

APPORT DE VARIABLES ISSUES DE LA SEGMENTATION D'ARBRES SUR DONNEES LIDAR AEROPORTE POUR L'ESTIMATION DES VARIABLES DENDROMETRIQUES DE PLACETTES FORESTIERES

Ana Cristina André^{1,2,3}, Jean-Pierre Renaud³, Cédric Véga⁴, Alain Munoz⁵, Jérôme Bock⁵, Laurent Saint-André⁶

1 : Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, Brasil
(anacristinandre.22@gmail.com)

2 : AgroParistech, 5400 Nancy

3 : Office National des Forêts, Département RDI, 54000 Nancy
(jean-pierre.renaud-02@onf.fr)

4 : IGN, Laboratoire de l'inventaire forestier, 54000 Nancy

5 : Office National des Forêts, Département RDI, 73023 Chambéry

6 : INRA, BEF, 54280 Champenoux

Résumé

Les récents développements dans le domaine du LiDAR aéroporté à balayage offrent de nouvelles possibilités pour estimer et cartographier différents attributs forestiers et améliorer la précision des inventaires. Dans cette étude, deux méthodes de segmentation des nuages de points LiDAR ont été utilisées afin d'obtenir de nouvelles variables et d'évaluer leur apport sur la précision d'estimation de la hauteur dominante, de la surface terrière et du volume, sur trois sites forestiers contrastés. L'originalité de cette approche vient de l'utilisation conjointe de variables de segmentation et de métriques standards de distribution de hauteur dans le cadre d'une approche « placette ». Les précisions obtenues à l'aide des variables de segmentation ont été comparées à celles obtenues par l'approche « placette » initialement proposée par Næsset (1997). Les résultats confirment l'avantage de combiner les 2 méthodes. Pour la surface terrière et le volume, des R^2 de 0.92 et 0.95, et une erreur quadratique moyenne de 4.6 m²/ha (16%) et 72.4 m³/ha (17%) ont été respectivement obtenus. Pour la hauteur dominante, l'erreur du modèle avoisine l'erreur de mesure sur le terrain (estimée inférieure à 1 m). Il semble donc que l'intégration de variables issues de la segmentation de houppiers permette d'améliorer l'estimation des attributs forestiers. La robustesse de ces résultats reste à évaluer sur un plus grand nombre de peuplements et sur des configurations d'acquisition LiDAR variées.

Mots clés : LiDAR, approche placette, segmentation, modèle d'estimation, surface terrière, volume

Abstract

Recent developments in the field of airborne laser scanning offer new possibilities to map different forest attributes and improve forest inventory accuracy. In this work, two methods of LiDAR point clouds segmentation were used to obtain new variables and assess the gain in accuracy to estimate dominant height, basal area and volume, in three contrasted forest stands. The originality of this approach is to combined segmentation methods and standards metrics based on height distribution summarized at the "plot" level. The information obtained using the segmentation variables were compared to those obtained by the "standard" method initially proposed by Næsset (1997). The results confirm the advantage of combining the two methods. For basal area and volume, R^2 of 0.92 and 0.95, and mean square errors of 4.6 m²/ha (16%) and 72.4 m³/ha (17%) were obtained respectively. For dominant height, model error is in the range of field measurement error (estimated to be less than 1 m). So, it seems that segmentation allows an improvement of forest attributes estimations. The robustness of these findings remains to be evaluated on a larger number of stands and varied configurations of LiDAR acquisitions.

Keywords : LIDAR, area based approach, segmentation methods, estimation model, basal area, volume

1. Introduction

Une meilleure connaissance de la ressource forestière au niveau d'un bassin d'approvisionnement permet d'en optimiser la production. Les récents développements dans le domaine du laser aéroporté à balayage (LiDAR = light detection and ranging) et de l'imagerie aérienne offrent de nouvelles possibilités pour cartographier différents paramètres forestiers et améliorer la précision des inventaires (Næsset *et al.*, 2013 ; White *et al.*, 2013). L'accès facilité à la connaissance des hauteurs de peuplements qu'offre le LiDAR, représente probablement un virage technologique qui pourrait nécessiter une révision des modèles allométriques standards en foresterie (Picard *et al.* 2012). Cette connaissance « nouvelle » des hauteurs devient disponible pour l'ensemble d'un couvert ou d'un massif, ouvrant la porte à des modèles plus globaux, dépassant même l'échelle de l'arbre ou de la placette.

Depuis la fin des années 1990, un grand nombre d'études a montré le potentiel du LiDAR pour fournir des estimations précises d'attributs forestiers, comme la hauteur des arbres, le volume de bois, ainsi que d'autres paramètres de structure ou de distribution (e.g. Næsset, 1997 ; Magnussen and Boudewyn, 1998 ; Means *et al.*, 2000 ; Zimble *et al.*, 2003 ; Maltamo *et al.*, 2006 ; Vauhkonen *et al.*, 2011). Ces développements, obtenus largement sur des forêts de conifères, représentent une avancée marquante dans le monde de la télédétection forestière, car ils fournissent des informations qui, sur le terrain, sont difficilement mesurées avec précision, telles que la hauteur des arbres ou la taille des couronnes de la canopée. Cette capacité du LiDAR à mesurer ces paramètres en fait un outil permettant d'améliorer le travail des gestionnaires forestiers (Vastaranta *et al.*, 2013 ; White *et al.*, 2013). Les systèmes LiDAR permettent d'obtenir des nuages de points 3D qui décrivent la structure de la surface terrestre, incluant le sol et les objets qu'il porte.

