Estimativa da altura dominante em povoamentos decíduos através de dados LIDAR com múltiplos retornos

Estimating dominant height in deciduous stands using multi-echo LIDAR data

Flora S.R.V. Martins[1]
Jérome Bock[2]
Ethienne Dambrine[3]
Grégory Dez[4]
Jean-Luc Dupouey[5]
Murielle Georges-Leroy[6]
Anne Jolly[7]
Jean-Pierre Renaud[8]

Resumo

A altura dominante (Ho) de 120 parcelas georreferenciadas ($600m^2$ cada) foi estimada através de diversas medidas de distribuição, máximas locais e densidade derivadas da varredura a laser com pequena impressão obtidas em um povoamento florestal decíduo irregular. As parcelas foram distribuídas de forma a contemplar toda a variabilidade de alturas dominantes e o tipo de regeneração existente nos $112km^2$ da área estudada, localizada na floresta de Haye, França. O modelo construído com variáveis de máxima local (Hmv5) e densidade de retornos (d9) foi capaz de estimar a Ho com alta acurácia, além de ser independente do tipo de povoamento, o que possibilitou sua aplicação em toda a floresta. A validação cruzada do modelo final mostrou que este explicou 98% da variabilidade observada nas parcelas em campo, com um RMSE de 0.77m (3.31%). Nenhum efeito drástico da escolha do MNT, densidade de retornos ou do posicionamento das parcelas foi detectado no modelo, sugerindo grande estabilidade.

Palavras-chave: sensoriamento remoto; LIDAR; índice de sítio; altura dominante; povoamentos decíduos.

1 Office National des Forêts, R&D Department Velaine-en-Haye, France and Forest Ecology & Ecophysiology Unit, INRA, Champenoux, France; E-mail: florasrvm@gmail.com
2 Office National des Forêts, R&D Department Velaine-en-Haye, France; E-mail: jerome.bock@onf.fr
3 Forest Ecosystems Biogeochemistry Unit, INRA, Champenoux, France; E-mail: dambrine@nancy.inra.fr
4 Office National des Forêts, R&D Department Velaine-en-Haye, France and Forest Ecology & Ecophysiology Unit, INRA, Champenoux, France.
5 Forest Ecology & Ecophysiology Unit, INRA, Champenoux, France; E-mail: dupouey@nancy.inra.fr
6 Service régional de l'Archéologie de Lorraine, Metz, France; E-mail: murielle.leroy@culture.gouv.fr
7 Office National des Forêts, R&D Department Velaine-en-Haye, France.
8 Office National des Forêts, R&D Department Velaine-en-Haye, France.

Recebido para publicação em 09/07/2010 e aceito em 31/07/2010

Abstract

The mean dominant heights of 120 georeferenced field sample plots (600m^2 each) were estimated from a range of canopy densities, distributions and local maxima metrics derived by a small-footprint laser scanner over various deciduous forest stands using regression analysis. The sample plots were distributed in order to better represent the variability in stand dominant height and regeneration practices throughout a 112km^2 study area in Haye forest, France. The model constructed with a local maxima (*Hmv5*) and a LIDAR density metric (*d9*) was able to estimate Ho with a very high accuracy and was not sensitive to stand types. Cross-validation showed that the final model explained 98% of the variability in ground-truth dominant height, with a RMSE of 0.77m (3.31%). No drastic effects of DTM, echo densities, or positioning errors were found in the models.

Key words: remote sensing; LIDAR; site index; dominant height; deciduous stands.

Introduction

In order to reduce the expenses of forest inventory and supply forest managers with accurate data, the use of Remote Sensing techniques have been widely studied during the last decades. The concept of using laser system to assess forest parameters is recent and attracts much attention as a rapid and efficient tool for forest inventories (LEEUWEN; NIEUWENHUIS, 2010). Airborne LIDAR (*Light Detection And Raging)* systems are now used operationally for forest inventory at various geographical scales (NÆSSET, 2002). However, to estimate parameters of interest, no generic equations are available owing to variability in data acquisition systems, stand composition and structures. Therefore, the use of LIDAR metrics requires to perform empirical equations from ground-truth plots. Nelson et al. (1988) recommended the use of laser-derived stand profiles for the retrieval of stand characteristics.

