

ESTIMATION DE PARAMÈTRES FORESTIERS PAR DONNÉES LIDAR AÉROPORTÉ ET IMAGERIE SATELLITAIRE RAPIDEYE : ÉTUDE DE SENSIBILITÉ

Jean-Matthieu Monnet^{1,2}, Émilie Chirouze^{1,2}, Éric Mermin^{1,2}

1: Irstea, UR EMGR, 2 rue de la Papeterie - BP 76, F-38402 St-Martin-d'Hères, France

2: Univ. Grenoble Alpes, F-38402 Grenoble, France

Résumé

Les différents critères qui influencent la précision de la méthode surfacique d'estimation de paramètres forestiers à partir de données LiDAR ont été étudiés le plus souvent séparément et dans des contextes forestiers variés. Cela limite les possibilités de comparaison en vue de la détermination des critères prépondérants du point de vue de la précision et du coût de ce type d'inventaire. Cet article présente une évaluation de l'influence des caractéristiques des relevés de terrain sur la précision des modèles d'estimation, ainsi que de l'intérêt de l'utilisation de l'information spectrale LiDAR ou issue de fusion avec des données satellitaires RapidEye. L'erreur obtenue en combinant les données LiDAR et RapidEye est de 16,5% pour la surface terrière, 23,7% pour la densité de tiges et 11,7% pour le diamètre moyen. Le choix de la taille de la placette de terrain et du diamètre de recensement ont un impact important sur l'erreur d'estimation. La précision du géoréférencement et le nombre de placettes apparaissent comme cruciaux dans la mesure où ils influencent non seulement l'erreur d'estimation mais également la précision de son évaluation.

Mots clés : LiDAR, Forêt, RapidEye

Abstract

Regarding the accuracy of forest estimates from airborne laser scanning data, several important parameters have already been investigated separately, and in various forest contexts. Comparing the results in order to determine the optimal setting for the accuracy and cost of LiDAR-based inventories thus remains difficult. This article presents an assessment of the sensitivity of prediction accuracy with regards to ground truth data and to the use of spectral information from LiDAR intensity or RapidEye satellite imagery. With data fusion the error of basal area estimates is 16.5%, 23.7% for stem density and 11.7% for mean diameter. The size and diameter threshold of the forest plots greatly impact prediction accuracy. The position accuracy and sample number of forest plots are crucial points as they influence both prediction error and its estimation.

Keywords : Airborne laser scanning, Forest, RapidEye

1. Introduction

Le scanner laser (ou LiDAR) aérien est une technique de télédétection permettant de caractériser les éléments de la surface terrestre sous la forme d'un nuage de points géoréférencés. Un point correspond à un élément de la surface qui a réfléchi vers l'aéronef une des impulsions laser émises par le capteur embarqué. Cette technique connaît depuis une dizaine d'années un essor important de part ses applications dans les domaines de la topographie et de l'inventaire forestier. En forêt, la méthode de traitement la plus répandue, dite « surfacique » (Næsset, 2002; Hollaus et al., 2009) consiste à combiner la description géométrique du couvert végétal par le nuage de points laser et un réseau de placettes terrain de référence pour produire des cartes continues de paramètres forestiers. La relation entre les descripteurs du nuage de points, également appelés « métriques », et le paramètre forestier d'intérêt est calibrée statistiquement à partir des relevés de terrain. Même si des modèles prédictifs « physiquement interprétables » ont été

proposés (Bouvier et al., 2015), ils restent dépendants d'une calibration locale et sont difficilement transférables à d'autres massifs forestiers ou acquisitions LiDAR.

Les nombreux résultats déjà publiés pour différents types de forêt donnent un ordre de grandeur de la précision attendue pour les principaux paramètres forestiers, comme le volume sur pied, la surface terrière ou la hauteur du peuplement. Cependant pour un même massif forestier, les choix effectués en terme de configuration d'acquisition LiDAR, d'échantillonnage de données de terrain et de chaîne de traitement pour la modélisation ont un effet à la fois sur la précision du résultat et le coût global de l'opération. Les valeurs de précision ne sont pas généralisables, ni directement comparables.

Concernant l'acquisition LiDAR, la configuration de vol influe directement sur la manière dont les impulsions laser interagissent avec le couvert végétal, et donc sur la qualité de la représentation géométrique par le nuage de points qui en résulte. Les comparaisons basées sur plusieurs vols au-dessus d'une même zone restent rares, du

fait du coût important lié aux multiples acquisitions. Deux études (Næsset, 2004b, 2009) ont montré que les estimations de hauteur, de densité de canopée et de volume étaient robustes par rapport à la hauteur de vol, alors qu'une autre (Morsdorf et al., 2008) a constaté des différences concernant l'estimation de la hauteur des arbres et le taux de couvert. Selon des simulations menées sur l'influence de l'angle de scan (Holmgren et al., 2003), il semble que certaines essences soient plus affectées par de grands angles d'incidence. Évalué à partir de données réelles, l'effet d'un angle de scan plus important sur la géométrie du nuage de points semble cependant limité (Morsdorf et al., 2008; Montagni, 2013).

Si la capacité du scanner laser à caractériser la structure géométrique du couvert forestier est un atout majeur pour l'estimation des paramètres forestiers, l'addition d'une information spectrale permet d'améliorer la précision. Lorsque le capteur les enregistre, les valeurs d'intensité peuvent être utilisées (Hudak et al., 2006). Cependant leur utilisation rigoureuse nécessite au préalable une calibration assez complexe (Wagner, 2010). L'utilisation de données spectrales obtenues par imagerie aérienne ou satellitaire est une alternative (Hudak et al., 2006; Wallerman et Holmgren, 2007; Packalén et Maltamo, 2007; Takahashi et al., 2010).

Les références de terrain ont également une grande importance dans la chaîne de traitement, tout en représentant une part importante du coût de l'inventaire. En tant que données d'apprentissage, elles doivent couvrir la variabilité du paramètre et l'extension spatiale de la zone d'étude (Junttila et al., 2013). En tant que données de validation, elles doivent être un échantillon représentatif de la zone d'application du modèle, et comporter un effectif suffisant pour que les intervalles de confiance ne soient pas sous-estimés (Strunk et al., 2012). La calibration des modèles par grands types de peuplements permet par ailleurs d'obtenir une meilleure précision qu'avec un modèle global (Næsset, 2004a), d'où l'importance des données pré-existantes pour préparer les relevés de terrain. À l'échelle de la placette, une partie de l'erreur des modèles LiDAR est due à des effets de bord. En effet, le scanner détecte essentiellement des houppiers, qui sur la bordure d'une placette n'appartiennent pas forcément à des arbres dont le tronc se trouve à l'intérieur de l'emprise d'inventaire (White et al., 2013). Ces effets sont plus importants pour les placettes de petite taille (Monnet et al., 2010; Watt et al., 2013; Ruiz et al., 2014), et sont aggravés lorsque le géoréférencement des données de terrain est approximatif (Gobakken et Næsset, 2009; Frazer et al., 2011). Ils peuvent cependant être réduits grâce à une détection préliminaire des arbres dans les données LiDAR (Packalén et al., 2015).

