. 5211–5228, 2009.

C. Véga, S. Durrieu and T. Allouis, “A Lidar filtering algorithm for terrain modeling in complex forested environments”, *Comput. Geosci.,* accepted, 2011.

# A New Method for Incorporating Hillslope Effects to Improve Canopy-Height Estimates From Large-Footprint LIDAR Waveforms

Tristan Allouis, *Student Member, IEEE*, Sylvie Durrieu, and Pierre Couteron

***Abstract*—Forest structure variables, such as the canopy height, are of central interest for the quantification of ecosystem functions and the assessment of biomass levels. The objective of this letter is to propose a new method for ridding canopy-height estimates from the influence of the hillslope within large-footprint (light detection and ranging) LIDAR waveforms. The method is based on modeling (using two generalized Gaussian functions) and the fitting of canopy and ground components to large-footprint (30 m) waveforms. The canopy heights were estimated for 27 waveforms: A root-mean-square error of 3.3 m was obtained using a high-resolution digital terrain model (DTM) to estimate the ground component (4.3 m using the 80-m-resolution Shuttle Radar Topography Mission DTM) and 3.5 m when self-estimating the ground component (hillslope) based on the large-footprint waveform. This approach opens new possibilities for waveform decomposition for natural resources and topography assessments based on large-footprint LIDAR waveforms in forest environments.**

***Index Terms*—Digital terrain model (DTM), forest structure, Gaussian, high resolution, nonlinear least squares (NLS) fitting, signal processing, tree height.**

## I. Introduction

In THE context of climate-change mitigation, monitoring and reporting technologies are required for the sustainable management of forest resources. The structural variables that characterize the 3-D configuration of forest biomass are fundamental for quantifying forest functions and assessing biomass levels. Among those variables, canopy height is known to correlate with stand production [1].

During the course of six years, the Geoscience Laser Altimeter System [GLAS on Ice, Cloud, and land Elevation Satellite (ICESat)] has provided useful data for the assessment of canopy height, and these data have been recently combined with Moderate Resolution Imaging Spectroradiometer images to generate a global forest height map [2]. The canopy height is typically retrieved from light detection and ranging (LIDAR) waveforms based on the distance between the canopy and ground echoes. The latter is identified as the last or second to the last of up to six Gaussian functions that are fitted on the raw waveform [3], [4]. However, the large LIDAR footprint size of GLAS (65 m) makes it particularly difficult to extract accurate canopy-height measurements on uneven or sloped terrain due to mixed ground and vegetation echoes [4]. A commonly used approach is to decrease the footprint size of spaceborne LIDAR for the recovery of vegetation structures, such as the 50-m footprint of the future IceSat-II mission [5], the 25–30-m footprint of the proposed Deformation, Ecosystem Structure and Dynamics of Ice system [6], and the 10–20-m footprint of the proposal LIDAR for Earth And Forest (LEAF) mission [7]. However, even with these smaller footprint sizes, the data acquired by these future spaceborne systems will still be affected by mixed echoes on sloped terrain.

A method that uses ancillary slope data to correct GLAS waveforms for ground-slope influences has been recently published [8]. In this method, the vegetation height was retrieved based on the waveform extent minus the ground elevation extent within the footprint size, which required *a priori* knowledge of the topography. Our objectives and approach are similar to those of Lee *et al.* [8], although our method uses a waveform-modeling step that is designed to self-estimate the hillslope based on the LIDAR waveform. Our model is based on two assumptions: 1) Both the shape of the ground and the vegetation components of the backscattered waveform can be approximated with a generalized Gaussian function, and (2) in regular tree stands, the canopy (upper layer) has a consistent height within the footprint. Two generalized Gaussian functions were fitted to the LIDAR waveform as initially proposed by Allouis *et al.* [9] for the measurement of water depth (i.e., the first Gaussian modeled the surface echo, and the second Gaussian modeled the ground echo).

## II. Materials

### A. Study Area

Twenty-seven reference plots were defined according to the French national forest inventory protocol (three circular nested plots with 6-, 9-, and 15-m radii) and were measured ($X$, $Y$, and $Z$ coordinates; tree diameter at breast height (DBH), total tree height and crown base height for each tree, and plot hillslope) in December of 2007 and November of 2008 in the Haute-Bléone forest of the southern French Alps (44.16° N, 6.32° W). This 108-ha badland area has a complex topography and was

Manuscript received September 16, 2011; revised October 17, 2011 and revised December 6, 2011; accepted December 7, 2011. This work was supported in part by Centre National d'Études Spatiales and in part by the French Languedoc-Roussillon region.

T. Allouis and S. Durrieu are with the National Research Institute of Science and Technology for Environment and Agriculture, Territories, Environment, Remote Sensing and Spatial Information Joint Research Unit, 34196 Montpellier, France (e-mail: tristan.allouis@teledetection.fr; sylvie.durrieu@teledetection.fr).

P. Couteron is with the Institut de Recherche pour le Développement, botAnique et bioinforMatique de l'Architecture des Plantes Joint Research Unit, 34398 Montpellier, France (e-mail: pierre.couteron@ird.fr).

Digital Object Identifier 10.1109/LGRS.2011.2179635

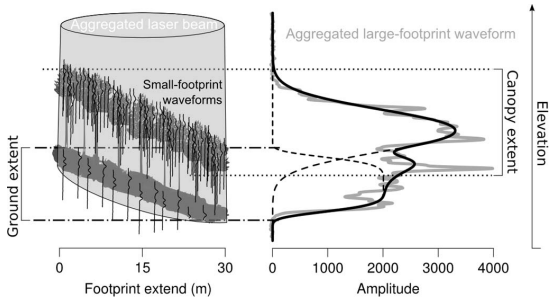



Fig. 1. (Left) Small-footprint waveforms were accumulated on a 30-m footprint. (Right) (Solid black line) Waveform model comprised (dashed black lines) two generalized Gaussian functions and was fitted on (gray) the cumulative waveform. (Top) First Gaussian function represents the canopy return, and (bottom) the second function represents the ground return. The dotted lines indicate the canopy and the ground extent.

afforested in the late nineteenth century with black pine (*Pinus nigra ssp. nigra [Arn.]*) to prevent intense soil erosion. The site is particularly suited for assessing LIDAR efficiency for various ground conditions and stand heights. The slopes of the reference plots range from 4° to 45°, and the tree heights range from 3.6 to 22.1 m. The mean standard deviation of the tree heights in each plot is 3.5 m, and the plot densities vary from 100 to 1075 stems/ha.

### B. LIDAR Data

The data were acquired in April of 2007 using an airborne RIEGL LMS-Q560 system. This sensor is a small-footprint full-waveform laser scanner that operates at a pulse rate of 111 kHz (see Wagner *et al.* [10] for the system design and specifications). The flight altitude was approximately 600 m above ground level, which resulted in a footprint diameter of approximately 0.30 m and a measurement density of 5 shots/m$^2$.