Several sources of variability may affect accuracy of estimates deducted from LIDAR metrics. A major one is the pulse density that affects directly the resolution of the returned signal, as well as the cost of the whole operation (LIM et al., 2008, NÆSSET, 2009). Field position error could also affect estimates accuracy (GOBAKKEN and NÆSSET, 2009). As LIDAR metrics are calibrated using field positions generally measured with GPS, a positioning variability could affects results, especially in heterogeneous stands, where large variations may occurs over short distances. Another source of variability may be the generating algorithm of digital terrain models (DTM) and it's faithful to the relief.

During the last 20-25 years, several experiments have been carried out in order to determine various forest stand parameters, such as tree height, density and timber volume by different airborne laser systems (NÆSSET, 1997; 2002; VEGA; ST-

ONGE, 2008). However, airborne LIDAR (ALS) has been the most widely used, partly because of terrestrial laser scanning (TLS) restriction by its short and limited working range and occlusion, phenomenon that is especially encountered in the upper canopy (LEEUWEN; NIEUWENHUIS, 2010). Consequently, researchers recommend ALS to estimate tree height rather than TLS.

Site fertility index represents crucial information for forest management and its calculation requires the knowledge of stand age and dominant height (Ho) which are both difficult to acquire. Whereas age could be deduced from plantation dates found in archives, measuring Ho is often tedious. Therefore, site index is rarely mapped over large areas with a fine scale resolution. In forestry, the dominant height (Ho) is defined as the average height of the 100 biggest trees per hectare. As this estimate is a rank metric, its estimation is generally biased for small sampling area (PIERRAT et al., 1995; GARCIA, 1998). In France, in order to alleviate this problem, it is a common practice to estimate Ho from the *n*-1 biggest trees given a plots area of n*100m^2 (DUPLAT; PERROTTE, 1981). With LIDAR data, one could imagine to find out Ho from the measurements of the 100 highest trees per hectare, based on the highest returned hits. However, for a given field plot, without a tree segmentation treatment of the LIDAR cloud, these returns could be produced from only a few trees resulting in an upward bias in height estimates (LOVELL et al., 2003). Coops et al. (2007), in a *Pseudotsuga menziesii* and *Tsuga sp.* mixed stand, found a strong relationship between Ho, observed from ground-truth plots and the average of the four highest LIDAR returns taken from 4 plot quadrants (R^2=0.82). In Norway, Næsset (2004) also revealed a very good prediction of Ho using LIDAR data, with only a small error associated to this estimate 3.0 - 7.6% (0.70 - 1.55m). Similar results were recently obtained on broadleaves by Heurich and Thoma (2008) (R^2 of 0.94 and an error of less then 6% for Ho).

So far, estimation of forest stand characteristics from airborne laser scanner data has focused mostly on coniferous forests. In this study we explored the use of multi-echo LIDAR metrics in estimating Ho for French deciduous stands of different structures. The impact of pulse densities was evaluated by thinning the original LIDAR dataset and the effect of field position error was examined both indirectly, by varying the LIDAR extraction area, and directly, on a subset of plots more precisely localised by triangulation from ground features easily observable on the DTM. The effect of the DTM was also examined by using different data adjustments. Finally, Ho was estimated using different LIDAR metrics including two algorithms calculating local maxima.

Material and Methods

Study area and stand delineation

This study was conducted in the Haye forest (48°41'39"N, 6°04'20"E), North-Eastern France. The total area was 112km^2. The altitudes range from 220 to 410m along a predominant regular topography (80%) broken by small valleys (20%). The massif covers a limestone plain with beech (*Fagus sylvatica*) and oak (*Quercus petraea* and *Quercus robur* L.) as the main tree species. The structure of the forest is characterized by three main stand typologies: high forest, coppice with standards, or intermediate between these two types. Coppice with standards is a two-story management system where among regularly cut hornbeam trees ("coppice"), some oak and beech trees called "standards" are left to grow as larger size timber.