La perméabilité des couverts forestiers aux signaux LiDAR permet de caractériser la distribution verticale de la végétation et d'échantillonner le niveau du sol avec une précision suffisante pour générer des modèles numériques de terrain de bonne précision, qui serviront de référence pour calculer la hauteur des éléments du sursol et analyser leur structure. L'analyse des structures de végétation et l'estimation de paramètres dendrométriques à partir de données LiDAR peut être envisagée à l'échelle de la placette ou de l'arbre.

Les approches « placettes » ou « surfaciques », visent à mettre en relation des métriques LiDAR, paramètres statistiques dérivés de la distribution 3D du nuage de points LiDAR, avec des mesures terrain de référence, afin d'établir des modèles empiriques de prévisions. De telles approches ont d'ores et déjà été intégrées dans des inventaires nationaux, notamment dans les pays nordiques (p.ex. Finlande) (Naesset *et al.* 2004). Ces approches souffrent toutefois d'un manque de généralité, et nécessitent un calibrage des modèles pour chaque acquisition et type de peuplement. Les approches « arbre », permettent quant à elle d'obtenir des informations sur les individus d'une zone donnée, et sont à ce titre, plus proches des méthodes d'inventaire de terrain. Cependant, la généralisation de telles approches se heurte à différents problèmes. La détection des arbres individuels à partir d'une segmentation du nuage de points LiDAR nécessite des densités de points élevées (un minimum de 2 à 10 pts/m² selon Kaartinen *et al.* 2012) qui impactent les coûts d'acquisition. Par ailleurs, la qualité des segmentations est fortement dépendante de la performance de l'algorithme de segmentation et du type de forêt (Kaartinen *et al.* 2012 ; Vauhkonen *et al.* 2011). Et surtout, le développement de modèles d'estimation de paramètres d'arbre individuel demande des informations additionnelles sur les arbres, comme l'espèce, qui sont difficiles à obtenir à partir de données LiDAR seules.

Malgré leurs divergences, ces deux approches offrent un certain nombre de complémentarités qui ont peu été étudiées (Lindberg *et al.* 2010; Breidenbach et Astrup 2014). En particulier, nous pensons que la combinaison de variables simples issues de la segmentation et de métriques standard donneraient des informations complémentaires qui, utilisées au niveau de la placette, permettrait d'améliorer la robustesse des modèles de prédiction de paramètres forestiers, sans pour autant nécessiter de calibrage important.

L'objectif de cet article est donc de proposer des indicateurs de structure de placettes à partir de produits de la segmentation des houppiers et d'évaluer l'apport de ces paramètres dans les estimations de variables dendrométriques.

2. Zone d'étude et données utilisées

Cette étude porte sur 3 zones forestières contrastées en France métropolitaine (Figure 2) :

1. Le site de l'Observatoire Permanent de l'Environnement (OPE) situé près de la commune de Bure (55), s'étend sur une superficie de 60 km² et porte une forêt feuillue dominée par le hêtre (*Fagus sylvatica*);

2. Le site de la Combe d'Aillon, situé dans le Parc Naturel Régional des Bauges, en Savoie (73), est une forêt de montagne au relief accidenté. Les peuplements irréguliers, majoritairement composés de sapins (*Abies alba*), couvrent une superficie d'environ 25 km² ;
3. Le site de St Symphorien, d'une surface d'environ 40 km², est situé sur le plateau Landais dans le département de la Gironde (33). Ses peuplements de pins maritimes (*Pinus pinaster*), issus de semis artificiels, poussent sur des sols sableux, pauvres et au relief plat.



Figure 1 : Localisation des 3 zones d'études : l'observatoire pérenne de l'environnement (OPE Bure) dans le Nord-Est ; La Combe d'Aillon dans les Alpes, et au sud-ouest, St Symphorien sur le plateau Landais.

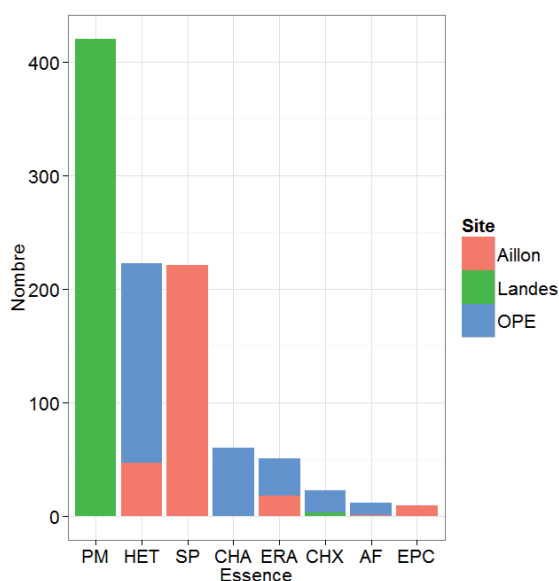


Figure 2 : Répartition du nombre de tiges par essence pour les placettes des trois sites d'étude. (PM : pin maritime ; HET : hêtre ; SP : sapin ; CHA : charmes ; ERA : érables ; CHX : chênes ; AF : autres feuillus ; EPC : épicéas)