The data set was composed of raw full-waveform signals (1-D intensity profiles) of both emitted and received energies that were digitized at 1 GHz (15 cm). The georeferenced 3-D point cloud generated from the result of fitting Gaussian functions over the received waveforms [11] was also provided.

## III. Methods

### A. Simulation of Large-Footprint LIDAR Waveforms

The raw small-footprint waveforms were georeferenced in a global coordinate system using the GPS position of the plane and the LIDAR scan angle. Waveforms that were located within the reference plot boundaries were then summed to obtain a cumulative large-footprint waveform (Fig. 1). The corresponding emitted waveforms were also summed to produce a cumulative emitted waveform.

Cumulative waveforms were expected to be similar to real large-footprint waveforms of spaceborne LIDARs, except that the atmospheric influences (additional noise) and the Gaussian distribution of energy within the footprint were not taken into account. Five footprints (30 cm in diameter) per square meter were considered to provide the same representative measurements of the forest structure as five footprints of 56-cm diameter (providing a 100% coverage of the 30-m plots). In addition, the digitization rate of the small-footprint waveforms provided cumulative waveforms with the same digitization rate as large-footprint waveforms (e.g., 1 GHz for GLAS).

### B. Approximation of Large-Footprint Waveforms Using the Sum of Two Gaussian Functions

The curve resulting from the interaction between the emitted laser pulse (Gaussian shape) and a sloped terrain is a flattened Gaussian curve where the width is dependent on the hillslope and the footprint size. The difference in elevation between the uphill and downhill ground sections within large footprints is called the "ground extent" (Fig. 1) and is given by the following equation:

$$GroundExtent\ (\mathrm{m}) = \tan\left(GroundSlope\ (^\circ)\right) * FootprintDiameter\ (\mathrm{m}). \quad (1)$$

To model the ground component of a large-footprint waveform, we convolved a series of unit impulse functions of lengths that correspond to the ground vertical extent [(1)] using a Gaussian function that corresponds to the cumulative emitted waveform. The resulting curve can be modeled using a generalized Gaussian function [11], which incorporates a parameter that flattens the traditional Gaussian shape

$$f(t) = A \cdot \exp\left(-\frac{|t-\mu|^{\alpha^2}}{2\sigma^2}\right) \quad (2)$$

where $A$ is the scale factor (amplitude), $\mu$ is the mean (time position), $\sigma$ is the standard deviation (width), and $\alpha$ is the flattening parameter ($\alpha > \sqrt{2}$).

The canopy waveform component is affected by the decrease in the intensity of the laser as it passes through the vegetation, which is, therefore, assumed to produce an asymmetrical waveform component. However, we modeled the vegetation backscatter using a generalized Gaussian function for the sake of simplicity.

The cumulative waveform was accordingly modeled as the sum of two generalized Gaussian functions. The first function corresponded to the canopy component, and the second function corresponded to the ground component (Fig. 1).

### C. Gaussian Parameter Fitting

A nonlinear least squares (NLS) algorithm [12] was used to fit the sum of the two Gaussian functions on the 27 cumulative waveforms. The objective of this fitting process was to find the model parameters ($A_{\mathrm{canopy}}, \mu_{\mathrm{canopy}}, \sigma_{\mathrm{canopy}}, \alpha_{\mathrm{canopy}}, A_{\mathrm{ground}}, \mu_{\mathrm{ground}}, \sigma_{\mathrm{ground}}$, and $\alpha_{\mathrm{ground}}$) that minimize the sum of the residuals given by the following equation: $Waveform(t) - (f_{\mathrm{canopy}}(t) + f_{\mathrm{ground}}(t))$.

The NLS algorithm was first performed using *a priori* knowledge of the width ($\sigma_{\mathrm{ground}}$) and the flattening ($\alpha_{\mathrm{ground}}$) parameters of the ground component model. These parameters were retrieved from Section III-B using an ancillary digital terrain model (DTM). The method was tested using a 1-m-pixel DTM with a vertical accuracy of tens of centimeters that was derived from the small-footprint LIDAR

point cloud [11]. The method was also tested using an 80-m-pixel digital elevation model (DEM) that was generated by the Shuttle Radar Topography Mission (SRTM). The SRTM DEM vertical resolution is less than 10 m [13], but it has been widely used for processing spaceborne LIDAR signals [8], [14]. The NLS algorithm was used to retrieve the $A_{\mathrm{canopy}}, \mu_{\mathrm{canopy}}, \sigma_{\mathrm{canopy}}, \alpha_{\mathrm{canopy}}, A_{\mathrm{ground}}$, and $\mu_{\mathrm{ground}}$ parameters, whereas ($\sigma_{\mathrm{ground}}$ and $\alpha_{\mathrm{ground}}$) were fixed with a tolerance of $\pm 10\%$ that accounted for the possibility of inaccurate hillslope estimation.

The NLS algorithm was also performed without any prior knowledge of the ground topography. In this case, all Gaussian parameters $(A_{\mathrm{canopy}}, \mu_{\mathrm{canopy}}, \sigma_{\mathrm{canopy}}, \alpha_{\mathrm{canopy}}, A_{\mathrm{ground}}, \mu_{\mathrm{ground}}, \sigma_{\mathrm{ground}}, \text{and } \alpha_{\mathrm{ground}})$ were retrieved. The local slope, defined here as the "self-estimated slope," was estimated based on the value of the $\sigma_{\mathrm{ground}}$ parameter at the end of the NLS process. The linear relationship between $\sigma_{\mathrm{ground}}$ and the hillslope was previously retrieved using simulations based on the convolution method presented in Section III-B.

### D. Computing the Canopy Height

The elevation of the canopy top was retrieved as the position where the waveform exceeded a given noise threshold (i.e., the waveform start) [4]. The threshold was set to 4.5 times the maximum noise level using a generate-and-test method. Thus, the retrieved canopy top corresponded to the highest tree-top elevation, which is typically located within the uphill portion of a large footprint if we assume that the trees are consistent in height.

To assess the canopy height, the tree-top elevation must be compared to the ground elevation below the trees. However, only the mean ground position within the large-footprint area was retrieved (i.e., from the $\mu_{\mathrm{ground}}$ parameter). Consequently, the difference between $\mu_{\mathrm{ground}}$ and the ground elevation below the highest tree top was assumed to be equal to half of the ground extent.

Finally, the canopy-top elevation was corrected by half of the emitted pulse stretch (i.e., a distance corresponding to the full-width at half-maximum (FWHM) of the emitted pulse). The correction of the canopy-top position is given by the following equation:

$$CanopyTop_{\mathrm{elev}} = WaveformStart - FWHM/2 - GroundExtent/2 \quad (3)$$

where $CanopyTop_{\mathrm{elev}}$, $WaveformStart$, $FWHM$, and $GroundExtent$ are given in meters. $GroundExtent$ is given by (1) using either an ancillary DTM or a self-estimate of the hillslope based on the large-footprint waveform (see Section III-C).