Ground-truth plots

As ground-truth, 120 field plots ($600m^2$ each) were installed and classified according to stand types. For each plot, Ho was calculated by averaging top height measurements of the 5 biggest trees per plot. Two opposed height measurements per tree were performed using a Vertex. For bent trees, corrections were performed to take into account the horizontal distance between the observer and the ground projection of tree apex. Plot position was recorded using a Geoexplorer XT GPS.

Laser scanner data

Multi-echo LIDAR data were acquired at leaf-off state in March 2007 over the $112km^2$ of forest. First, last and a maximum of two intermediates echoes were recorded by pulse. LIDAR return densities ranged from 10 to 64 points/m^2 which were classified as ground or vegetation by the vendor using a proprietary algorithm. No filtering was applied and so, all the registered echoes were considered in computations. Flight characteristics are shown in table 1.

Table 1. LIDAR data and flight characteristics.

Parameters	Values
Equipment	Leica ALS50-II
Flight altitude	700m
Speed	$100m.s^{-1}$
Band width	428m
Lateral overlapping	30%
Pulse frequency	115.8kHz
Mean ground pulse density	5.26 points.m^{-2}
Scanning angle	34° (+/- 17°)
Measured echoes	4 maximum (first, last, intermediates)
Altitude precision	8 cm
X Y precision	25 - 30 cm

LIDAR metrics

LIDAR-based predictors were generated by subtracting vegetation points by linear interpolation from DTM. At plot level, height percentiles (90, 95, 99, etc.), return densities (absolute and relative per layer) and distribution parameters were calculated. Local maxima metrics were also derived using two methods (Figure 1). The first one subdivided the plot area into "*n*" concentric rings of equal areas, and the final metric was an average of the maximum heights obtained for each ring. This LIDAR metric was further identified with the prefix Hdoman-*n* (*n* being the number of rings, from 2 to 5). The other method started by finding a maximal point for a given plot and used an exclusion area of a given radius away from this point prior to restart the operation. This looped until the desired number of local maxima was obtained (from 2 to 5) and these heights were then averaged. This metric

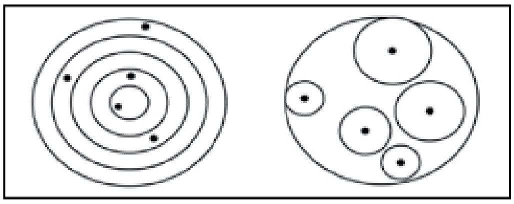

Figure 1. Derived metrics of local maxima: 2 to 5 highest points in a $600m^2$ plot were selected and then averaged. In (a), approach of rings and in (b), maxima using "crown-like" exclusion zones

partly avoid the possible bias of measuring several times the same tree. It mimicked tree crowns, and the exclusion radius varied as a function of pulse height. The regression equation has been obtained from crown radius measurements of 800 beech trees (R^2=0.54). This variable was further identified with the prefix Hmv*n* (where *n* is a function of the number of selected points in one plot, which is 2, 3, 4 or 5).

Assessing variability sources

DTM

DTM is mandatory to calculate vegetation height. As different adjustment methods to the micro relief can produce different DTM and so, affecting canopy heights estimations, the effect of three DTM was assessed. The first one was provided by the vendor (resolution of 0.5 cm), and two others were modelled from a linear and a quadratic regression using ground points. We have also tested the use of heights in the DSM (subtraction of terrain model from elevation model) that have one soil point and one vegetation point per 0,25 cm^2.

Positioning error

The impact of the LIDAR extraction area used to compare to the ground-trough and GPS uncertainties were examined. LIDAR data were extracted from different radii: 5, 10, 13.82 (the same of field plots), 15 and 20m, centered on the ground plot coordinates. The second approach for assessing the impact of positioning error was to relocalise plots (*n=35*) based on characteristic features from the DTM. Thirty plots were precisely relocalised in such a way using laser telemeters and triangulation.