Les métriques décrivant le nuage de points sont très variées, depuis les percentiles de hauteur usuels aux métriques bi- ou tri-dimensionnelles comme celles dérivant d'une segmentation préalable. Ces métriques sont ensuite utilisées telles quelles, ou après transformation et sélection, dans des modèles d'estimation paramétriques (Næsset, 2002) ou non paramétriques (Packalén et Mal-

tamo, 2007; Niska et al., 2010; Monnet et al., 2011; Penner et al., 2013).

Il apparaît que les différentes étapes influençant le coût et la précision des estimations par la méthode LiDAR surfacique ont été bien étudiées, mais dans des contextes forestiers et avec des chaînes de traitement très différents, si bien que la comparaison de leurs effets respectifs reste difficile. L'objectif de cet article est, sur la base d'un jeu de données de 68 placettes inventoriées dans un massif forestier de montagne, d'évaluer individuellement les effets des relevés de terrain (rayon des placettes, diamètre précomptable, précision du géoréférencement, nombre de placettes) et de l'utilisation de données annexe (stratification à partir des données externes, fusion avec de l'imagerie satellitaire RapidEye) sur la précision des estimations de surface terrière, densité de tiges et diamètre dominant.

2. Données

2.1. Zone d'étude

La zone d'étude (55.7 km²) est située dans la communauté de communes des Quatre Montagnes, dans le massif du Vercors en Isère, France (figure 1). L'altitude est comprise entre 930 et 1710 m. Le relief est relativement peu accidenté mais certaines zones karstiques présentent un fort micro-relief. Environ un tiers des forêts appartiennent à des propriétaires privés, principalement sur d'anciennes terres agricoles, alors que les forêts publiques sont gérées par l'Office National des Forêts. Les peuplements sont majoritairement résineux, à dominante de sapin pectiné (*Abies alba*) ou d'épicéa commun (*Picea abies*), mais on trouve également des mélanges de feuillus et résineux ainsi que quelques peuplements de pin. Les peuplements résineux sont traités en futaie irrégulière dans les forêts publiques, alors que dans le privé il s'agit plutôt de peuplements réguliers assez capitalisés.

2.2. Données télédétection

L'acquisition LiDAR aérienne a eu lieu le 20 septembre 2010 sur la partie ouest, et environ un an plus tard, en août 2011, pour les deux autres parties (figure 1). Le capteur est un scanner à retour d'onde complète RIEGL LMS-Q560. La vitesse de vol était de 170 km.h⁻¹ à une hauteur de 600 m au-dessus du sol. Le recouvrement inter-bandes était de 65%, et la densité d'émission de 170 kHz avec un angle de scan de ± 30 degrés. La densité moyenne d'impulsions est de $11,2 \pm 4,7$ m⁻², calculée à une résolution de 10 m. Le traitement des données brutes a été effectué par le prestataire : les échos ont été extraits et géoréférencés avec la suite logicielle RIEGL, puis le nuage de points a été classé selon les catégories « sol » et « végétation » avec TerraScan et livré au format LAS 1.2.

Les données satellitaires RapidEye sont fournies par la plate-forme Geosud. On utilise l'image brute orthorectifiée pour le département de l'Isère. La période d'acquisition est mai à octobre 2010. L'image comporte cinq

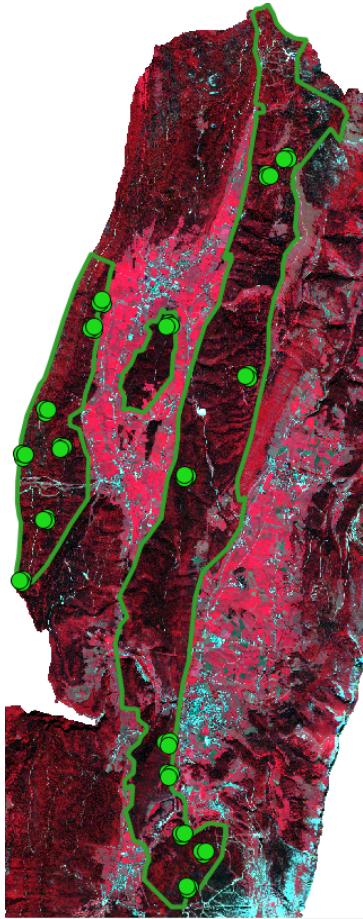


FIGURE 1: Présentation de la zone d'étude : emprise du vol LiDAR (bordure verte) et position des grappes de placettes (points). Fond : image RapidEye, canaux vert, rouge et proche infra-rouge en fausses couleurs.

bandes (bleu, vert, rouge, red-edge et proche infra-rouge) correspondant aux longueurs d'onde 440-510, 520-590, 630-685, 690-730 et 760-850 nm. La résolution spatiale après orthorectification est de 5 m.

2.3. Placettes forestières

68 placettes forestières ont été installées en deux phases. Les 32 placettes relevées lors de la première phase sont un sous-échantillon d'un inventaire statistique effectué sur les forêts publiques de l'ensemble de la communauté de communes (Le Ferrec, 2011). 36 placettes supplémentaires ont été inventoriées pendant l'été 2012, selon un protocole identique mais incluant également les forêts privées. L'échantillonnage a été réalisé par grappe de quatre placettes afin de limiter le temps de déplacement. La placette centrale est rejointe par navigation GPS. Les placettes périphériques ont leur centre à 60 m de celui de la placette centrale, dans les directions $\{az, az + 120, az + 240\}$ où az est un azimut tiré au sort entre 0 et 360°. Les placettes sont circulaires de rayon 15 m. Pour toutes les tiges vivantes de diamètre à 1,3 m supérieur à 7,5 cm, et dont les centres sont inclus dans l'emprise, sont relevés :

- l'azimut, la pente et la distance au sol jusqu'au centre de la placette ;
- le diamètre à 1,3 m ;
- l'essence.

Les coordonnées du centre de la placette sont relevées par GPS. Le géoréférencement des placettes a été amélioré en ajustant manuellement la localisation des arbres et le modèle numérique de hauteur. Trois paramètres forestiers sont calculés pour chaque placette (tableau 1) :

- surface terrière (G) : somme de la surface horizontale du tronc des arbres, mesurée à hauteur de poitrine (1,3 m), divisée par la surface de forêt occupée par les arbres, ici 707 m² ;
- densité de tiges (N) : nombre d'arbres divisé par la surface de forêt qu'ils occupent.
- diamètre moyen (D) : moyenne des diamètres.