Finally, the canopy height is given by the following equation:

$$CanopyHeight = CanopyTop_{\mathrm{elev}} - \mu_{\mathrm{ground}}. \quad (4)$$

### E. Comparison of LIDAR-Derived Canopy Heights With Reference Data

Considering that canopy trees are consistent in height within the footprint, we compared the LIDAR-derived measurements with the dominant height of the field-measured trees. The dominant height corresponded to the mean height of the seven largest trees (i.e., trees with the largest DBH) per 0.07-ha plot [16].

The accuracy of the canopy-height estimates was assessed using the measurement mean error (bias), standard deviation (dispersion), root-mean-square error (RMSE), and the mean of absolute values of all percentage errors (MAPE).

The results from our method were compared to those obtained using the Lee *et al.* [8] method with the same data set. Identical noise thresholds and both the high-resolution LIDAR DTM and the SRTM DEM were used.

The performance of our method was also assessed using the self-estimated hillslope.

### F. Performance of Slope Self-Estimation in Variable Local Slopes

A flattened Gaussian model can provide an accurate description of the LIDAR energy that is backscattered from the ground when the slope is homogeneous within the footprint; however, the model may be inappropriate when the slope is dissected or otherwise complex. The accuracy of the LIDAR DTM was previously demonstrated to be less than 35 cm [11], and it was therefore used in this section to compute a reference digital slope model. We then computed the standard deviation of the pixel values of the digital slope model (i.e., the local variability of the hillslope within the large-footprint waveform) and tested its correlation with the estimation error of the mean slope (retrieved from $\sigma_{\mathrm{ground}}$).

## IV. Results

The results are summarized in Table I and Fig. 2. Using the proposed method (method 1) with a hillslope derived from the high-resolution LIDAR DTM, the canopy height was assessed with a 3.3-m rmse, which corresponded to a 16.1% MAPE. The bias of $-2.1$ m and the standard deviation of 2.5 m are lower than the standard deviation of the reference tree heights, and the $R^2$ value is 0.86 (Fig. 2). Using the SRTM DEM, the fitting algorithm failed to process one of the 27 LIDAR waveforms. For 26 footprints (a failure rate of approximately 4%), we obtained a 4.3-m rmse that corresponded to a MAPE of 27.0%. We observed a lower bias (1.4 m) but a higher standard deviation (4.1 m) and an $R^2$ of 0.66. When self-estimating the hillslope based on the fitted model, the results were more accurate than those obtained using the SRTM DEM. Despite a slightly higher bias ($-1.6$ m), we observed a standard deviation that was lower than the standard deviation of the reference tree heights (3.2 m), a lower rmse (3.5 m), and a lower MAPE (21.6%).

The canopy-height estimations obtained using the Lee *et al.* [8] method (method 2) with the SRTM DEM displayed stronger bias, standard deviation, rmse, and MAPE values than those by method 1, although a lower bias was observed using the LIDAR DTM. It is noteworthy that the accuracy of the canopy-height estimates obtained with the self-estimated hillslope in method 1 was similar to the accuracy obtained using method 2 with the high-resolution DTM (Table I).

TABLE I
ERROR IN THE TREE-HEIGHT RETRIEVALS USING HILLSLOPE DATA DERIVED FROM THE LIDAR DTM, SRTM DEM, AND SELF-ESTIMATED SLOPE. EACH LIDAR-DERIVED ESTIMATION WAS COMPARED TO THE MEAN OF THE CORRESPONDING REFERENCE TREE HEIGHTS (ESTIMATED–FIELD MEASURED). METHOD 1 REFERS TO THE METHOD PRESENTED IN THIS LETTER, AND METHOD 2 IS THE METHOD PRESENTED BY LEE *ET AL.* [2011]

| | LIDAR DTM | | SRTM DEM | | Self-estimation |
|---|---|---|---|---|---|
| | Method 1 | Method 2 | Method 1 | Method 2 | Method 1 |
| Number of estimates | 27 | 27 | 26 | 27 | 27 |
| Estimates bias (m) | -2.1 | -1.4 | 1.4 | 3.8 | -1.6 |
| Estimates standard deviation (m) | 2.5 | 3.4 | 4.1 | 4.5 | 3.2 |
| RMSE (m) | 3.3 | 3.6 | 4.3 | 5.8 | 3.5 |
| MAPE (%) | 16.1 | 20.5 | 27.0 | 38.8 | 21.6 |

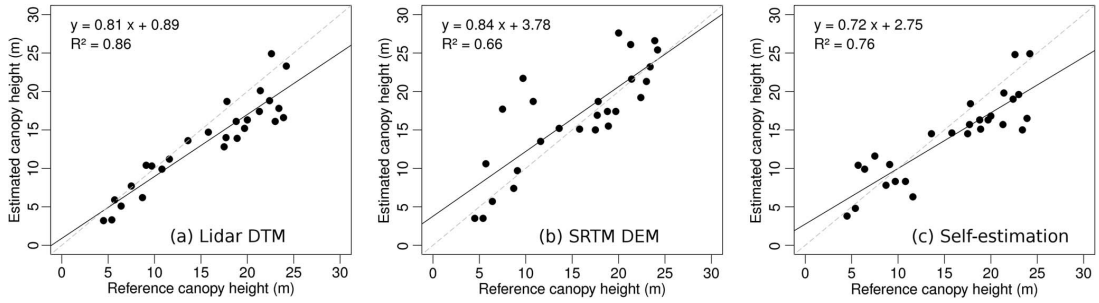



Fig. 2. (Black) Regression lines between the estimated and reference tree heights using (a) the LIDAR DTM on 27 footprints, (b) the SRTM DEM for the hillslope approximation on 26 footprints, and (c) the self-estimation of the hillslope on 27 footprints. The gray dotted lines are the 1:1 slopes.

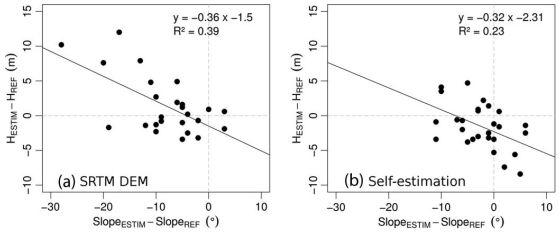



Fig. 3. (Black) Regression lines between the error in the canopy-height estimations and the error in the hillslope estimations using (a) the SRTM DEM and (b) the retrieved Gaussian stretch parameter. The gray dotted lines indicate $x = 0$ and $y = 0$.