Les données de terrain ont été obtenues sur des placettes circulaires de 15 m de rayon (horizontal) où tous les arbres de 17.5 cm de diamètre et plus ont été mesurés en circonférence à 1.3 m de hauteur. Pour évaluer la hauteur dominante (Ho) des placettes (700 m²), la hauteur des 6 plus gros arbres a été mesurée sur le terrain en faisant la moyenne de deux mesures prises dans des directions opposées, à l'aide d'un dendromètre Vertex. Pour le site Aillon, une seule mesure a été effectuée, en visée vers l'aval, en raison des fortes pentes. La surface terrière a été calculée directement à partir des circonférences. Les volumes (totaux et à découpe 7 cm), ont été obtenus en utilisant les équations du projet ANR Emerge, telles que décrites dans St-André *et al.* (2013). Enfin, la structure des diamètres observée sur le terrain a été résumée en calculant un indice de Gini (Valbuena *et al.* 2013) tel que :

$$CG = \sum_{j=1}^n \frac{(2j - n - 1)D_j}{D_j(n - 1)}$$

où j représente le rang en fonction de la taille en ordre ascendant; n le nombre total d'arbre, et D_j le diamètre à 1.3 m de l'arbre j . Cet indice est un descripteur de l'inégalité de la distribution des diamètres. Il varie de 0 à 1, où 0 signifie l'homogénéité parfaite et 1 l'hétérogénéité

totale. Ce coefficient est utilisé pour décrire différents types de structures forestières (Valbuena *et al.* 2013). La position des centres de placettes a été mesurée en utilisant un récepteur GNSS différentiel avec une précision estimée entre 1 et 4 m. La qualité du positionnement a été améliorée visuellement, en appariant les couronnes d'arbres observées sur le LiDAR avec les positions de ces mêmes arbres telles que dérivées de mesures terrain d'angle et de distance entre le centre de la placette et le tronc. Au total 32 placettes terrains ont été utilisées pour cette étude, 10 à l'OPE (Bure) mesurées en février 2010 ; 12 dans la forêt départementale de Combe d'Aillon, mesurées en avril 2011 ; et 10 dans les Landes, mesurées d'avril à juin 2011. Une description dendrométrique de ces différents sites est donnée au Tableau 1. Elle résume la hauteur dominante (Ho), la surface terrière (G), les volumes (total : Vtot et à découpe 7cm : V7) ainsi que l'indice de Gini pour chacun des sites d'étude.

Les acquisitions LiDAR ont été réalisées par la société Sintégra, dont le siège social est situé à Meylan (38). Le scanner utilisé est un Riegl LMS Q560 monté à bord d'un avion. Les vols sur Bure, Combe d'Aillon et Landes ont eu lieu le 6 octobre 2010, le 7 juillet 2011 et le 5 avril 2011 respectivement, à une altitude moyenne de vol de 550 m. Le calcul et le pré-traitement des données LiDAR ont été effectués par le prestataire à l'aide des suites logicielle Riegl et Terrasolid.

Paramètre	Site	Min	Max	Médiane	Moyenne	Coeff. var. (%)
Ho (m)	Aillon	21.5	35.4	32.6	31.4	14
	OPE	19.3	40.9	22.2	23.8	27
	Landes	15.2	30.5	22.2	22.9	26
G (m ² /ha)	Aillon	18.2	66.9	40.5	43.5	31
	OPE	7.6	40.6	15.3	19.2	49
	Landes	0.0	42.2	22.9	19.4	64
Vtot (m ³ /ha)	Aillon	238	1157	651	713	37
	OPE	101	914	202	289	80
	Landes	0	584	235	220	76
V7 (m ³ /ha)	Aillon	200	1055	563	630	39
	OPE	68	776	153	220	92
	Landes	0	543	210	200	78
Gini	Aillon	0.2	0.3	0.2	0.2	19
	OPE	0.1	0.3	0.2	0.2	37
	Landes	0.1	0.2	0.1	0.1	43

Tableau 1 : Résumé par site des variables dendrométriques mesurées sur le terrain. Ho : hauteur moyenne des 6 plus gros arbres par placette de 15m de rayon ; G : surface terrière des arbres de diamètre ≥ 17.5 cm ; Vtot : volume total des arbres de diamètre ≥ 17.5 cm ; V7 : volume « bois fort tige » à la découpe 7 cm de diamètre, pour les arbres de diamètre ≥ 17.5 cm ; Gini : indice d'hétérogénéité des diamètres à 1.3 m. (Coeff.var représente le coefficient de variation exprimé en %.)

4. Méthodologie

3.1. Métriques standard calculés à partir des données LiDAR (Std)

Préalablement au calcul des métriques, une première étape consiste à normaliser le nuage de points LiDAR en remplaçant l'élévation de chaque point par sa hauteur au-dessus du sol. Pour cela, la distance verticale entre chaque point LiDAR et un TIN (« triangular irregular network ») créé à partir des points classés sol est calculée en utilisant le logiciel Lastools (rapilasso GmbH) (Figure 3).

Une fois cette normalisation réalisée, les points correspondants à l'emprise exacte des placettes terrain sont extraits sur 15m de rayon. Après cette extraction, différentes métriques sont calculées à l'aide du logiciel Lasmanager ©ONF (plus de 500 métriques). Pour les premiers retours, les derniers et les retours intermédiaires, des indicateurs de moments (moyenne des hauteurs, écart-type, skewness...) ainsi que des indicateurs de distribution de hauteurs (déciles, densités de points par déciles ou par tranche fixe de hauteur, taux de pénétration) sont calculés. Quelques métriques basées sur l'extraction de maxima locaux sont également calculées (voir p.ex. Bock *et al.* 2009).

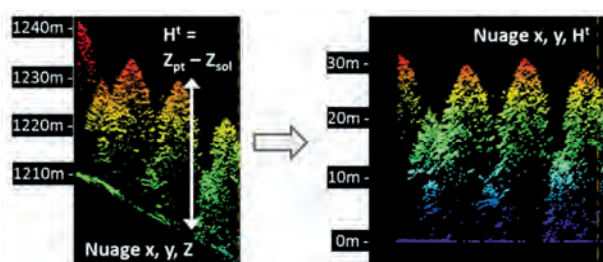


Figure 3 : Nuage de points LiDAR dans une zone de pente, illustrant la normalisation des données (en altitudes à gauche) par soustraction avec le TIN créé à partir des points classés sol pour obtenir des hauteurs (à droite).

3.2. Segmentation et métriques associées

Deux méthodes de segmentation ont été utilisées dans le cadre de cette étude.