	Surface terrière (m ² .ha ⁻¹)	Densité de tiges (ha ⁻¹)	Diamètre moyen (cm)	Densité d'impulsions (m ⁻²)
Min.	15,7	180	14,7	5,9
Moy.	37,2	860	22,0	11,8
Max.	73,4	1990	40,5	23,2

Tableau 1: Statistiques des placettes ($n = 68$).

3. Méthode

3.1. Calcul des métriques

Pour chaque placette, le nuage LiDAR se trouvant sur l'emprise inventoriée est extrait. Les hauteurs des points par rapport au sol sont calculées en retranchant à l'altitude des points l'altitude du sol obtenue par projection verticale sur une surface TIN (triangulated irregular network) constituée par les points classés « sol ». Le Modèle Numérique de Hauteur (MNH) est un raster de résolution 0,5 m calculé en retenant pour chaque pixel la hauteur du point le plus haut qu'il contient. Les pixels sans points reçoivent la valeur 0. Pour le calcul des métriques, les points de hauteur inférieure à 2 m sont ensuite exclus pour éviter l'influence de la végétation basse. 16 métriques de hauteur sont calculées : les 20^{ème}, 40^{ème}, 60^{ème}, 80^{ème} and 99^{ème} percentiles de hauteur pour les types d'échos selon qu'ils sont arrivés « seul », « premier de plusieurs » et « dernier », ainsi que la hauteur moyenne de tous les points. Les percentiles sont abrégés h_{gx} avec $g \in \{s, p, d\}$ le type d'écho et x le percentile. Trois métriques de densité sont ajoutées : la proportion des points qui se trouvent sous les seuils $\frac{h_{99}}{2}$, $\frac{2h_{99}}{3}$ et $\frac{5h_{99}}{6}$ où h_{99} est le 99^{ème} percentile de hauteur. Ils sont notés d_x avec x la proportion appliquée à h_{99} . Deux métriques sont calculées à partir du MNH : la moyenne du MNH sur la placette (MNH_m) et le pourcentage de valeurs du MNH supérieures à 2 m ($MNH_{>2}$). Six métriques d'intensité sont ajoutées : la moyenne et l'écart-type des valeurs d'intensité pour les points de hauteur supérieure à 2 m dans chaque groupe. Ils sont abrégés respectivement i_{gm} et i_{gsd} avec $g \in \{s, p, d\}$ le type d'écho.

Pour chaque placette, sont sélectionnés dans l'image RapidEye les pixels avec au moins 90% de leur surface dans l'emprise circulaire. Pour chaque pixel sont extraites les valeurs des canaux rouge (r), red-edge (re) et proche infrarouge (pir). On effectue ensuite les ratios suivants $ndvi = \frac{pir-r}{pir+r}$, $pir2r = \frac{pir}{r}$, $ndve = \frac{re-r}{re+r}$, $re2r = \frac{re}{r}$. Les 14 métriques RapidEye sont la moyenne et l'écart-type sur quatre ratios et des trois canaux sélectionnés sur chaque placette, et abrégés sur le modèle *ratio.g* où g désigne la moyenne (m) ou l'écart-type (sd).

3.2. Sélection des métriques et évaluation

Pour chaque paramètre forestier, une transformation Box-Cox est appliquée pour normaliser sa distribution avec l'ensemble des données. La sélection des métriques à inclure dans les modèles d'estimation s'opère selon le critère du coefficient de corrélation ajusté (R_a^2) le plus élevé en comparant pour chaque combinaison d'au plus trois métriques, le résultat d'un ajustement par moindres carrés. Une fois le jeu de métriques sélectionné, la précision du modèle est évaluée par 100 répétitions de validation croisée. Pour chaque répétition, dix groupes d'effectif égal sont constitués. Tour à tour, tous les groupes sauf un sont utilisés pour ajuster les coefficients de la régression avec les métriques sélectionnées, puis l'erreur d'estimation est calculée pour les placettes du groupe restant. L'erreur quadratique moyenne (eqm) est calculée pour chaque répétition, ainsi que son coefficient de variation (CV_{eqm}) obtenu en la divisant par la moyenne des valeurs mesurées sur le terrain.

3.3. Analyse par paramètre

Sauf mention contraire, le scénario de référence pour la chaîne de traitement est basé sur : utilisation de tous les points LiDAR, extraction des métriques LiDAR géométriques et intensité et de métriques RapidEye, utilisation de l'ensemble des données de terrain (géoréférencement amélioré manuellement, rayon de placette de 15 m et diamètre précomptable de 7,5 cm).

Information spectrale. La chaîne de traitement est répétée pour les modalités suivantes : les métriques LiDAR géométriques, d'intensité et les métriques RapidEye sont utilisées seules ou en combinaison pour constituer le jeu de variables du modèle.

Rayon des placettes. La chaîne de traitement est répétée pour les rayons de placette $r \in \{8, 9, \dots, 15\}$. Pour chaque placette, le calcul des paramètres forestiers est alors effectué en excluant les arbres situés à plus de r mètres du centre, et le calcul des métriques LiDAR et RapidEye en considérant l'emprise circulaire de rayon r m.

Diamètre précomptable. La chaîne de traitement est répétée pour les diamètres précomptables d de 7,5 à 22,5 cm par pas de 2,5. Pour chaque placette, le calcul des paramètres forestiers est alors effectué en excluant les arbres de diamètre inférieur à d cm.

Erreur de géoréférencement. Une erreur aléatoire est ajoutée aux coordonnées des placettes. L'erreur est supposée normale, d'écart-type $\sigma \in \{0, 1, 2, 4, 8\}$ et centrée en 0. Les erreurs étant ajoutées indépendamment aux deux coordonnées pour chaque placette, l'espérance pour la distance absolue entre la position d'origine de la placette et sa position perturbée est $\sigma\sqrt{\frac{\pi}{2}}$, et l'écart-type $\sigma\sqrt{\frac{4-\pi}{2}}$. Pour chaque valeur de σ , la chaîne de traitement suivante est répétée 50 fois : tirage au sort d'une erreur pour chaque coordonnée de chaque placette, extraction des métriques sur les nouvelles emprises, sélection des variables, évaluation de l'erreur d'estimation par deux répétitions de validation croisée. Deux erreurs d'estimation sont calculées : l'erreur apparente en calibrant et validant le modèle avec les métriques issues des emprises modifiées, et l'erreur réelle en validant avec les métriques issues des emprises initiales.