The relationship between the error in the canopy-height estimation and the error in the hillslope estimation is shown in Fig. 3. A portion of the error in the height estimation using the SRTM DEM is derived from the error in the hillslope estimation (Fig. 3). Fewer scattered values were observed using the self-estimated slope (retrieved from $\sigma_{\mathrm{ground}}$), which was more accurate than the slope estimated using the SRTM DEM. The error ranged from $-11^\circ$ to $6^\circ$ for the self-estimated slope and from $-28^\circ$ to $3^\circ$ for the SRTM DEM-estimated slopes (Fig. 3). A similar relationship was not observed when the LIDAR DTM was used because the slopes that were derived from the LIDAR DTM closely matched the reference measurements.

We observed that the standard deviation of the tree-height estimates decreased with increasing accuracy of the slope estimation (i.e., the methods can be sorted by increasing accuracy as follows: SRTM DEM, self-estimation, and LIDAR DTM). However, the bias in the height estimation did not follow the same trend. In addition, we did not observe any significant relationship (random distribution) between the slope variability within the footprints (standard deviation ranging from $\pm 2^\circ$ to $\pm 11^\circ$ with a mean of $\pm 6^\circ$) and the error of the self-estimated hillslope. A significant relationship between the stem density and the error in the hillslope estimates was not observed.

## V. DISCUSSION AND CONCLUSION

Both the theoretical and empirical studies presented here have demonstrated that a Gaussian laser pulse reflected from a sloped surface produces a flattened Gaussian return, which we approximated using a generalized Gaussian function. This approximation of the ground return facilitated the mean ground positioning, whereas a multi-Gaussian fit (such as in the GLA14 product that was derived from ICESat waveforms) led to uncertainties in the correspondence between a Gaussian peak and the ground position [17]. The model was not sensitive to the hillslope variability within the footprint. The shadow that was produced on the ground by the tree crown led to the recording of an erratic ground waveform component (see Fig. 1). However, the fit of the generalized Gaussian model was not affected.

In theory, the canopy waveform component does not have the same shape as a generalized Gaussian; however, the width and flattening parameters of this function were found to provide a sufficient extent of shape modification to fit the waveforms.

Furthermore, a slight gap between the modeled canopy component and the actual waveform would not have influenced the canopy-top positioning because it was retrieved using a threshold method.

It was determined to be more correct to compare estimates using the dominant height rather than the mean height of all trees; some of which may not have belonged to the upper layer. Because some of reference trees did not belong to the upper layer, we compared the height estimates derived from LIDAR signals with the dominant height rather than the average of all the heights of trees in the plot. The retrieved canopy heights were generally lower than the actual dominant heights, although an overestimation was observed when using the SRTM DEM that may be related to the slope underestimation (Fig. 3). Method 2 provided the tallest tree height measurements, although it could be reduced by 1 m by taking into account the ringing effect [18] that can broaden the received waveform.

A high-resolution DTM is not always available to provide the optimal correction of canopy-height estimates. Self-estimation of the hillslope based on large-footprint waveforms was shown here to be a more valuable solution for correcting canopy-height measurements than the use of a coarser ancillary DTM, such as the SRTM DEM (80-m pixel with a 10-m vertical accuracy) or the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER DEM, 30-m resolution). Despite its higher resolution, the ASTER DEM did not provide estimates that were more accurate than the results using the SRTM (tested but not shown here).

Although the validity of the proposed model requires further confirmation in a more complex forest environment, such as multilayered or irregular stands, the fit of two generalized Gaussian functions was shown to be a promising alternative to existing methods for processing large-footprint LIDAR waveforms. By separating the canopy and ground waveform components and self-estimating the hillslope within large-footprint LIDAR waveforms (even in plots with low vegetation heights and high slopes, i.e., with a mixture of both components), this approach offers new perspectives for refining biomass estimates using large-footprint LIDAR data. In addition, this method offers a new approach for improving hydrological predictions in forested areas in which global DEMs are known to provide an inaccurate assessment of the ground elevation.

## ACKNOWLEDGMENT

The authors would like to thank the GIS Draix for providing the full-waveform light detection and ranging data and the National Institute of Sciences of the Universe (INSU) for its support to GIS Draix through the ORE program. The authors would also like to thank L. Albrech, J. Baron, and C. Véga (Irstea/Territories, Environment, Remote Sensing and Spatial Information Joint Research Unit) for their assistance with the ground truth surveys. This work is a part of the ExFOLIO project.

## REFERENCES

[1] H. C. Muller-Landau, R. S. Condit, J. Chave, S. C. Thomas, S. A. Bohlman, S. Bunyavejchewin, S. Davies, R. Foster, S. Gunatilleke, N. Gunatilleke, K. E. Harms, T. Hart, S. P. Hubbell, A. Itoh, A. R. Kassim, J. V. LaFrankie, H. S. Lee, E. Losos, J. R. Makana, T. Ohkubo, R. Sukumar, I. F. Sun, M. N. Nur Supardi, S. Tan, J. Thompson, R. Valencia, G. V. Muñoz, C. Wills, T. Yamakura, G. Chuyong, H. S. Dattaraja, S. Esufali, P. Hall, C. Hernandez, D. Kenfack, S. Kiratiprayoon, H. S. Suresh, D. Thomas, M. I. Vallejo, and P. Ashton, "Testing metabolic ecology theory for allometric scaling of tree size, growth and mortality in tropical forests," *Ecol. Lett.*, vol. 9, no. 5, pp. 575–588, May 2006.

[2] M. A. Lefsky, "A global forest canopy height map from the Moderate Resolution Imaging Spectroradiometer and the Geoscience Laser Altimeter System," *Geophys. Res. Lett.*, vol. 37, no. 15, p. L15 401, Aug. 2010.

[3] A. L. Neuenschwander, T. J. Urban, R. Gutierrez, and B. E. Schutz, "Characterization of ICESat/GLAS waveforms over terrestrial ecosystems: Implications for vegetation mapping," *J. Geophys. Res.*, vol. 113, no. 2, p. G02 S03, Apr. 2008.

[4] Q. Chen, "Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry," *Remote Sens. Environ.*, vol. 114, no. 7, pp. 1610–1627, 2010.

[5] W. Abdalati, H. Zwally, R. Bindschadler, B. Csatho, S. Farrell, H. Fricker, D. Harding, R. Kwok, M. Lefsky, T. Markus, A. Marshak, T. Neumann, S. Palm, B. Schutz, B. Smith, J. Spinhirne, and C. Webb, "The ICESat-2 laser altimetry mission," *Proc. IEEE*, vol. 98, no. 5, pp. 735–751, May 2010.

[6] J. Sauber, M. Hofton, R. Bruhn, S. Luthcke, and B. Blair, "DESDynI lidar for solid Earth applications," in *Proc. IEEE IGARSS*, 2010, pp. 1903–1906.