La première méthode (BV), utilise des modèles numériques de hauteur (MNH) afin d'effectuer une segmentation de type bassin versant à l'aide des outils hydrologiques de l'extension « Spatial Analyst » d'ArcGis (méthode et outil développés au sein du département R&D de l'ONF). Il est important de noter que le MNH en entrée est généré à partir d'un jeu de différents MNH créés par triangulation des premiers retours LiDAR : en prenant en compte tous les premiers retours, puis en excluant les points au-dessous de certains seuils de hauteur (p.ex. 5m, 10m, 15m). Le MNH raster final est obtenu par combinaison de ce jeu de données en conservant la hauteur maximale pour chaque cellule (Khosravipour *et al.* 2013).

A partir du MNH résultant, deux segmentations différentes ont été réalisées. La première (v1) a été opérée directement sur le MNH. La seconde (v2) a été réalisée en utilisant une version lissée du MNH, en appliquant un filtre moyen 3x3. La segmentation se fait sur les MNH inversés (outil « Negate » d'ArcGis) en utilisant les outils « Basin » et « FlowDirection » d'ArcGis.

La seconde méthode de segmentation (P) est basée sur une analyse directe du nuage de points en Z (sans normalisation du nuage de points). Cette segmentation est multi-échelles et multi-critères. La normalisation ($H = Z - MNT$) est réalisée après la segmentation, en calculant tout d'abord la hauteur H de l'apex des arbres ; puis la hauteur des points de chaque couronne est déduite de l'altitude de référence de la « base » de l'arbre. Cette méthode permet de limiter les déformations des houppiers liées au relief. Les détails de l'algorithme sont donnés dans Véga *et al.* (2014). La figure 4 présente le résultat de ces segmentations pour une placette donnée.

Ces deux méthodes donnent à la fois des segments de superficies connues et des points de hauteurs maximales par segment, qui sont assimilés à des « apex ». Ces surfaces (S) et ces hauteurs d'apex (H), ainsi que leurs produits ($S \cdot H$) ont été utilisés pour générer différentes métriques par placettes. Des moyennes, des écart-types, des sommes, ainsi que des indices de Gini ont été calculés pour ces paramètres, ainsi que les modes de leurs distributions, calculés suivant la méthode décrite par Jaskierniak *et al.* (2011) avec la fonction `gamlssMX` de la librairie « `gamlss.mx` » de R.

3.3. Sélection des variables statistiques

Pour établir des modèles de prédiction de H_0 , G, V_{tot} et V_7 , sachant qu'un nombre restreint de paramètres doit être retenu parmi un grand nombre de métriques, nous avons recherché la meilleure combinaison de variables pour un nombre fixe de paramètres (1 ou 3) en utilisant la fonction « `regsubsets` » de la librairie « `leaps` » de R. Nous nous sommes limités à 3 paramètres, afin de limiter les problèmes de multicollinéarité et toutes les combinaisons de variables (méthode exhaustive) ont été comparées.

Nous avons retenu les modèles sur la base du critère d'AIC (Akaike 1974), qui permet de sélectionner le modèle le plus parcimonieux (dont l'AIC est la plus faible). Comme la multicollinéarité est très fréquente dans les métriques LiDAR, nous avons utilisé la fonction « `vif` » de la librairie « `car` » pour s'assurer de l'indépendance des variables utilisées dans les modèles multivariés ($VIF < 4$). La précision des modèles retenus a été évaluée par validation croisée en utilisant la fonction « `cv.lm` » de la librairie « `DAAG` ». Les observations ont été séparées aléatoirement en 3 groupes pour estimer l'erreur de prédiction. L'erreur moyenne de validation croisée tend à être moins optimiste qu'en calibration.

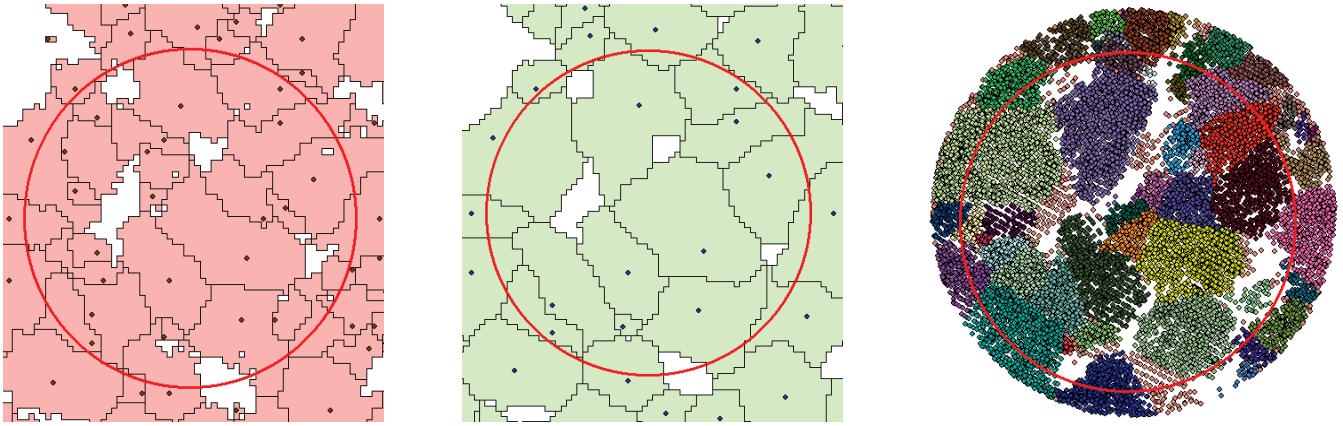


Figure 4 : Représentation des 2 méthodes de segmentations abordées dans cette étude. La première image (à gauche) montre la segmentation BV sans lissage préalable du MNH (BVv1), la deuxième la segmentation BV avec lissage du MNH (BVv2) et la dernière la segmentation P.