Nombre de placettes. La chaîne de traitement est répétée 100 fois pour chaque valeur du nombre de placettes de calibration $n \in \{40, 50, 60, 68\}$: tirage au sort de n placettes pour sélectionner les variables et calibrer le modèle ; estimation de l'erreur sur l'ensemble des placettes.

Stratification des placettes. On répète la chaîne de traitement pour deux stratifications des placettes. La première comprend deux groupes basés sur l'essence : résineux ($n_{E,r} = 22$) et mélange ou feuillus ($n_{E,m} = 46$). La seconde comprend deux groupes basés sur la structure : futaie irrégulière ($n_{S,fi} = 36$) et autre ($n_{S,a} = 32$). Ces stratifications ont été établies au préalable, la première à partir des données RapidEye, et la seconde par interprétation des orthophotos de la zone et par analyse des documents de gestion existants (Redon, 2012). Les étapes de sélection et calibration sont réalisées pour chaque sous-groupe et l'erreur est estimée pour chaque placette par validation croisée leave-one-out. L'erreur globale pour chaque stratification est calculée afin de pouvoir la comparer à l'erreur du modèle global.

4. Résultats et discussion

4.1. Information spectrale

L'apport d'une information spectrale dans les modèles d'estimation dépend du paramètre forestier considéré (figure 2). Pour la surface terrière, les métriques d'intensité LiDAR ne sont pas sélectionnées (tableau 2), par contre l'addition de métriques RapidEye permet de diminuer l'erreur moyenne, de 17,3 à 16,5% (figure 2). Il s'agit de métriques dérivées des canaux red-edge et rouge. Pour la densité de tiges, les métriques RapidEye ne sont pas sélectionnées, alors que les métriques d'intensité LiDAR contribuent à faire baisser l'erreur de 24,6% à 23,7% en moyenne. Enfin, le paramètre diamètre moyen bénéficie principalement des métriques RapidEye, avec une diminution de l'erreur de 12,8 à 11,7%. L'apport des métriques d'intensité en plus des métriques RapidEye et

	source	métriques	R_a^2
G	L	$h_{s60} + h_{s99} + MNH_{>2}$	71,8
	Li	$h_{s60} + h_{s99} + MNH_{>2}$	71,8
	R	$ndvi.m + ndve.m + re.sd$	52,8
	L+R	$d_{0,5} + ndve.m + re.sd$	73,8
	Li+R	$d_{0,5} + ndve.m + re.sd$	73,8
N	L	$h_{d60} + d_{0,5}$	78,6
	Li	$h_{d80} + i_{sm} + d_{0,5}$	81,2
	R	$pir.m + re2r.m + re.sd$	59,1
	L+R	$h_{d60} + d_{0,5}$	78,6
	Li+R	$h_{d80} + i_{sm} + d_{0,5}$	81,2
D	L	$h_{s40} + h_{d80} + d_{0,5}$	69,0
	Li	$h_{d60} + ism + d_{0,5}$	71,5
	R	$ndvi.m + r.m + r.sd$	59,6
	L+R	$h_{d40} + d_{0,5} + pir.m$	74,1
	Li+R	$h_{s99} + i_{sm} + pir.m$	74,3

Tableau 2: Apport des données spectrales : métriques sélectionnées et R_a^2 du modèle. Source : L (métriques géométriques LiDAR), Li (métriques géométriques et intensité LiDAR), R (métriques RapidEye).

géométriques LiDAR n'a pas un effet significatif sur l'erreur (test de Wilcoxon sur les 100 répétitions, $p = 0.6$).

L'apport complémentaire de l'information spectrale pourrait concerner la nature du couvert : présence de trouées et proportion feuillus/résineux. Ces aspects ne sont pas ou peu modélisés par la distribution en hauteur des points LiDAR, mais sont importants pour les modèles d'estimation qui portent sur des peuplements mélangés.

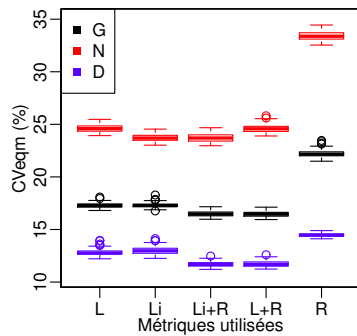


FIGURE 2: Boîtes à moustaches des valeurs d'erreur relative (CV_{eqm}) d'estimation, obtenues pour 100 répétitions de validation croisée, selon l'utilisation ou non de métriques RapidEye, pour les trois paramètres forestiers.

4.2. Données de terrain

4.2.1. Taille des placettes

Dans tous les cas, on constate une forte augmentation de l'erreur moyenne avec la diminution du rayon de placette de 15 à 8 m (figure 3). Pour la surface terrière, l'augmentation est de 16,4 à 24,4%, pour la densité de tiges de 23,7 à 31,9% et pour le diamètre dominant de 11,6 à 19,4%. Ces résultats sont conformes à ceux publiés dans la littérature et ils s'expliquent par l'effet de

	r (m)	métriques	R_a^2
G	15	$d_{0,5} + ndve.m + re.sd$	73,8
	11	$h_{l60} + d_{0,67} + MNH_{>2}$	72,3
	8	$h_{l99} + MNH_m + ndvi.m$	62,8
	15	$h_{l80} + i_{fm} + d_{0,5}$	81,2
N	11	$h_{s40} + d_{0,5} + MNH_{>2}$	80,2
	8	$h_{s60} + d_{0,67} + MNH_{>2}$	77,7
	15	$h_{s99} + i_{sm} + pir.m$	74,3
D	11	$h_{l60} + i_{sm} + d_{0,5}$	69,2
	8	$h_{l60} + i_{sm} + d_{0,67}$	64,1

Tableau 3: Métriques sélectionnées et R_a^2 du modèle selon le rayon r de la placette.

	d (cm)	métriques	R_a^2
G	7.5	$d_{0,5} + ndve.m + re.sd$	73,8
	15	$d_{0,5} + pir.m + re.sd$	69,2
	22.5	$d_{0,5} + MNH_m + pir.m$	59,9
N	7.5	$h_{l80} + i_{fm} + d_{0,5}$	81,2
	15	$h_{s80} + d_{0,67} + MNH_{>2}$	79,2
	22.5	$h_{f99} + MNH_m + ndve.m$	74,5
D	7.5	$h_{s99} + i_{sm} + pir.m$	74,3
	15	$h_{f20} + h_{l80} + i_{sm}$	73,0
	22.5	$h_{f80} + d_{0,5} + d_{0,83}$	74,6

Tableau 4: Métriques sélectionnées et R_a^2 du modèle selon le diamètre précomptable d .

bord plus important en proportion dans le cas de petites surfaces d'inventaire. La perte de précision est particulièrement importante pour les rayons en dessous de 12 m. Cette taille minimale est toutefois à mettre en relation avec la structuration horizontale du peuplement. Dans le cas d'une structure irrégulière où il existe une forte variabilité des hauteurs d'arbre au sein de la placette, les effets de bords seront plus importants que dans un peuplement régulier avec un couvert continu.