[7] S. Durrieu, "Design of a European space borne lidar system for vegetation mapping," presented at the SilviLaser Int. Conf. LiDAR Appl. Assessing Forest Ecosyst., Freiburg, Germany, Sep. 14–17, 2010.

[8] S. Lee, W. Ni-Meister, W. Yang, and Q. Chen, "Physically based vertical vegetation structure retrieval from ICESat data: Validation using LVIS in White Mountain National Forest, New Hampshire, USA," *Remote Sens. Environ.*, vol. 115, no. 11, pp. 2776–2785, Nov. 2011.

[9] T. Allouis, J.-S. Bailly, Y. Pastol, and C. Le Roux, "Comparison of LiDAR waveform processing methods for very shallow water bathymetry using Raman, near-infrared and green signals," *Earth Surf. Process Landf.*, vol. 35, no. 6, pp. 640–650, 2010.

[10] W. Wagner, A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka, "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner," *ISPRS J. Photogramm.*, vol. 60, no. 2, pp. 100–112, 2006.

[11] A. Chauve, C. Vega, S. Durrieu, F. Bretar, T. Allouis, M. Pierrot-Deseilligny, and W. Puech, "Advanced fullwaveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest," *Int. J. Remote Sens.*, vol. 30, no. 19, pp. 5211–5228, Oct. 2009.

[12] D. M. Bates and D. G. Watts, *Nonlinear Regression Analysis and Its Applications*. Hoboken, NJ: Wiley, 1988.

[13] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, L. Roth, D. Seal, S. Shaffer, J. Shimada, J. Umland, M. Werner, M. Oskin, D. Burbank, and D. E. Alsdorf, "The Shuttle Radar Topography Mission," *Rev. Geophys.*, vol. 45, no. 2, 2007.

[14] M. A. Lefsky, D. J. Harding, M. Keller, W. B. Cohen, C. C. Carabajal, F. Del Bom Espirito-Santo, M. O. Hunter, and R. de Oliveira, Jr., "Estimates of forest canopy height and above-ground biomass using ICESat," *Geophys. Res. Lett.*, vol. 32, no. 22, pp. 1–4, 2005.

[15] J. Boudreau, R. Nelson, H. Margolis, A. Beaudoin, L. Guindon, and D. Kimes, "Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3876–3890, 2008.

[16] O. Garcia and A. Batho, "Top height estimation in lodgepole pine sample plots," *Western J. Appl. Forestry*, vol. 20, no. 1, pp. 64–68(5), Jan. 2005.

[17] J. A. B. Rosette, P. R. J. North, and J. C. Suárez, "Vegetation height estimates for a mixed temperate forest using satellite laser altimetry," *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1475–1493, 2008.

[18] C. Mallet, U. Soergel, and F. Bretar, "Analysis of full-waveform lidar data for classification of urban areas," *Int. Archives Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 37, no. 3A, pp. 85–92, 2008.

# Conclusion et perspectives

L'OBJECTIF de cette thèse était de déterminer des configurations de capteurs lidars dédiés à l'étude de la végétation forestière, et de proposer des méthodes d'extraction de paramètres forestiers adaptés aux différentes configurations. La capacité de différentes résolutions (taille d'empreinte et échantillonnage spatial), longueurs d'onde et modes d'enregistrement du signal retour à mesurer des paramètres forestiers (hauteurs et densité d'arbres, taille des couronnes et indirectement volume et biomasse) a été évaluée. Les études ont été menées de l'échelle de l'arbre jusqu'à celle du peuplement, sur des données expérimentales ou simulées.

## Échantillonage du signal retour

L'information la plus complète sur la scène observée est obtenue par la numérisation du signal retour avec un pas de 15 cm (1 GHz). Ce mode d'enregistrement est couramment utilisé sur le lidar satellitaire GLAS et sur les nouveaux capteurs aéroportés *full-waveform*. Bien que le laser soit atténué tout au long de son parcours dans la végétation, les signaux *full-waveform* apportent des informations quasi continues sur la répartition verticale des éléments ayant réfléchi le signal. La correction de cette atténuation a permis, dans la seconde partie de la thèse, d'obtenir des informations plus précises sur le profil vertical des arbres et du peuplement. Un pas d'enregistrement encore plus précis est envisageable, mais son apport ne semble pas justifier son surplus de coût.

Les capteurs aéroportés multiéchos n'enregistrent, quant à eux, que les échos les plus significatifs de l'onde retour, perdant au passage une certaine quantité d'information. Ce mode d'enregistrement est le plus courant, puisque l'information apportée semble néanmoins fiable. En effet, la hauteur des arbres, la taille des couronnes (voir section 1.2) et le modèle numérique de terrain (voir section 3.2) ont été extraits respectivement à 57 cm, 1.48 m et 35 cm près, en accord avec les études précédentes (*Hollaus et al., 2006; Popescu et al., 2003; Chauve et al., 2009*).

L'apport de données *full-waveform* pour l'estimation du volume et de la biomasse à l'échelle de l'arbre individuel a été évalué. Les résultats ont montré que les données *full-waveform* ne permettent pas d'améliorer les estimations du volume des troncs. Celles-ci sont obtenues avec une précision de 16% en utilisant la taille de la couronne, la hauteur d'arbre et volume englobant de l'arbre dérivés des données multiéchos. À l'inverse, l'énergie totale rétrodiffusée par l'arbre, dérivée des signaux *full-waveform*, a permis une amélioration des estimations de biomasse. Cette amélioration est de l'ordre de 9% par rapport aux estimations effectuées dans cette étude ou dans les précédentes (*e.g. Popescu (2007)*) en utilisant des données multiéchos. Elle semble être liée à une meilleure prise en compte par les signaux complets de la biomasse présente dans les branches et les feuilles.

Dans le cadre d'une expérimentation d'un prototype lidar, la numérisation du signal retour à un pas de 1,5 m (100 MHz) a été testée (voir section 2.2). Ce mode d'enregistrement se situe entre les deux précédents, produisant des échos intermittents, mais régulièrement répartis. Il n'a pas semblé affecter les mesures de hauteurs d'arbres, puisque les estimations ont été données avec une erreur moyenne inférieure à 70 cm. Néanmoins, cette précision a été obtenue dans des peuplements homogènes avec une

densité de mesures permettant d'établir une moyenne calculée sur l'ensemble de la parcelle, au lieu d'une évaluation arbre à arbre. Bien que la précision obtenue soit convenable, l'apport d'un enregistrement à 100 MHz ne présente pas de réel avantage. Le matériel permettant une numérisation du signal à 1 GHz est aujourd'hui abordable et le stockage d'un nombre plus important de mesures n'est plus critique.