Pulse density

In order to appreciate the impact of pulse density, ground and vegetation returns were thinned up to 5% per plot. LIDAR metrics were recalculated and the thinning effect on the residual error was examined (30 repetitions).

Stand structure and robustness

The behaviour of the LIDAR metrics on Ho estimations was observed using regression analyses. The effect of stand typology was also examined during the model construction. The chosen model robustness was checked by cross validation. One of the 120 experimental plots was removed (leave-one-out) from the dataset at a time and the selected model was fitted to the data from the 119 remaining plots. The mean height of the dominant trees was then predicted for the removed observation. This procedure was repeated until predicted values were obtained for all plots.

Results and Discussion

The results obtained using univariate models using different LIDAR metrics to estimate Ho are presented in table 2. Nine of these models had a good accuracy, with a residual error smaller than 1m, which is probably close to the measurement error from ground plots. Most of these models used local maxima variables, which suggested that spatial information from LIDAR returns improved the accuracy of Ho estimates. Among the LIDAR metrics associated to the return distribution, the heights of the 95th percentile (hlid 95), followed by the 99th (hlid 99) percentile have also showed low residual errors (Table 2). The standard deviation of the height distribution of LIDAR returns

Table 2. RMSE and R^2 from univariated models where independent variables are LIDAR metrics

LIDAR metric	RMSE	R^2
hlid 95	0.83	0.99
hdoman5	0.86	0.99
hmv5	0.89	0.99
hdoman4	0.89	0.99
hdoman3	0.92	0.99
hmv3	0.94	0.98
hlid 99	0.95	0.98
hdoman2	0.96	0.98
hmv2	0.99	0.98
hlid 99.5	1.01	0.98
hlid max	1.10	0.98
std	1.11	0.98

(std) yielded a poor residual error (>1m). The three best models were those retained in further analyses.

Comparing plot localisations using GPS measurements and field localization with LIDAR DTM and triangulation, we found an averaged distance of 2.46m (0.6 to 6.62m) between these plot centres. Thus, it seems that the impact of positioning error on Ho accuracy was only of minor importance given the precision of the GPS used in this study. Corrections for these positioning errors only slightly improved the model accuracy based on a 13.8m extraction radius, which corresponds to the ground plot area (Figure 2). Varying the extraction radius to 15m had also a minor impact. However, a clear degradation in precision was observed for extraction radii much larger (*e.g.* 20m) or much smaller (*e.g.* 10m) than the one used in ground sampling.

The use of different DTM showed no drastic changes in the precision obtained for Ho estimation, except when using only DTM and DSM delivered with one vegetation point per 0.25 cm^2. This result suggests