Il est possible également que l'effet de bord soit pénalisant pour les métriques RapidEye quand les placettes sont petites, du fait de la taille de pixel de 5 m. Ces métriques sont moins sélectionnées pour les rayons plus faibles (tableau 3). Pour la surface terrière, les métriques $d_{0,67}$ et $MNH_{>2}$ sont sélectionnées pour un rayon entre 9 et 14 m. Pour la densité de tiges, une métrique de hauteur (percentile 40 ou 60) est associée à une métrique de densité ($d_{0,5}$ sauf dans un cas). La troisième métrique est une métrique d'intensité LiDAR pour r inférieur ou égal à 12 m et $MNH_{>2}$ sinon. Les métriques sélectionnées pour le diamètre moyen montrent moins de constance mais il apparaît que dans l'ensemble les métriques pertinentes pour l'estimation de ces paramètres forestiers sont peu dépendantes de la taille de la placette.

4.2.2. Diamètre précomptable

Pour la surface terrière, l'erreur moyenne absolue (eqm) reste stable autour de 6,2 $m^2 \cdot ha^{-1}$, sauf quand d vaut 22.5 cm, avec une valeur de 6,7 (figure 4). L'erreur moyenne relative d'estimation (CV_{eqm}) augmente progressivement de 16,5% à 23,8% lorsque le diamètre précomptable passe

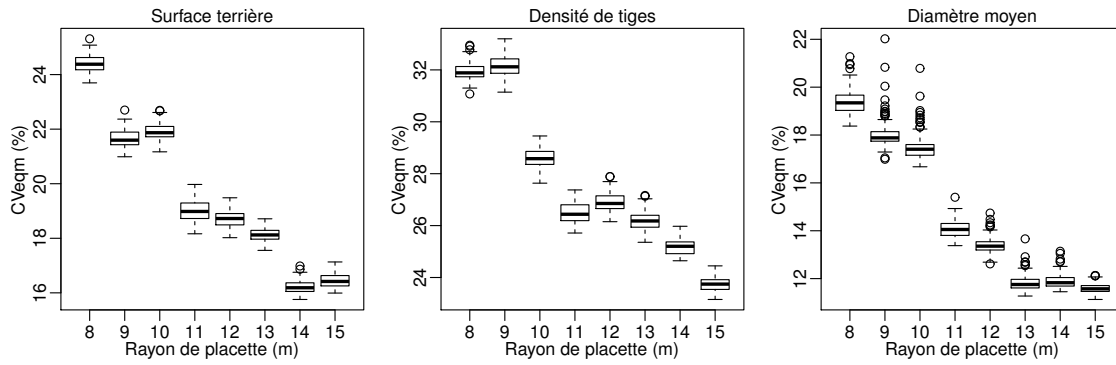


FIGURE 3: Boîtes à moustaches des valeurs d'erreur relative (CV_{eqm}) d'estimation obtenues pour 100 répétitions de validation croisée, pour différents rayons de placette.

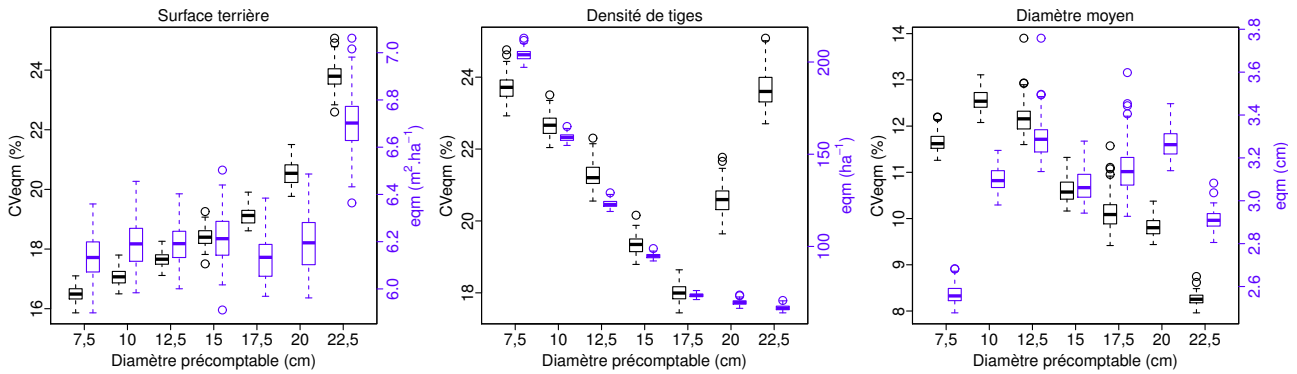


FIGURE 4: Boîtes à moustaches des valeurs d'erreur relative (CV_{eqm}) et absolue (eqm) d'estimation, obtenues pour 100 répétitions de validation croisée, pour différents diamètres de recensement.

de 7.5 à 22.5 cm, car la surface terrière moyenne pour les placettes passe de 37,2 à 28,2 $m^2 \cdot ha^{-1}$. La surface terrière sert en pratique à quantifier l'espace occupé par la végétation forestière. En augmentant le diamètre de recensement une partie de cette végétation est occultée alors qu'elle est potentiellement détectée par les impulsions laser. La précision traduit la qualité de la mise en relation entre ce qui est détecté par le scanner et ce que mesure le forestier. Les tiges de 7,5 à 17,5 cm représentent en effet presque 5 $m^2 \cdot ha^{-1}$ en moyenne, et autant pour les tiges de 17,5 à 22,5 cm. Le fait que l'erreur reste stable en absolu alors qu'elle augmente en relatif suggère que les modèles peinent à séparer la contribution des petites tiges de celle des grosses tiges à la surface terrière. Parmi les métriques sélectionnées (tableau 4), figure toujours la densité de la moitié supérieure de la canopée $d_{0,5}$, ainsi que deux métriques RapidEye, dont une est remplacée par la valeur moyenne du MNH pour d supérieur ou égal à 20 cm. Cette métrique représente la hauteur moyenne du toit de la canopée, qui est constituée par les arbres dominants et co-dominants. Dans un peuplement irrégulier, ceux-ci représentent l'essentiel de la surface terrière, a fortiori quand les tiges de faible diamètre ne sont pas comptabilisées.