## Longeur d'onde du laser

Le proche infrarouge est particulièrement adapté aux applications forestières, car la végétation possède dans ce domaine une transmittance et une réflectance importantes (*Asner, 1998*). C'est donc la longueur d'onde privilégiée pour les lidars topographiques aéroportés utilisés sur la forêt. Il est cependant possible d'utiliser d'autres longueurs d'onde, en compensant les différences de réflectance et de transmittance par une modification de la puissance du signal émis.

Le prototype lidar présenté dans la seconde partie fonctionne avec un laser ultraviolet initialement développé pour la mesure des concentrations de l'atmosphère en aérosols (*Chazette et al., 2007*). Cette longueur d'onde n'avait jusqu'à présent pas été testée sur un lidar dédié à la mesure de paramètres forestiers. La végétation possède une réflectance et une transmittance environ dix fois plus faibles dans l'utraviolet que dans le proche infrarouge (*Stam, 2008*). L'oeil humain étant moins sensible à la lumière ultraviolette qu'au proche infrarouge, des impulsions de plus forte énergie peuvent être émises sans risque afin de pénétrer plus profondément dans des forêts denses. Aussi, une réflectance plus faible de la végétation dans l'ultraviolet rend le laser ultraviolet moins sensible aux réflexions multiples à l'intérieur du couvert forestier, ce qui limite les erreurs de positionnement des photons dont le trajet a été dévié plusieurs fois avant de revenir au capteur. En validant la capacité du laser ultraviolet à restituer des paramètres de peuplements forestiers, l'expérience a ouvert la voie au développement de lidars bifonctions pour l'étude conjointe de l'atmosphère et de la végétation.

## Taille d'empreinte

La plus forte résolution spatiale est obtenue avec les capteurs disposant d'une petite taille d'empreinte (quelques dizaines de centimètres de diamètre) et un fort taux d'échantillonnage spatial (plusieurs mesures par mètre carré). Les données lidars à haute résolution spatiale sont évidemment les plus adaptées pour mesurer en détail la structure forestière jusqu'à l'échelle de l'arbre individuel, de même que pour produire des modèles numériques de terrain à très haute résolution (50 cm voire moins si la densité de mesure le permet) (*Chauve et al., 2009*).

Cependant, la cartographie à l'échelle de l'arbre individuel n'est généralement pas nécessaire à l'échelle régionale. Bien qu'elle soit techniquement possible, elle nécessiterait un temps de calcul important et poserait un problème de coût d'opération depuis un vecteur aéroporté. De plus, un lidar à haute résolution ne peut pas être embarqué sur satellite essentiellement pour des raisons de sécurité oculaire et de durée de vie du laser. Il est donc nécessaire de choisir la résolution du capteur en fonction de l'échelle d'observation, des paramètres recherchés, et de certaines contraintes de coût ou de faisabilité.

Dans la première partie de la thèse, des résolutions différentes ont indirectement été testées au travers de l'agrégation, à l'échelle de l'arbre, des signaux complets à haute résolution spatiale. Ces données agrégées ont apporté de nouvelles informations sur la quantité d'énergie réfléchie sur l'arbre et donc sur sa structure (voir section 1.2). Elles ont ainsi permis d'améliorer les estimations de biomasse. Cependant, elles se sont avérées moins performantes que les données multiéchos à haute résolution dans l'extraction des hauteurs d'arbres, la détection de l'altitude du sol étant imprécise dans les signaux issus d'empreintes de plusieurs mètres de diamètre. La méthode de détection des arbres individuels a montré sa limite dans les peuplements dont la densité dépasse 500 arbres par hectare, mais probablement toutes les méthodes de délinéation des couronnes, automatiques ou manuelles, basées sur des données de télédétection acquises verticalement trouveront leurs limites face aux couverts fermés (*Koch et al., 2006; Rowell et al., 2006*).

Pour la cartographie de grandes surfaces, l'utilisation de tailles d'empreintes supérieures au mètre a été envisagée. Une expérimentation a donc été menée dans les Landes avec une empreinte de 2,4 m qui n'avait jusqu'alors pas été testée dans le domaine forestier. Bien que le positionnement des mesures dans l'espace n'ait pas été plus précis que 5 m, les variations de hauteurs des arbres et leur espacement moyen sur la parcelle ont pu être extrait en utilisant des statistiques spatiales (variogrammes). Les variations de la structure de la forêt sur la parcelle ont donc été mesurées, alors que la résolution des mesures ne permettait pas d'accéder à l'arbre individuel. Un autre avantage d'une taille d'empreinte métrique est sa moindre sensibilité à un terrain en pente, en comparaison à des empreintes plus larges.

Le problème des empreintes décamétriques est leur forte sensibilité à la topographie, rendant délicate l'identification de la position du haut des arbres et du sol (*Lee et al., 2011*). En effet, plus l'empreinte sera large, plus l'énergie réfléchie sur les arbres situés en bas de la pente se confondra avec l'énergie réfléchie sur le sol en haut de la pente. Pour des raisons de sécurité oculaire, une empreinte décamétrique est cependant la seule solution envisageable pour les lidars satellitaires si l'on se cantonne au laser proche infrarouge couramment utilisé. Une méthode novatrice a donc été proposée dans cette thèse afin d'estimer la pente du terrain depuis le signal lidar et ainsi corriger les mesures de hauteurs de canopée. De cette manière, les hauteurs d'arbres ont pu être mesurées avec une précision de 21.6% jusqu'alors inenvisageable. Aussi, cette méthode a ouvert de nouvelles perspectives dans l'estimation de la topographie sous la forêt, à partir de données lidars à larges empreintes.

## Vers des lidars satellitaires dédiés à l'étude de la forêt

La connaissance de la structure en 3D de la végétation peut améliorer la compréhension du cycle du carbone et de son impact sur le réchauffement climatique. Le besoin d'un inventaire de la biomasse forestière à l'échelle globale est donc de plus en plus fort, afin d'améliorer les prévisions de concentration de dioxyde de carbone dans l'atmosphère. Aussi, la spatialisation à grande échelle de la structure 3D des forêts ouvrirait de grandes perspectives pour la modélisation de la dynamique des peuplements en écologie de la végétation forestière, ainsi qu'en foresterie pour le suivi de l'exploitation et des interventions sylvicoles.

Des données lidars satellitaires ont par le passé été utilisées pour mesurer la structure forestière, malgré certains inconvénients liés à la configuration du capteur qui n'avait pas été choisie pour étudier spécifiquement la forêt (*Chen, 2010; Lee et al., 2011*). De nouvelles missions satellitaires sont donc à l'étude. À côté du choix de la taille d'empreinte, abordé précédemment, un des enjeux centraux est de calibrer l'énergie d'émission afin que la dynamique du signal retour permette accéder à la position du sol, même dans les forêts les plus denses. À défaut d'avoir apporté une quantification précise de cette énergie à émettre, une partie de ces travaux de thèse se sont appliqués à simuler la forme du signal retour dans différents contextes forestiers, par l'agrégation de données aéroportées.