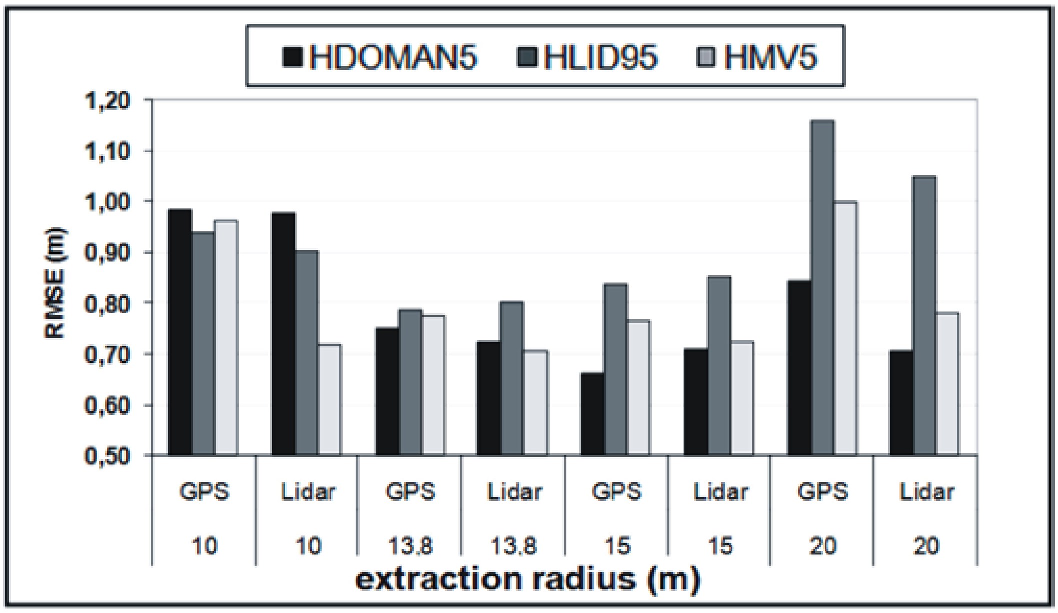

Figure 2. Impact of positioning uncertainties on the precision of Ho estimations (based on RMSE values). Four extraction radii were used, ranging from 10 to 20m. The impact of uncertainties associated to GPS and DTM positioning using triangulation (Lidar) is also compared for each LIDAR extraction radii. Ho was estimated using 3 models, 2 based on local maxima metrics (*i.e.* Hmv5 and Hdoman5) and a third based on the 95th height percentile of the returns distribution

that in our study, which was predominantly under flat terrain, the use of either the DTM supplied by the provider, or the ones built from linear or quadratic regressions was of minor importance in terms of Ho accuracy. So, as a simple method, DTM performed by a plan regression in SAS® was used in the rest of the experiment, as it allowed us to thin the LIDAR cloud together with the DTM. Heights were then obtained by subtracting the altitudes of the LIDAR cloud from this DTM.

In way of eliminating the relationship between model residues and the different physiognomies - once this characteristic is not known among all forest – multivariate models were constructed. The final composition considered not just the significance level of metrics, but also their pertinence.

than the ones tested here (Gobakken and Naesset 2008, Lim et al. 2008).

Finally, the model with the lowest RMSE (0.75m) that had a R^2 of 0.99 was retained (Equation 1):

$$Hdom = 9{,}09 + 0{,}99Hmv5 - 0{,}09d9 \quad (1)$$

Where: *Hdom* is the estimated Ho (in meters); *Hmv5* is the local maxima metric that averaged the highest five points (in meters) and *d9* is the percentage of vegetation returns found up to the 90^{th} percentile based on the returns height distribution.

This model met our expectations. It had not only a good accuracy, was independent of the forest types, but the Hmv5 metric allows a certain robustness of the results, precluding

Table 3. RMSE and R^2 from multivariate models. d9 = percentage of points until 90^{th} percentile, n = number of vegetation points registered in the plot, dif99_95 = vegetation points registered between the 99^{th} and 95^{th} percentile, std = standard deviation and d6 = percentage of points registered until de 60^{th} percentile. NS = non-significant relationship, * = significant relationship

Independent variables			R^2	RMSE (m)	Stand type effect in residues
Hmv5	d9		0.99	0.75	NS
Hmv5	d9	n	0.99	0.75	NS
Hdoman5	dif99_95	std	0.99	0.76	NS
Hlid95	std		0.99	0.81	*
Hlid95	d6	std	0.98	1.01	NS

The impact of thinning up to 5% the LIDAR signal did not deteriorate the precision of Ho estimates and no metrics related to density was significant in the models (Figure 3). These results may be caused by the very high pulse density obtained in this experiment. Others have shown degradation in accuracy occurring only at very low signal density, much lower

aberrant points for example. Cross-validation (Figure 4) of selected model

revealed that the mean difference between predicted and observed mean dominant heights was 0.77m (3.31 %), with a minimum of 0.01m (0.04 %) and a maximum of 2.79m (12 %), and a R^2 of 0.98. This model was then applied for all forest and the result is presented in figure 5.