4. Résultats et discussion

Nous avons d'abord évalué la qualité des 3 méthodes de segmentation en fonction des omissions et des fausses détections des arbres par comparaison avec les données de terrain (Tableau 2). On s'aperçoit que la méthode des bassins versants non filtrés (BVv1) conduit à un nombre important de fausses détections (exactitude de 42%), et que même en filtrant (BVv2), le nombre de fausses détections diminue, mais les omissions restent plus importantes que par la méthode Ptree (P) qui obtient les meilleurs scores en termes d'omissions et de fausses détections. Cette méthode de segmentation se rapproche donc le plus de la réalité de terrain.

Pour l'estimation de Ho, les résultats montrent que ce paramètre est facilement prédit par LiDAR. Le Tableau 3 présente des modèles aux R^2 élevés (≥ 0.98) et aux erreurs de validation croisées faibles ($\leq 4\%$). Par analogie avec la mesure de terrain, on aurait pu penser que la moyenne des 6 plus hauts segments serait ressortie parmi les meilleurs modèles de prédiction de Ho, ce qui n'est pas le cas. Le meilleur modèle à une seule variable utilise le 90^{ième} percentile de hauteur des derniers retours supérieurs à 2 m (H90_1st.sup2m). La valeur moyenne de cette métrique est de 24.6 m pour l'ensemble de notre dispositif alors que pour les premiers retours cette moyenne est de 26.7 m, la différence entre ces métriques est donc très ténue.

Dans la littérature (e.g. Naesset *et al.* 2004, Magnussen *et al.* 2012), la hauteur maximale est rarement utilisée pour prédire la hauteur des peuplements, car cette métrique représente la queue de la distribution et est plus instable que le 90^{ième} percentile par exemple. Cette instabilité explique peut-être pourquoi les métriques basées sur les hauteurs des « apex » des segments ne sont pas ressorties en premier pour prédire Ho. La variable de segmentation qui entre dans un modèle à 3 variables est le coefficient d'hétérogénéité des hauteurs des « apex » extraits par la méthode des Bassins versants non-filtrés (BVv1_H.gini). Dans notre cas, utiliser un modèle à 3 variables pour prédire Ho n'améliore que de très peu l'erreur moyenne du modèle (moins de 30cm) comparativement à un modèle à une seule variable.

Pour Ho, une erreur d'environ 0.9 m avoisine probablement l'erreur de mesure terrain.

En 2012, Järnstedt *et al.* (2012) ont comparé la précision d'estimation de divers attributs forestiers à partir de mesures LiDAR et photogrammétriques. Ils ont obtenu une précision voisine pour l'estimation de Ho à partir de ces 2 sources de données. Ils ont également souligné la contribution des derniers retours pour estimer Ho.

Méthode	N arbres	Recall	Exactitude
BVv1	670	81%	42%
BVv2	670	72%	72%
P	670	82%	89%

Tableau 2 : Qualité de la détection des arbres (par comparaison avec les relevés de terrain), en fonction des méthodes de segmentation. Les pourcentages de « recall » et d'exactitude sont calculés selon Véga *et al.* (2014). (Recall prend en compte les bonnes détections divisées par la somme des bonnes détections et des omissions, alors que pour l'Exactitude, la division est se fait sur la base de la somme des bonnes et des fausses détections.)

Les résultats de modélisation de G, Vtot et V7 sont également présentés au Tableau 3. Pour ces 3 paramètres, les meilleurs modèles uni-variés se basent sur la somme par placette du produit entre la surface des segments et leur hauteur d'« apex » (BVv1_H.S.sum). Ce résultat est à la fois intéressant et contradictoire. En effet, il laisse supposer que la somme des volumes de segments est bien en lien étroit avec la biomasse observée sur le terrain. En ce sens, ce résultat est conforme avec le fait que les arbres qui ont sur le terrain un large houppier et qui sont hauts, ont souvent une biomasse importante. Par contre, il est étonnant de constater que la méthode qui génère le plus de fausses détections (BVv1, Tableau2) puisse prédire le mieux ces paramètres dendrométriques. On se serait attendu à ce que la meilleure segmentation produise les meilleurs résultats. Ce résultat suggère donc que c'est probablement le volume sous la surface du MNH qui a un rôle important à jouer pour prédire la biomasse et que le processus de segmentation joue un rôle secondaire. Les erreurs moyennes de validation croisée associées à cette métrique varient entre 25% pour G et Vtot et 30% pour V7. Les meilleurs modèles sont toutefois obtenus avec 3 variables (avec des erreurs de validations croisées inférieures à 19%).

Variable	Equations	R ²	AIC	Erreur (EQM)	% Erreur	% Erreur.cv
Ho (m)	H90_lst.sup2m	0.98	81	0.9	3	4
	BVv1_H.gini + H95_fst + dcum07_lst	0.99	68	0.7	3	3
	Hmean_sup2m + d3_fst + dHmean_fst.sup1m	0.99	62	0.6	2	3
G (m ² /ha)	BVv1_H.S.sum	0.82	211	6.9	24	25
	BVv1_H.S.sum + BVv1_H.Q5 + TxPen1m_lst	0.92	189	4.6	16	17
	Nrel.12m24m + Nrel.sup24m + TxPen0.8H	0.78	219	7.6	26	28
Vtot (m ³ /ha)	BVv1_H.S.sum	0.89	380	104	24	25
	BVv1_H.S.sum + BVv1_S.Q10 + Volume_12m_22m	0.95	359	72	17	18
	Nrel.2m12m + Nrel.sup24m + dcum07	0.89	381	103	24	25
V7 (m ³ /ha)	BVv1_H.S.sum	0.86	382	107	28	30
	P_H.Q5 + BVv1_S.Q99 + Nrel.sup24m	0.95	353	65	17	18
	Hmean_lst.sup1m + Nrel.2m24m + dcum08	0.89	375	94	25	26