Une taille d'empreinte d'une largeur maximale de 30 m semble être un compromis approprié à l'étude de la forêt depuis un satellite. Une taille plus faible, de l'ordre d'une dizaine de mètres, permettrait d'atteindre des estimations encore plus précises car moins influencées par la pente du terrain. Mais, afin d'effectuer un échantillonnage le plus complet possible, il est nécessaire de trouver un compromis entre la taille de l'empreinte et l'espacement des mesures le long de la trajectoire du satellite ainsi qu'entre deux passages contigus. L'espacement des mesures le long de la trajectoire dépend de la fréquence de répétition du laser (limitée par la durée de vie du laser) et de la vitesse du satellite. L'espacement entre deux transects contigus dépend de la vitesse du satellite et de la rotation de la Terre sur elle-même. Ces espacements sont respectivement de 172 m et 14.5 km (à l'équateur sur une période de 183 jours) pour GLAS (*Zwally et al., 2002*). La fréquence de passage sur satellite au dessus du même point doit aussi être choisie, soit pour favoriser le suivi temporel d'une même surface (cas de GLAS avec une revisite de 183 jours), soit pour répartir les mesures le plus uniformément possible sur l'ensemble de la Terre (cas de LEAF avec une revisite tous les trois ans (*Durrieu, 2010*)).

## Pistes de recherches complémentaires

L'estimation de la biomasse effectuée à l'échelle de la placette – et plus à l'échelle de l'arbre individuel comme dans la première partie de la thèse – permettrait d'approcher le potentiel d'une empreinte de 30 m de diamètre. Il serait intéressant de comparer les estimations de biomasse effectuées à l'aide de données multiéchos à haute résolution spatiale avec celles effectuées en utilisant des signaux *full-waveform* agrégés à l'échelle de la placette. Comparée à des approches à l'arbre individuel, celle-ci permettrait de prendre en compte, dans une mesure certes difficile à préciser a priori, la biomasse des arbres dominés ainsi que de la végétation du sous-bois. Elle permettrait aussi de traiter plus aisément le cas de peuplements à couvert plus fermé qui rendent l'identification des couronnes difficile. Une question à laquelle il sera enfin nécessaire de répondre est la capacité d'une telle méthode à traiter le cas de forêts composées de différentes essences d'arbres.

Le potentiel d'un lidar satellitaire repose sur sa capacité à pénétrer le couvert forestier jusqu'au sol. Il est cependant nécessaire d'étudier soigneusement la dynamique du signal afin d'éviter la saturation du capteur en cas de faible couvert ou de surface uniforme tels un sol nu ou le haut d'un couvert très dense. Les modèles de simulation de signaux à larges empreintes présentés dans ce mémoire ont permis d'obtenir rapidement des résultats, compte tenu des développements effectués pour les autres parties de la thèse. Afin d'être entièrement opérationnels et d'approcher de plus près la forme d'un signal lidar satellitaire réel, ces modèles doivent cependant prendre en compte l'influence de l'atmosphère sur le signal. Ils doivent aussi être validés sur des peuplements plus complexes, composés de plusieurs étages de végétation et d'essences d'arbres différentes. Ce type d'approche est néanmoins prometteur et fourni une alternative aux complexes modèles optico géométriques et de transfert radiatif (*Ni-Meister et al., 2001*) ainsi qu'aux simulations par lancer de rayon sur des maquettes informatiques de plantes (*Disney et al., 2000*). En effet, ces dernières nécessitent d'importants investissements pour la mise au point de maquettes réalistes pour chaque essence et pour différents stades de développement des arbres, et sont difficiles à valider.

# Bibliographie

Allouis, T., Bailly, J.-S., Pastol, Y., and Le Roux, C. (2010). Comparison of lidar waveform processing methods for very shallow water bathymetry using raman, near-infrared and green signals. *Earth Surface Process and Landforms*, 35(6):640–650.

Asner, G. P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, 64(3):234 – 253.

Boudreau, J., Nelson, R., Margolis, H., Beaudoin, A., Guindon, L., and Kimes, D. (2008). Regional aboveground forest biomass using airborne and spaceborne lidar in québec. *Remote Sensing of Environment*, 112(10):3876–3890.

Brassel, P. (2001). *Swiss national forest inventory: methods and models of the second assessment*. Birmensdorf: Swiss Federal Research Institute WS.

Bruce, J. P., ed., Hoesung, L., ed., Haites, E. F., and ed. (1996). *Climate change 1995 : economic and social dimensions of climate change*. Cambridge University Press / Intergovernmental Panel on Climate Change and Working Group III, Cambridge.

Chauve, A., Vega, C., Durrieu, S., Bretar, F., Allouis, T., Pierrot-Deseilligny, M., and Puech, W. (2009). Advanced fullwaveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest. *International Journal of Remote Sensing*, 30:19:5211–5228.

Chazette, P., Sanak, J., and Dulac, F. (2007). New approach for aerosol profiling with a lidar onboard an ultralight aircraft: application to the african monsoon. *Environmental Science & Technology*, 41(24):8335–8341.

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*, 114(7):1610 – 1627.

Disney, M. I., Lewis, P., and North, P. R. J. (2000). Monte carlo ray tracing in optical canopy reflectance modelling. *Remote Sensing Reviews*, 18(2):163–196.

Dubayah, R. O. and Drake, J. B. (2000). Lidar remote sensing for forestry. *Journal of Forestry*, 98(6):44–52. Cited By (since 1996): 64.

Durrieu, S. (2010). Design of a european space borne lidar system for vegetation mapping. In *Silvilaser International Conference on LiDAR Applications for Assessing Forest Ecosystems, Freiburg, Germany, September 14th - 17th, 2010.*

Ellerman, A. D. and Buchner, B. K. (2007). The european union emissions trading scheme: Origins, allocation, and early results. *Review of Environmental Economics and Policy*, 1(1):66–87.

FAO (2006). *Global Forest Resources Assessment 2005. Progress towards sustainable forest management*, volume 147. Rome.

Gouvernement Français (2005). *Changements climatiques : guide explicatif des accords internationaux*. Mission interministérielle de l'effet de serre du Gouvernement Français.

Hall, F. G., Bergen, K., Blair, J. B., Dubayah, R., Houghton, R., Hurtt, G., Kellndorfer, J., Lefsky, M., Ranson, J., Saatchi, S., Shugart, H., and Wickland, D. (2011). Characterizing 3d vegetation structure from space: Mission requirements. *Remote Sensing of Environment*, 115(11):2753 – 2775.

Heinzel, J. and Koch, B. (2011). Exploring full-waveform lidar parameters for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, 13(1):152 – 160.