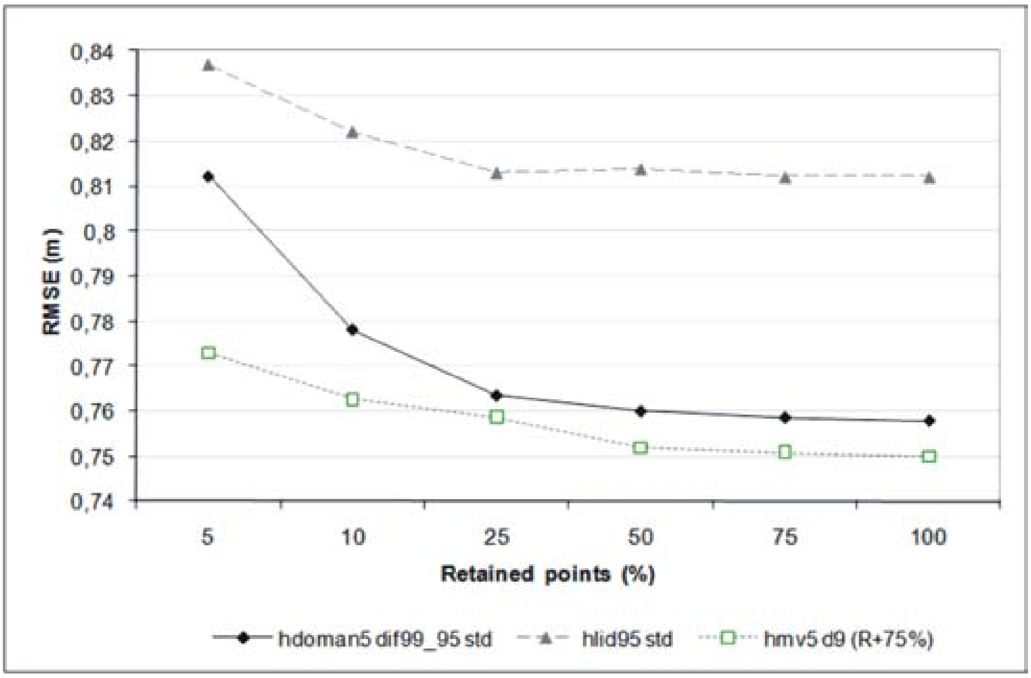

Figure 3. Density effect in RMSE of multivariated models

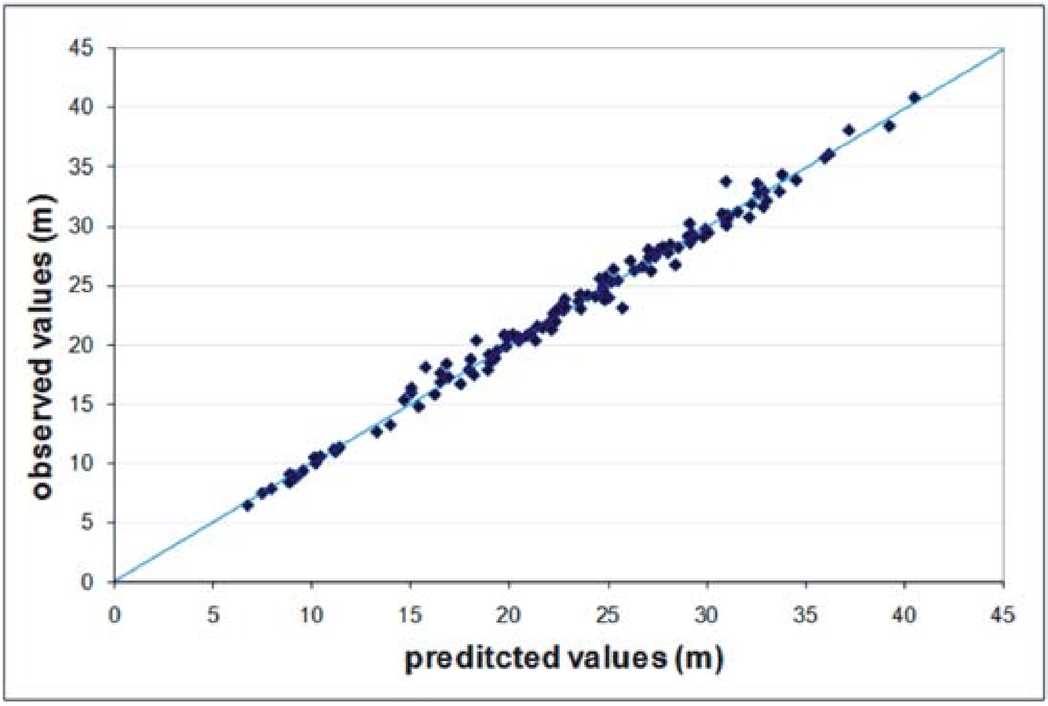

Figure 4. Cross-validation of predicting Ho model: predicted x observed values in meters