Pour la densité de tiges, quand le diamètre passe de 7,5 à 17,5 cm, les erreurs moyennes relative et absolue diminuent respectivement de 23,7 à 18% et de 203

à 73 ha^{-1} . L'erreur moyenne relative augmente ensuite à 23,7% pour 22,5 cm, alors que l'erreur moyenne absolue se stabilise à 66 ha^{-1} . La densité de tiges est fortement influencée par les petits arbres, avec une moyenne de 860, 490 et 280 ha^{-1} pour des diamètres précomptables de 7,5, 15 et 22,5 cm. Par contre, la contribution des petites tiges au nuage de point est faible dans un peuplement irrégulier où le signal laser est fortement atténué par les arbres dominants de plus fort diamètre. Ceci explique le fait que la précision des prédictions est moins bonne quand les petites tiges sont comptées. Le changement de tendance à partir de 17,5 cm s'expliquerait alors par le fait que les arbres d'un diamètre supérieur constituent la partie haute de la canopée et sont donc bien représentés dans le nuage de points mais absents de l'inventaire terrain. Les métriques sélectionnées comportent toujours un percentile 80 de la hauteur des points et une métrique de densité de la partie supérieure de la canopée (sauf pour le diamètre précomptable de 22.5 cm). Comme pour la surface terrière, les métriques issues du MNH sont sélectionnées lorsque le diamètre précomptable augmente. Le taux de trouées ($MNH_{>2}$) est sélectionné pour 12,5 et 15 cm, alors que MNH_m est sélectionné pour les valeurs supérieures.

Pour le diamètre moyen, l'erreur moyenne relative est autour de 12% pour des diamètres précomptables de 7,5 à 12,5 cm puis descend à 8,2% pour 22,5 cm. Pour l'er-

reur absolue, les valeurs sont entre 2,9 et 3,3 cm, sauf pour un diamètre précomptable de 7,5 cm avec 2.5 cm. Lorsqu'on augmente le seuil de recensement, le diamètre moyen tend à se rapprocher du diamètre des plus gros arbres de la placette, c'est-à-dire du diamètre dominant. Pour d inférieur ou égal à 10 cm les métriques sélectionnées sont les mêmes (h_{s99} , i_{sm} et $pir.mean$). Pour d entre 12,5 et 20 cm, la métrique d'intensité LiDAR est conservée et associée à des percentiles de hauteur (20 et 80%). Pour $d = 22.5$ cm, deux métriques de densité correspondant à la partie haute de la canopée sont sélectionnées et associées au percentile de hauteur h_{s80} . Ces changements de métriques sont plus difficiles à interpréter.

4.2.3. Erreur de géoréférencement

Si les métriques sélectionnées dans les modèles varient peu selon les différents tirages au sort lorsque l'erreur est faible, les combinaisons obtenues sont plus variables pour des erreurs plus importantes. Pour la surface terrière, les précisions apparente et réelle restent similaires pour σ jusqu'à 4 m (figure 5). À partir de 4 m, la précision diminue et pour 8 m, la précision apparente sous-estime la précision réelle (18,2% contre 17,5% en moyenne). Pour la densité de tiges, des tendances similaires se dessinent. L'erreur apparente augmente en moyenne légèrement avec l'erreur de position, et une forte dégradation s'opère pour $\sigma = 8$ m. L'erreur apparente reste cependant peu affectée, avec des valeurs moyennes pour CV_{eqm} entre 23,6 et 24,4%. Pour le diamètre moyen, les précisions moyennes apparente et réelle sont très similaires. Elles sont légèrement et progressivement dégradées avec l'augmentation de l'erreur de positionnement (de 11,6 à 12,6%).

Pour tous les paramètres forestiers, la précision apparente se dégrade en moyenne pour des erreurs d'écart-type supérieur à 4 m. Pour des valeurs inférieures, l'apparente stabilité pourrait être liée au fait que l'erreur ajoutée est d'un ordre de grandeur proche de celle des données de référence. Par ailleurs, l'utilisation des coordonnées avec erreur pour la validation conduit en moyenne à sous-estimer la précision pour les fortes erreurs de positionnement, ce qui est conforme aux résultats obtenus sur deux autres sites en France (Monnet et Mermin, 2014). Dans tous les cas il s'avère que la variabilité de l'estimation de la précision augmente fortement avec l'erreur de positionnement.

4.2.4. Nombre de placettes de calibration

On observe la même tendance pour tous les paramètres forestiers (figure 6) : la moyenne et la variabilité de l'erreur augmentent quand le nombre de placettes diminue. Il est possible que le nombre initial de 68 placettes soit déjà trop faible pour vraiment pouvoir quantifier ce phénomène.

4.3. Stratification des modèles

Les métriques utilisées pour un même paramètre forestier changent selon les groupes (tableau 5). En général, l'erreur totale diminue quand on passe d'un modèle

strate		métriques	R_a^2
<i>G</i>		$d_{0,5} + ndve.m + re.sd$	73,8
<i>N</i>		$h_{180} + i_{fm} + d_{0,5}$	81,2
<i>D</i>		$h_{s99} + i_{sm} + pir.m$	74,3
<i>G</i>	S.a	$d_{0,5} + ndvi.sd + pir.sd$	79,8
<i>N</i>	S.a	$h_{160} + d_{0,5}$	83,7
<i>D</i>	S.a	$h_{160} + d_{0,5} + ndve.m$	77,8
<i>G</i>	S.fi	$h_{f20} + h_{am} + re.sd$	75,0
<i>N</i>	S.fi	$h_{180} + d_{0,5}$	71,2
<i>D</i>	S.fi	$h_{180} + i_{lsd} + ndvi.sd$	79,6
<i>G</i>	E.m	$d_{0,5} + pir.sd + re2r.sd$	74,9
<i>N</i>	E.m	$h_{140} + d_{0,5}$	81,3
<i>D</i>	E.m	$h_{140} + d_{0,5} + pir.m$	72,7
<i>G</i>	E.r	$d_{0,67} + MNH_{>2} + ndve.m$	84,3
<i>N</i>	E.r	$h_{s40} + i_{lm} + d_{0,67}$	86,0
<i>D</i>	E.r	$h_{s60} + h_{s80} + d_{0,5}$	77,7

Tableau 5: Influence de la stratification des modèles : métriques sélectionnées et R_a^2 du modèle pour différentes strates. Stratification « Structure » : S.i (irrégulier), S.a (autre). Stratification « Essence » : E.m (mélange ou feuillus), E.r (résineux).

global à des modèles stratifiés. Pour la surface terrière, elle est de 16.4, 14.2 et 15.3% respectivement pour les modèles global, stratifié selon l'essence et stratifié selon la structure. Pour la densité de tiges, les valeurs sont de 23.6, 22.8 et 24.9%. Pour le diamètre moyen, elles sont de 11.6, 10.8 et 10.4%. Étant donné les résultats précédents montrant la forte variabilité des estimations d'erreur obtenues pour des nombres de placettes faibles, ces valeurs doivent être considérées avec précaution. Il semble cependant que la stratification selon l'essence bénéficie aux modèles de surface terrière et de densité de tiges, alors qu'on obtient de meilleurs résultats avec la structure pour le diamètre moyen.