Hollaus, M., Wagner, W., Eberhöfer, C., and Karel, W. (2006). Accuracy of large-scale canopy heights derived from lidar data under operational constraints in a complex alpine environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(5):323–338.

Hyde, P., Dubayah, R., Walker, W., Blair, J. B., Hofton, M., and Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (lidar, sar/insar, etm+, quickbird) synergy. *Remote Sensing of Environment*, 102(1-2):63 – 73.

IFN (2010). Un inventaire annuel sur la france entière. Inventaire Forestier National. http://www.ifn.fr/spip/IMG/pdf/100506-2p-newmetho.pdf.

Kato, A., Moskal, L. M., Schiess, P., Swanson, M. E., Calhoun, D., and Stuetzle, W. (2009). Capturing tree crown formation through implicit surface reconstruction using airborne lidar data. *Remote Sensing of Environment*, 113(6):1148 – 1162.

Koch, B., Heyder, U., and Welnacker, H. (2006). Detection of individual tree crowns in airborne lidar data. *Photogrammetric Engineering and Remote Sensing*, 72(4):357–363.

Le Guen, J. (2010). *Rapport de la mission : "Protection des forêts tropicales et de leur biodiversité contre la dégradation et la déforestation"*. Présidence de la République Française.

Le Toan, T., Quegan, S., Davidson, M., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., and Ulander, L. (2011). The biomass mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment*, 115(11):2850 – 2860.

Lee, S., Ni-Meister, W., Yang, W., and Chen, Q. (2011). Physically based vertical vegetation structure retrieval from icesat data: Validation using lvis in white mountain national forest, new hampshire, usa. *Remote Sensing of Environment*, In Press, Corrected Proof:–.

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).

Lefsky, M. A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., and Harding, D. (1999). Lidar remote sensing of the canopy structure and biophysical properties of douglas-fir western hemlock forests. *Remote Sensing of Environment*, 70(3):339–361.

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 Jr., R. (2005). Estimates of forest canopy height and aboveground biomass using icesat. *Geophysical Research Letters*, 32(22):1–4.

Lim, K., Treitz, P., Wulder, M., St-Onge, B., and Flood, M. (2003). Lidar remote sensing of forest structure. *Progress in Physical Geography*, 27(1):88–106. Cited By (since 1996): 91.

Means, J. E., Acker, S. A., Fitt, B. J., Renslow, M., Emerson, L., and Hendrix, C. J. (2000). Predicting forest stand characteristics with airborne scanning lidar. *Photogrammetric Engineering and Remote Sensing*, 66(11):1367–1371.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., Harmon, M. E., and McKee, W. A. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western cascades of oregon. *Remote Sensing of Environment*, 67(3):298–308. Cited By (since 1996): 146.

Minang, P. A., Jungcurt, S., Meadu, V., Murphy, D., and International Institute for Sustainable Development (IISD) (2009). *The REDD negotiations : moving into Copenhagen*. IISD, Winnipeg, MB.

Muss, J. D., Mladenoff, D. J., and Townsend, P. A. (2011). A pseudo-waveform technique to assess forest structure using discrete lidar data. *Remote Sensing of Environment*, 115(3):824 – 835.

Nadeau, C. (2000). Analyse des effets atmosphériques dans les données en télédétection du moyen infrarouge sur la classification des minéraux de surface en milieu aride. Maîtrise en géographie (télédétection), Université de Sherbrooke.

Nelson, R., Krabill, W., and Tonelli, J. (1988). Estimating forest biomass and volume using airborne laser data. *Remote Sensing of Environment*, 24(2):247 – 267.

Ni-Meister, W., Jupp, D. L. B., and Dubayah, R. (2001). Modeling lidar waveforms in heterogeneous and discrete canopies. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9):1943–1958.

Næsset, E. (1997). Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 61(2):246 – 253.

Popescu, S. C. (2007). Estimating biomass of individual pine trees using airborne lidar. *Biomass and Bioenergy*, 31(9):646 – 655. Multiple benefits from sustainable bioenergy systems - Proceedings of a Joint Workshop of IEA Bioenergy Task 30 and Task 31, August 2005, Perth, Western Australia, IEA Bioenergy T31:2007:01.

Popescu, S. C., Wynne, R. H., and Nelson, R. F. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29(5):564–577.

Prabhu, R., Colfer, J. C., and Dudley, R. (1999). *Guidelines for developing, testing and selecting criteria and indicators for sustainable forest management.*, volume 1. Center for International Forestry Research, Indonesia.

Proisy, C., Couteron, P., and Fromard, F. (2007). Predicting and mapping mangrove biomass from canopy grain analysis using fourier-based textural ordination of ikonos images. *Remote Sensing of Environment*, 109(3):379–392.

Rowell, E., Selelstad, C., Vierling, L., Queen, L., and Shepperd, W. (2006). Using laser altimetry-based segmentation to refine automated tree identification in managed forests of the black hills, south dakota. *Photogrammetric Engineering and Remote Sensing*, 72(12):1379–1388.

Sexton, J. O., Bax, T., Siqueira, P., Swenson, J. J., and Hensley, S. (2009). A comparison of lidar, radar, and field measurements of canopy height in pine and hardwood forests of southeastern north america. *Forest Ecology and Management*, 257(3):1136 – 1147.

Shugart, H. H., Saatchi, S., and Hall, F. G. (2010). Importance of structure and its measurement in quantifying function of forest ecosystems. *Journal of Geophysical Research G: Biogeosciences*, 115(4).

Stam, D. M. (2008). Spectropolarimetric signatures of earth-like extrasolar planets. *Astronomy and Astrophysics*, 482(3):989–1007.

Sun, G., Ranson, K., Kimes, D., Blair, J., and Kovacs, K. (2008). Forest vertical structure from glas: An evaluation using lvis and srtm data. *Remote Sensing of Environment*, 112(1):107 – 117.

Tonolli, S., Dalponte, M., Vescovo, L., Rodeghiero, M., Bruzzone, L., and Gianelle, D. (2010). Mapping and modeling forest tree volume using forest inventory and airborne laser scanning. *European Journal of Forest Research*, 130(4):569–577. Article in Press.

van der Werf, G., Morton, D., DeFries, R., Olivier, J., Kasibhatla, P., Jackson, R., and et al (2009). Co2 emissions from forest loss. *Nature Geoscience*, 2:737–738.

Zwally, H. J., Schutz, B., Abdalati, W., Abshire, J., Bentley, C., Brenner, A., Bufton, J., Dezio, J., Hancock, D., Harding, D., Herring, T., Minster, B., Quinn, K., Palm, S., Spinhirne, J., and Thomas, R. (2002). Icesat's laser measurements of polar ice, atmosphere, ocean, and land. *Journal of Geodynamics*, 34(3-4):405–445.