Tableau 3 : Modèles de prédiction de la hauteur dominante, (Ho) de la surface terrière (G), du volume total (Vtot) et du volume à découpe 7 cm (V7). Les modèles à une et trois variables sont présentés avant ceux n'utilisant que les métriques standards uniquement (en grisé). (voir l'annexe pour la définition des variables)

Pour G, le meilleur modèle obtenu intègre le volume des segments (BVv1_H.S.sum), le 5ième percentile de hauteur des « apex » (BVv1_H.Q5) et le taux de pénétration des derniers retours ayant atteint le seuil de 1m au-dessus du sol (TxPen1m_lst). Ce modèle présente une erreur quadratique moyenne de validation croisée de 4.9 m²/ha (17%).

Pour Vtot, le meilleur modèle intègre également BVv1_H.Asum, ainsi que le 10ème percentile des surfaces de segments (BVv1_S.Q10) et le volume de canopée situé dans la tranche de 12 à 22 m de hauteur (Volume_12m_22m). Cette dernière métrique est issue d'Hollaus *et al.* (2009). Ce modèle présente une erreur quadratique moyenne de validation croisée de 78.6 m³/ha (18%), ce qui est mieux que pour un modèle équivalent ne tenant pas compte de métriques de segmentation (25%).

Pour V7, le meilleur modèle à 3 variables utilise les métriques de segmentation et présente une erreur quadratique moyenne inférieure à 70 m³/ha (18%). Contrairement aux autres, ce modèle utilise des métriques issues de 2 méthodes de segmentation (P et BVv1). Le 5ième percentile des hauteurs d'apex de la méthode P (P_H.Q5) est associé au 99ième percentile de la surface des segments (BVv1_S.Q99), ainsi que la densité relative des retours au-dessus d'un seuil de 24 m (Nrel.sup24m). Encore une fois, ce modèle est supérieur à un équivalent qui n'utiliserait pas de métriques de segmentation (erreur quadratique de validation croisée de 26%).

Pour l'ensemble des résultats présentés au Tableau 3, aucun effet de « site » n'a été observé en ajoutant cette variable aux modèles ($p > 0.05$).

Les métriques de segmentation semblent de bons prédicteurs de la surface terrière et des volumes. Effectivement, en modèle univarié, la somme par placette des produits des surfaces de segment par leur

hauteur (H.S) explique entre 70 et 75% de la variance associée à G et aux volumes. Comme en foresterie, on associe facilement la surface des houppiers au diamètre des arbres, on aurait pu s'attendre à ce que la surface des segments soit intuitivement considérée comme la métrique la plus liée à G. Cependant, il semble que dans nos modèles, ce soit plutôt la combinaison de la hauteur et de la surface des segments qui soit le plus liée à G. Allouis *et al.* (2013) ont également observé ce résultat dans des forêts ouvertes du sud de la France et Salas *et al.* (2010) ont montré que la hauteur était également nécessaire pour bien prédire les diamètres des arbres à partir de diamètres de couronnes. Pour leur part, Hollaus *et al.* (2009) ont souligné l'importance des métriques associées aux volumes de canopée, pour prédire les volumes, dans une étude portant sur des forêts de conifères des Alpes Autrichiennes.

Les résultats que nous avons obtenus sont donc comparables à ceux d'autres études effectuées dans des forêts de conifères dont les structures sont plus simples (Magnussen *et al.*, 2012 ; Vastaranta *et al.*, 2012 ; Järnstedt *et al.*, 2012 ; Vastaranta *et al.*, 2013). Ils mettent cependant l'accent sur l'importance des métriques de segmentation, utilisées non pas comme substituts à l'identification des arbres mesurés sur le terrain, mais bien comme porteuses d'une information jusque-là négligée par les métriques LiDAR couramment utilisées dans la littérature et qui résument le nuage LiDAR en termes de percentiles de hauteurs et de densités (e.g. Naesset *et al.* 2004) .

5. Conclusion

Notre hypothèse concernant l'amélioration des modèles de prévision d'attributs forestiers par l'utilisation de métriques de segmentation est vérifiée. A première vue, il existe un lien étroit entre les surfaces des

« houppiers » issus de la segmentation et la surface terrière ou le volume des arbres. Ce lien n'est cependant pas simple, car il intègre la hauteur des « apex » à ces surfaces de « houppiers ». Leur produit semble une variable clé dans la prévision de ces attributs forestiers. La précision des modèles de prévision de la surface terrière et du volume avec trois variables se voit nettement améliorée par l'ajout de métriques de segmentation et les AIC diminuent fortement (de plus de 10 points).

Pour ce qui est de la méthode de segmentation et du filtrage des MNH, il semble que la méthode qui aurait pu paraître la moins bonne pour obtenir un nombre réaliste d'arbres, car elle tend à beaucoup sur-segmenter, soit celle qui produise les métriques les plus intéressantes pour prédire la surface terrière et le volume. Ce résultat laisse penser que c'est plutôt le volume situé sous la canopée qui est une métrique intéressante en non pas la stricte segmentation des houppiers. Sur cette base, il sera intéressant de tester si l'utilisation directe des volumes sous le MNH permet d'améliorer les modèles de prévision des variables dendrométriques.

Enfin, même si les résultats obtenus sont intéressants, il est essentiel de réaliser une étape de validation sur un plus grand nombre de placettes, car il est fréquent d'obtenir des résultats trop optimistes avec un nombre restreint de placettes d'étude (32 placettes dans notre cas). La robustesse de nos résultats devrait donc être évaluée sur un plus grand nombre de sites et de placettes, avec des structures forestières plus diversifiées.

La précision des modèles obtenus est de l'ordre de 20% (erreur de validation croisée). Une piste d'amélioration pourrait venir d'un raffinement des métriques de segmentation utilisées. Le volume d'un polygone (surface de segment x hauteur) ne représente que grossièrement la surface assimilatrice des arbres. Un raffinement par la prise en compte du volume ou de la surface « réelle » des houppiers pourrait être envisagé. De même, si l'on considère que le volume de polygones que nous avons obtenus est un paramètre clé pour la prévision d'attributs forestiers, alors on peut penser qu'en utilisant des MNH, le produit de la hauteur des pixels par leur surface pourrait raffiner les volumes des polygones de segments que nous avons utilisé ici. Il serait intéressant d'examiner l'importance de la somme des volumes des voxels (pixels en 3 dimensions) issus des MNH de différentes résolutions. Ceci permettrait d'éviter une étape de segmentation, augmenterait la rapidité des traitements et permettrait d'aborder la question de la résolution (« optimale ») des MNH.

Enfin, une partie de l'imprécision des modèles obtenus pour les volumes est possiblement liée au fait que nous utilisons des volumes calculés et non pas mesurés, pour calibrer nos modèles. Malgré le fait que nous utilisons les tarifs de volumes les plus récents, ces données présentent une imprécision. L'utilisation du LiDAR terrestre pourrait être considérée dans des travaux futurs, s'il s'avère que cet instrument puisse donner des « mesures » de volumes de qualité supérieure aux tarifs de cubage.

Remerciements

Ce travail a bénéficié d'un soutien financier de l'Agence Nationale de la Recherche (ANR) dans le cadre du projet ANR " FORESEE " (ANR-10-BIOE-08-07), « Caractérisation de la ressource forestière pour les bioénergies », ainsi que d'une aide de l'État gérée par l'Agence Nationale de la recherche au titre du programme Investissements d'avenir portant la référence n° ANR-11-LABX-002-01 (Laboratoire d'Excellence ARBRE). Nous remercions également le Conseil Général de Savoie (CG73) pour la mise à disposition des données LiDAR de la Combe d'Aillon, ainsi que les reviewers pour les améliorations qu'ils ont apportées à ce manuscrit.

Références

- Akaike, H., 1974. *A new look at the statistical model identification*. IEEE Trans. Autom. 19:716–723.
- Allouis, T., Durrieu, S., Vega, C., Couteron, P., 2013. *Stem Volume and Above-Ground Biomass Estimation of Individual Pine Trees from LiDAR Data: Contribution of Full-Waveform Signals*. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6:924–934.
- Bock, J., Dambrine, E., Dez, G., Dupouey, J.L., Georges-Leroy, M., Jolly, A., Martins, F.S.R.V., Renaud, J.P., 2009. *Towards site index mapping in deciduous stands using multi-echo LiDAR data*. Proceeding Extending For. Inventory Monit. IUFRO Div. 4 Quebec City, Canada.9p.
- Breidenbach, J., et R. Astrup. 2014. *The Semi-Individual Tree Crown Approach*. In *Forestry Applications of Airborne Laser Scanning*, édité par M. Maltamo, E. Næsset, et J. Vauhkonen, 113–133. *Managing Forest Ecosystems 27*. Springer Netherlands.
- Järnstedt, J., Pekkarinen, A., Tuominen, S., Ginzler, C., Holopainen, M., Viitala, R., 2012. *Forest variable estimation using a high-resolution digital surface model*. ISPRS J. Photogramm. Remote Sens. 74:78–84.
- Jaskierniak, D., Lane, P.N.J., Robinson, A., Lucieer, A., 2011. *Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions*. Remote Sens. Environ. 115:573–585.
- Kartinen, H., Hyypä, J., Yu, X., Vastaranta, M., Hyypä, H., Kukko, A., Holopainen, M. et al. 2012. *An International Comparison of Individual Tree Detection and Extraction Using Airborne Laser Scanning*. Remote Sensing 4 (4): 950–974.
- Khosravipour, A., Skidmore, A.K., Isenburg, M., Wang, T., Hussin, Y.A., 2013. *Development of an algorithm to generate a LiDAR pit-free canopy height model*. SilviLaser. 30:125 – 128.
- Hollaus, M., W. Wagner, K. Schadauer, B. Maier, et K. Gabler. 2009. *Growing stock estimation for alpine forests in Austria: a robust LiDAR-based approach*. Can. J. For. Res. 39 (7): 1387–1400.
- Lindberg, E., J. Holmgren, K. Olofsson, J. Wallerman, et H. Olsson. 2010. *Estimation of Tree Lists from Airborne Laser Scanning by Combining Single-Tree and Area-Based Methods*. International Journal of Remote Sensing 31 (5): 1175–1192.

- Magnussen, S., Boudewyn, P., 1998. *Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators*. Can. J. For. Res. 28:1016–1031.
- Magnussen, S., Næsset, E., Gobakken, T., Frazer, G., 2012. *A fine-scale model for area-based predictions of tree-size-related attributes derived from LiDAR canopy heights*. Scand. J. For. Res. 27:312–322.
- Maltamo, M., Malinen, J., Packalén, P., Suvanto, A., Kangas, J., 2006. *Nonparametric estimation of stem volume using airborne laser scanning, aerial photography, and stand-register data*. Can. J. For. Res. 36:426–436.
- 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*. Photogramm. Eng. Remote Sensin 66:1367–1371.
- Næsset, E., 1997. *Determination of mean tree height of forest stands using airborne laser scanner data*. ISPRS J. Photogramm. Remote Sens. 52:49–56.
- Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, A., Söderman, U., 2004. *Laser scanning of forest resources: the nordic experience*. Scandinavian Journal of Forest Research 19: 482–499.
- Næsset, E., Gobakken, T., Bollandsås, O.M., Gregoire, T.G., Nelson, R., Ståhl, G., 2013. *Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway*. Remote Sens. Environ. 130:108–120.
- Picard, N., Saint-André, L., Henry, M., 2012. *Manual for building tree volume and biomass allometric equations. From field measurement to prediction*. Food and Agricultural Organization of the United Nations, Rome, and Centre de Coopération Internationale en Recherche Agronomique pour le Développement. Montpellier.
- Salas, C., Ene, L., Gregoire, T.G., Næsset, E., Gobakken, T., 2010. *Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models*. Remote Sens. Environ. 114:1277–1285.
- St-André, L., Munoz, A., Bock, J., Jolly, A., Renaud, J.P., Deleuze, C., Monnet, J.M., Mermin, E., Bouvier, M., Durrieu, S., 2013. *Modèles allométriques utilisés pour estimer les volumes des arbres*. Projet ANR-10-BIOE-008, ANR Foresee. 9p.
- Valbuena, R., Packalen, P., Mehtätalo, L., García-Abril, A., Maltamo, M., 2013. *Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning*. Can. J. For. Res. 43:1063–1074.
- Vastaranta, M., Kankare, V., Holopainen, M., Yu, X., Hyyppä, J., Hyyppä, H., 2012. *Combination of individual tree detection and area-based approach in imputation of forest variables using airborne laser data*. ISPRS J. Photogramm. Remote Sens. 67:73–79.
- Vastaranta, M., Wulder, M.A., White, J.C., Pekkarinen, A., Tuominen, S., Ginzler, C., Kankare, V., Holopainen, M., Hyyppä, J., Hyyppä, H., 2013. *Airborne laser scanning and digital stereo imagery measures of forest structure: comparative results and implications to forest mapping and inventory update*. Can. J. Remote Sens. 39: 382–395.
- Vauhkonen, J., Ene, L., Gupta, S., Heinzl, J., Holmgren, J., Pitkänen, J., Solberg, S., Wang, Y., Weinacker, H., Hauglin, K.M., Lien, V., Packalén, P., Gobakken, T., Koch, B., Næsset, E., Tokola, T., Maltamo, M., 2011. *Comparative testing of single-tree detection algorithms under different types of forest*. Forestry 85:27–40.
- Vega, C., A. Hamrouni, S. El Mokhtari, J. Morel, J. Bock, J. -P. Renaud, M. Bouvier, et S. Durrieu. 2014. *PTrees: A point-based approach to forest tree extraction from LiDAR data*. International Journal of Applied Earth Observation and Geoinformation 33 : 98–108.
- White, J.C., Wulder, M.A., Varhola, A., Vastaranta, M., Coops, N.C., Cook, B.D., Pitt, D., Woods, M., 2013. *A best practices guide for generating forest inventory attributes from airborne laser scanning data using an area-based approach*. Can. For. Serv. Inf. Rep. FI-X, 50.
- Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C., Gerard, P.D., 2003. *Characterizing vertical forest structure using small-footprint airborne LiDAR*. Remote Sens. Environ. 87:171–182.

Annexe : Définition des variables présentées au Tableau 3

Métrique	Définition
BVv1_H.gini	Indice de Gini basé sur les maxima locaux de la méthode des "bassins versants" non filtrés
BVv1_H.Q5	5ième quantile de hauteur des maxima locaux de la méthode des "bassins versants" non filtrés
BVv1_H.S.sum	Somme des produits Hauteur*Surface des "bassins versants" non filtrés
BVv1_S.Q10	10ième quantile des surfaces de la méthode des "bassins versants" non filtrés
BVv1_S.Q99	99ième quantile des surfaces de la méthode des "bassins versants" non filtrés
d3_fst	Densité relative de tous les premiers situés dans le 4ième tranche de hauteur (l'épaisseur des tranches est de $((H95-2)/10)$ m)
dcum07	Densité relative de tous les retours situés entre les tranches de hauteur 1 et 8 (l'épaisseur des tranches est de $((H95-2)/10)$ m)
dcum07_1st	Densité cumulée de tous les derniers retours entrés dans le 7ième tranche de hauteur (l'épaisseur des tranches est de $((H95-2)/10)$ m)
dcum08	Densité cumulée de tous les retours entrés dans le 8ième tranche de hauteur (l'épaisseur des tranches est de $((H95-2)/10)$ m)
dHmean_fst.sup1m	Moyenne des écarts à la hauteur moyenne, des hauteurs de tous les premiers retours supérieurs à 1 m
H90_1st.sup2m	Hauteur du 90ième quantile de Hauteur des derniers retours supérieurs à 2 m
H95_fst	Hauteur du 95ième quantile de Hauteur des premiers retours
Hmean_1st.sup1m	Hauteur moyenne des derniers retours supérieurs à 1 m
Hmean_sup2m	Hauteur moyenne de tous les retours supérieurs à 2 m
Nrel.12m24m	Proportion de retours dans la couche située entre 12 et 24 m
Nrel.2m12m	Proportion de retours dans la couche située entre 2 et 12 m
Nrel.2m24m	Proportion de retours dans la couche située entre 2 et 24 m
Nrel.sup24m	Proportion de retours supérieurs à 24 m
P_H.Q5	5ième quantile de hauteur des maxima locaux de la méthode Ptree
TxPen0.8H	taux de pénétration de tous les retours au seuil du 80ième percentile de hauteur
TxPen1m_1st	Taux de pénétration des derniers retours au seuil de 1 m
Volume_12m_22m	Volume de canopée située entre 12 et 22 m