5. Conclusion

À partir de ce jeu de données issues de forêts de montagne, l'analyse de sensibilité de l'approche surfacique a permis de confirmer les tendances constatées dans la littérature, mais aussi d'apporter de nouvelles informations.

L'utilisation de l'information spectrale, que ce soit via les valeurs d'intensité ou la fusion avec les données RapidEye, est bénéfique aux modèles d'estimation pour tous les paramètres testés. Cependant il faudra vérifier que l'utilisation des données d'intensité LiDAR ne génère pas lors de la cartographie des erreurs fortement corrélées avec l'angle de scan. Pour les données RapidEye qui sont a priori plus homogènes, il faudra cependant porter attention à l'effet des métriques qui ne sont pas des ratios et donc sensibles aux variations d'illumination dues à la topographie.

Le diamètre précomptable, qui influe fortement sur la vision de terrain, a un impact indirect sur la précision atteinte par les modèles, notamment dans le cas de peuplements irréguliers où les jeunes arbres représentent

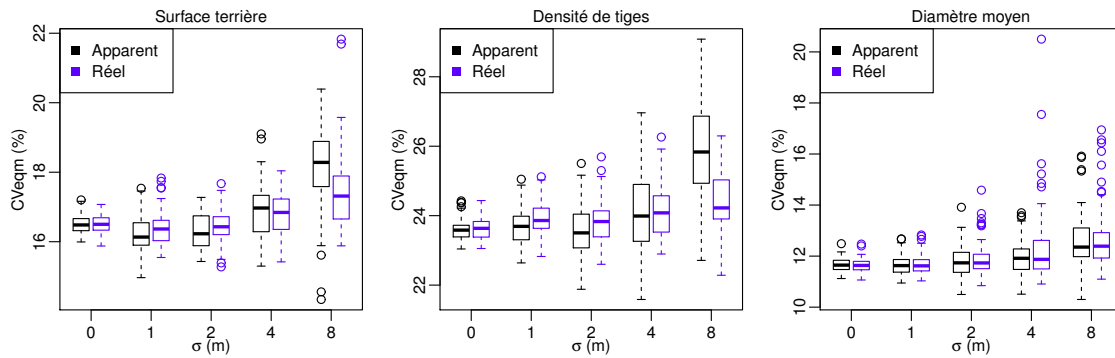


FIGURE 5: Boîtes à moustaches des valeurs d'erreur relative (CV_{eqm}) d'estimation, obtenues pour différentes valeurs d'écart-type σ avec 50 simulations d'une erreur de géoréférencement de distribution normale $\mathcal{N}(0, \sigma^2)$ pour les coordonnées de chaque placette.

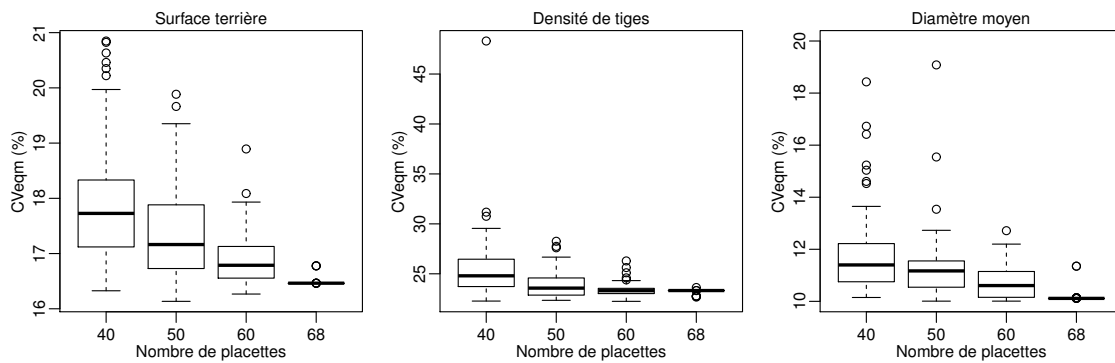


FIGURE 6: Boîtes à moustaches des valeurs d'erreur relative (CV_{eqm}) d'estimation, obtenues par validation croisée pour 100 tirages au sort d'un nombre n de placettes.

des proportions variables du matériel végétal. Dans le cas de forêts équiennes, il serait probablement plus judicieux d'adopter des diamètres de recensement adaptés à chaque classe d'âge et de calibrer des modèles stratifiés selon ce critère. Le rayon de la placette a également un effet important, via la limitation des effets de bord. Comme pour le diamètre de recensement il pourrait être adapté selon la structure des peuplements et leur homogénéité. Une solution possible est le protocole basé sur des placettes concentriques, comme utilisé en France pour l'Inventaire Forestier National (Robert et al., 2010). Un tel type d'inventaire pourrait être simulé à partir des données de cette étude afin de comparer la perte éventuelle de précision avec le gain de temps que permet ce protocole.

Les critères précédents influencent surtout la valeur moyenne de l'erreur estimée par les répétitions de validation croisée. L'erreur de positionnement et le nombre de placettes influencent également sa variabilité, c'est à dire qu'ils dégradent la précision en moyenne, mais également la qualité de son estimation. L'effet de l'erreur de positionnement est faible pour des erreurs d'écart-type inférieur à 4 m, et ce d'autant que l'utilisation de coordonnées faussées lors de la validation conduit à sous-estimer la précision. Il est donc crucial lors des relevés de terrain d'utiliser du matériel professionnel permettant d'atteindre un géoréférencement précis. Le nombre de

placettes de terrain est également un point important car il influe fortement sur la variabilité de l'estimation de l'erreur. Le nombre optimal dépend probablement de la variabilité des peuplements forestiers et de celle des données LiDAR. La stratification des modèles selon des critères permettant d'avoir des strates plus homogènes du point de vue de l'interaction entre les impulsions laser et la végétation permet d'améliorer la précision, mais le compromis optimal entre le nombre de strates et l'effectif de placettes par strate semble difficile à trouver a priori, et ce d'autant plus qu'on a souligné les problèmes liés aux faibles effectifs.