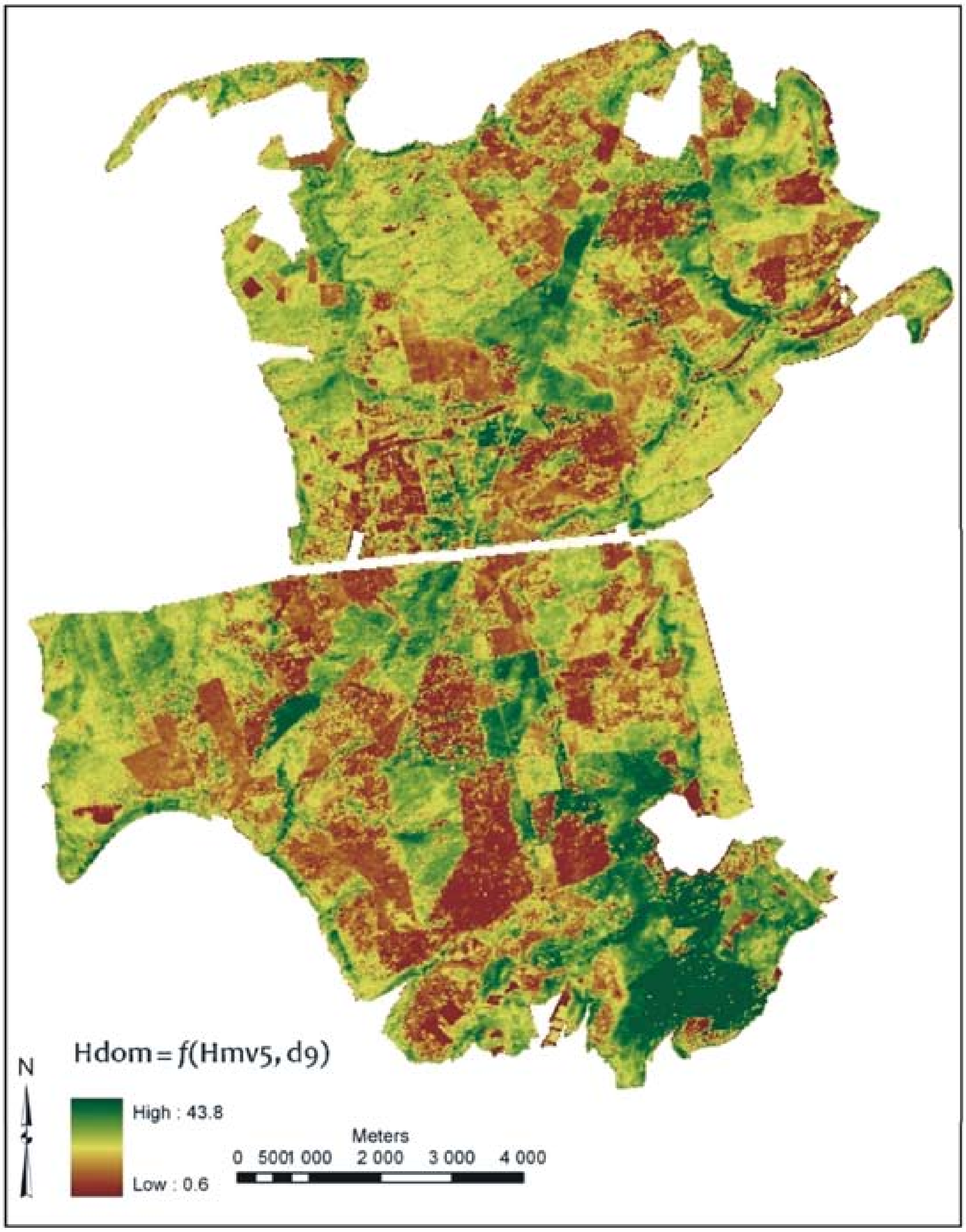

Figure 5. Distribution map of dominant height trought Haye's Forest

Conclusion

In this experiment, Ho was estimated with a residual error smaller than 77 cm. This accuracy probably reflected ground measurement error. Moreover, the use of local maxima and density metrics in the models allowed to remove stand typology effects in the residues, which provided more reliability for the mapping stage. Even though maximal height returns have been shown to be a highly variable LIDAR metric, affected for example by pulse density (LIM et al., 2008, NÆSSET, 2009), we observed that local maxima metrics were particularly interesting to estimate Ho, giving more robustness to the model toward the different stand type studied. Our results also suggest that the uncertainties associated to GPS positioning were only of minor importance in Ho estimations. This could probably be due to the relatively precise instrument used. Finally, a finding of particular importance is also that the LIDAR density could be reduced without any apparent impact on Ho accuracy. This result could help reducing the costs of data acquisition in further studies.

Literature cited

COOPS, N. C.; HILKER, T.; WULDER, M. A.; ST-ONGE, B.; NEWNHAM, G.; SIGGINS, A.; TROFYMOW, J. A. Estimating canopy structure of Douglas-fir forest stands from discrete return LIDAR. **Trees**, v. 21, n. 3, p. 295-310, 2007.

DUPLAT, P.; PERROTTE, G. **Inventaire et estimation de l'accroissement des peuplements forestiers**. Paris : Office National des Forêts (ed.). Paris, 1981. 432 p.

GARCIA, O. Estimating top height with variable plot size. **Canadian Journal of Forest Research**, v. 28, p. 1509-1517, 1998.

GOBAKKEN, T.; E. NAESSET. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. **Canadian Journal of Forest Research,** v.38, p. 1095-1109, 2008.

______. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. **Canadian Journal of Forest Research**, v. 39, n.5, p. 1036-1052, 2009.

HEURICH, M.; THOMA, F. Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (*Fagus sylvatica*) and Norway spruce (*Picea abies*) forests. **Forestry**, v. 81, p. 645-661, 2008.

LEEUWEN, M. VAN; NIEUWENHUIS, M. Retrieval of forest structural parameters using LIDAR remote Sensing. **European Journal of Forest Research**, v.129, p. 749–770, 2010.

LIM, K.; HOPKINSON, C.; TREITZ, P.. Examining the effects of sampling point densities on laser canopy height and density metrics. **Forestry Chronicle**, v. 84, p. 876-884, 2008.

LOVELL, J. L.; JUPP, D. L. B.; CULVENOR, D. S.; COOPS, N. C. Using airborne and ground based ranging LIDAR to measure canopy structure in Australian forests. **Canadian Journal of Remote Sensing**, v. 29, n. 5, p. 607-622, 2003.

NÆSSET, E. Determination of mean tree height of forest stands using airborne laser scanner data. **Journal of Photogrammetry and Remote Sensing**, *v.*52, p. 49-56, 1997.

______. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. **Remote Sensing of Environment**, v. 80, p. 88-99, 2002.

______. Practical large-scale forest stand inventory using a small footprint airborne scanning laser. **Scandinavian Journal of Forest Research**, n. 19, p. 164-179, 2004.

______. 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,** v. 113, p. 148-159, 2009.

NELSON, R.; KRABILL, W.; TONELLI, J. Estimating forest biomass and volume using airborne laser data. **Remote Sensing of Environment**, v. 24, n. 2, p. 247-267, 1988.

PIERRAT, J. -C.; HOULLIER, F.; HERVÉ, J. –C.; SALAS GONZALES, R. Estimation de la moyenne des valeurs les plus élevées d'une population finie: application aux inventaires forestiers. **Biometrics**, v. 51, p. 679-686, 1995.

VEGA, C.; ST-ONGE, B. 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**, v. 112, p. 1784–1794, 2008.