Les résultats obtenus pour cette zone d'étude ne sont pas généralisables, cependant ils permettent de comparer l'impact des différents critères sur un cas unique. Il serait intéressant de reproduire cette méthodologie sur un massif forestier présentant une structure et des essences différentes pour comparer les tendances. Par ailleurs, l'utilisation de répétitions de validation croisée a mis en évidence le fait que certains critères, au-delà de l'effet sur la précision, avait également un impact important sur la qualité de l'estimation de la précision.

Remerciements

Cette étude a été financée par la Commission Européenne (programme Alpine Space, projet 2-3-2-FR NEW-

FOR) et par l'Agence Nationale de la Recherche (projet ANR-2010-BIOE-008 FORESEE).

Références

- Bouvier, M., Durrieu, S., Fournier, R. A., Renaud, J.-P., 2015. Generalizing predictive models of forest inventory attributes using an area-based approach with airborne lidar data. *Remote Sensing of Environment* 156, 322–334.
- Frazer, G. W., Magnussen, S., Wulder, M. A., Niemann, K. O., 2011. Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of lidar-derived estimates of forest stand biomass. *Remote Sensing of Environment* 115 (2), 636–649.
- Gobakken, T., Næsset, E., 2009. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research* 39 (5), 1036–1052.
- Hollaus, M., Dorigo, W., Wagner, W., Schadauer, K., Höfle, B., Maier, B., 2009. Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. *International Journal of Remote Sensing* 30 (19), 5159–5175.
- Holmgren, J., Nilsson, M., Olsson, H., 2003. Simulating the effects of lidar scanning angle for estimation of mean tree height and canopy closure. *Canadian Journal of Remote Sensing* 29 (5), 623–632.
- Hudak, A. T., Crookston, N. L., Evans, J. S., Falkowski, M. J., Smith, A. M. S., Gessler, P. E., Morgan, P., 2006. Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. *Canadian Journal of Remote Sensing* 32 (2), 126–138.
- Junttila, V., Finley, A. O., Bradford, J. B., Kauranne, T., 2013. Strategies for minimizing sample size for use in airborne lidar-based forest inventory. *Forest Ecology and Management* 292, 75–85.
- Le Ferrec, A., 2011. Élaboration d'une méthode d'initialisation de simulations de dynamique forestière à l'échelle du massif dans le secteur des quatre montagnes, Vercors. Rapport de Master, ENITA Bordeaux.
- Monnet, J.-M., Chanussot, J., Berger, F., 2011. Support vector regression for the estimation of forest stand parameters using airborne laser scanning. *IEEE Geoscience and Remote Sensing Letters* 8 (3), 580–584.
- Monnet, J.-M., Mermin, É., Chanussot, J., Berger, F., 2010. Estimation of forestry parameters in mountainous coppice stands using airborne laser scanning. Dans : *Proceedings of Silvilar 2010, 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems*, Freiburg, Allemagne, 14-17/09/2010, 109–116.
- Monnet, J.-M., Mermin, É., 2014. Cross-correlation of diameter measures for the co-registration of forest inventory plots with airborne laser scanning data. *Forests* 5 (9), 2307–2326.
- Montaghi, A., 2013. Effect of scanning angle on vegetation metrics derived from a nationwide airborne laser scanning acquisition. *Canadian Journal of Remote Sensing* 39, S152–S173.
- Morsdorf, F., Frey, O., Meier, E., Itten, K. I., Allgower, B., 2008. Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *International Journal of Remote Sensing* 29 (5), 1387–1406.
- Niska, H., Skön, J. P., Packalén, P., Tokola, T., Maltamo, M., Kohlainen, M., 2010. Neural networks for the prediction of species-specific plot volumes using airborne laser scanning and aerial photographs. *IEEE Transactions on Geoscience and Remote Sensing* 48 (3), 1076–1085.
- Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80 (1), 88–99.
- Næsset, E., 2004a. Accuracy of forest inventory using airborne laser scanning : Evaluating the first nordic full-scale operational project. *Scandinavian Journal of Forest Research* 19 (6), 554–557.
- Næsset, E., 2004b. Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sensing of Environment* 91 (2), 243–255.
- 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.
- Packalen, P., Strunk, J. L., Pitkanen, J. A., Temesgen, H., Maltamo, M., 2015. Edge-tree correction for predicting forest inventory attributes using area-based approach with airborne laser scanning. *IEEE Journal of Selected Topics In Applied Earth Observations and Remote Sensing* 8 (3), 1274–1280.
- Packalén, P., Maltamo, M., 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. *Remote Sensing of Environment* 109 (3), 328–341.
- Penner, M., Pitt, D. G., Woods, M. E., octobre 2013. Parametric vs. nonparametric lidar models for operational forest inventory in boreal Ontario. *Canadian Journal of Remote Sensing* 39 (5), 426–443.
- Redon, M., 2012. Biodiversité potentielle dans les forêts du vercors : une approche hiérarchique pour la conservation des espaces forestiers. Thèse de doctorat, université de Grenoble.
- Robert, N., Vidal, C., Colin, A., Jean-Christophe, H., Hamza, N., Cluzeau, C., 2010. National forest inventories reports : France. Dans : Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R. E. (Eds.), *National Forest Inventories : Pathways for Common Reporting*. Springer, pp. 207–221.
- Ruiz, L. A., Hermosilla, T., Mauro, F., Godino, M., 2014. Analysis of the influence of plot size and lidar density on forest structure attribute estimates. *Forests* 5 (5), 936–951.
- Strunk, J., Temesgen, H., Andersen, H. E., Flewelling, J. P., Madsen, L., 2012. Effects of lidar pulse density and sample size on a model-assisted approach to estimate forest inventory variables. *Canadian Journal of Remote Sensing* 38 (5), 644–654.
- Takahashi, T., Awaya, Y., Hirata, Y., Furuya, N., Sakai, T., Sakai, A., 2010. Stand volume estimation by combining low laser-sampling density lidar data with QuickBird panchromatic imagery in closed-canopy Japanese cedar (*cryptomeria japonica*) plantations. *International Journal of Remote Sensing* 31 (5), 1281–1301.
- Wagner, W., 2010. Radiometric calibration of small-footprint full-waveform airborne laser scanner measurements : Basic physical concepts. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (6), 505–513.
- Wallerman, J., Holmgren, J., 2007. Estimating field-plot data of forest stands using airborne laser scanning and SPOT HRG data. *Remote Sensing of Environment* 110 (4), 501–508.
- Watt, M. S., Adams, T., Aracil, S. G., Marshall, H., Watt, P., 2013. The influence of lidar pulse density and plot size on the accuracy of New Zealand plantation stand volume equations. *New Zealand Journal of Forestry Science* 43, 15.
- White, J., Wulder, M., Varhola, A., Vastaranta, M., Coops, N.C. and Cook, B., Pitt, D., Woods, M., 2013. A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach.