Processing Full-Waveform Lidar Data to Extract Forest Parameters and Digital Terrain Model: Validation in an Alpine Coniferous Forest

Adrien Chauve, Sylvie Durrieu, Frédéric Bretar, Marc Pierrot-Deseilligny, William Puech

PROCESSING FULL-WAVEFORM LIDAR DATA TO EXTRACT FOREST PARAMETERS AND DIGITAL TERRAIN MODEL: VALIDATION IN AN ALPINE CONIFEROUS FOREST

Adrien CHAUVE[1,2,3], Sylvie DURRIEU[1], Frédéric BRETAR[2], Marc PIERROT-DESEILLIGNY[1], William PUECH[3]

[1]UMR TETIS Cemagref/Cirad/ENGREF-AgroParisTech, Maison de la Télédétection 500, rue J.F. Breton, 34093 Montpellier Cedex 5, France, 04 67 54 87 32, 04 67 54 87 00, adrien.chauve@teledetection.fr, sylvie.durrieu@teledetection.fr, pierrot@teledetection.fr

[2]Laboratoire MATIS - Institut Géographique National, 2-4 Avenue Pasteur, 94165 Saint Mandé cedex, France, 01 43 98 84 36, 01 43 98 85 81, frederic.bretar@ign.fr

[3]Laboratoire LIRMM, UMR CNRS 5506, Université Montpellier II, 161, rue Ada, 34392 Montpellier Cedex 05, France, 04 67 41 86 85, 04 67 41 85 00, william.puech@lirmm.fr

ABSTRACT

Small footprint discrete return lidar data have already proved useful for providing information on forest areas. During the last decade, a new generation of airborne laser scanners, called full-waveform (FW) lidar systems, has emerged. They digitize and record the entire backscattered signal of each emitted pulse. Full-waveform data hold large potentialities. In this study, we investigated the processing of raw full-waveform lidar data for deriving Digital Terrain Model (DTM) and Canopy Height Model (CHM). The main objective of this work was to compare geometric information derived from full-waveform and multi-echo data for various stands. An enhanced peak detection algorithm developed in a previous study was used to extract target positions from full-waveform data on plots under different stand characteristics. The resulting 3D point clouds were compared to the discrete return lidar observations provided by the lidar operator. Ground points were then identified using an original classification algorithm. They were used to derive DTMs which were compared to ground truth. Digital Surface Models were obtained from first echoes and canopy height models were then computed. Detecting weak echoes, when processing full-waveform data, enabled to better describe the canopy shape and to penetrate deeper into forest cover. However DTM was not significantly improved.

Keywords: waveform analysis, signal modelling, DTM, lidar, forest

1 INTRODUCTION

Airborne laser scanning is an active remote sensing technique providing range measurements between the laser scanner and the Earth topography. Based on direct georeferencing using both GPS and inertial measurements, such distance measurements are mapped into 3D point clouds with high accuracy and relevancy. Standard small footprint airborne multi-echo laser scanner systems can detect up to six echoes along the travel path of the laser pulse: the first echo is associated with the top of the canopy and the last pulse with the ground. Discrete return lidar observations have already proved useful for providing information on forest areas: individual tree extraction (Brandtberg *et al.*, 2003), height and crown diameter measurement (Persson *et al.*, 2002; Naesset and Bjerknes, 2001), large scale automatic extraction of tree tops (Nilsson *et al.*, 2003). Many methods and algorithms have been developed for forest measurements (Hyyppä *et al.*, 2004). During the last decade, a new generation of airborne laser scanners, called full-waveform (FW) lidar systems, has appeared. They digitize and record the entire backscattered signal of each emitted pulse (see figure 1). Full-waveform data hold large potentialities. In addition to an improvement of range measurements, physical properties of the targets included in the diffraction cone are likely to be derived from waveform analysis. Studies have been carried out on forestry applications to measure the canopy height (Lefsky et al., 1999), and vertical distribution of canopy material (Dubayah and Blair, 2000) using data acquired with large footprint experimental lidar systems. Modelling of raw lidar signal recorded by recent small footprint industrial systems has already proved efficient in increasing the number of detected targets in comparison with data provided by multi-echo lidar systems for which real-time point extraction method is unknown to the end user (Persson *et al.*, 2005; Chauve *et al.*, 2007).

Detecting weak echoes allows to better describe 3D vegetation structure and ground.. As a consequence, Digital Terrain Model (DTM), Digital Surface Model (DSM), and the derived Canopy Height Model (CHM) are expected to be significantly improved.

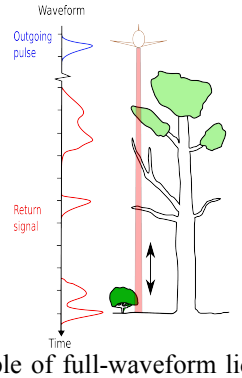

Figure 1. Principle of full-waveform lidar system. Laser emitted pulse (in blue) and backscattered signal (in red).

However real potentialities of small footprint full-waveform lidar systems for forest characterization has been little studied until now. In this study, we investigated the processing of raw full-waveform lidar data for extracting more points than classical multi-echo data, and studied the influence on resulting DTMs and CHMs.

2 AVAILABLE DATA

2.1 AREA OF INTEREST

In this study, the surveyed area was an alpine coniferous forest near Digne-les-Bains, France.

2.2 FULL-WAVEFORM LIDAR DATA

The data acquisition was performed in April 2007 using a RIEGL LMS-Q560 system. The main technical characteristics of this sensor are presented in (Wagner *et al.*, 2006). The lidar system operated at a pulse rate of 111 kHz. The flight height was around 500 m leading to a footprint size of about 0.25 m. The system temporal sampling is 1 ns (0.30 m). The point density was about 5 pts/m2. Each return waveform was made of one or two sequences of 80 samples corresponding to an altimetric profile of 24 or 48 m. For each emitted pulse, both emitted and return waveforms as well as the 3D point cloud computed by the lidar operator were provided.

2.3 FIELD DATA

In order to evaluate the potential of full-waveform lidar data in various stand conditions, 3 plots were selected with different characteristics:

- plot 1: (71 m x 47 m) low-density Black pine stand originating from a seed cutting, including a bare soil area; 9 m difference in height; around 66 stem/ha;
- plot 2: (32 m x 22 m) old dense old Black pine plantation on sloping terrain; 12 m difference in height; around 440 stem/ha;
- plot 3: (35 m x 21 m) very dense old Sylvester pine plantation; around 449 stem/ha.

Accurate positions, diameters at breast height (DBH), total heights and crown dimensions (heights and diameters) were measured for all the trees of the arboreal strata. The underlayer vegetation was also described. For each plot, ground coordinates, measured using tacheometers and DGPS, were available for a set of points. Unfortunately, because of georeferencing issues, we could only process the data of the first two plots for DTM quality assessment.

3 METHODOLOGY

3.1 WAVEFORM PROCESSING

Waveform processing consists in decomposing the waveform into a sum of components or echoes, where each component characterizes the contribution of a target to the backscattered signal. Many studies have already been carried out to perform full-waveform lidar data processing and analysis. Non-linear least-squares (NLS) methods (Hofton *et al.*, 2000, Reitberger *et al.*, 2006) or maximum likelihood estimation using the Expectation Maximization (EM) algorithm (Persson *et al.*, 2005) are typically used to fit the signal to a mixture of Gaussian functions to parametrize the peaks. It was found that small-footprint lidar waveforms could be generally well modelled with a sum of Gaussian pulses (Wagner *et al.*, 2006).

We focused in this study on maximising the number of peaks detected from the waveforms: the issue is to extract as much information as possible above the noise level while limiting erroneous peak detection. The optimization step is well-known and efficient and the critical step relies on the assessment of the right number of components. The main sources of ill-detections are both the noise and the ringing effect. They are taken into account as follows: (1) the background noise is thresholded; (2) only one peak is kept when two very close echoes are detected under the lidar system resolution; (3) and finally the peaks due to the ringing effect are removed based on an amplitude ratio criterion.

In this study waveforms were decomposed into sums of Gaussian functions and the optimization method was a non-linear least-squares algorithm. To

detect the number of components, we used an improved peak detection method developed in a previous study (Chauve *et al.*, 2007). The main steps are:

- A basic detection method, based on zero crossings of the first derivative, is used at first to estimate the number and the position of the components;
- Using these values as initialisation values, a first estimation of the signal is computed;
- An iterative process is performed to find missing peaks by detecting echoes on the difference between the modelled and initial signals. If new peaks are detected, the fit is performed again. This process is repeated until no improvement is found.

This enhanced peak detection method is useful to model complex waveforms with overlapping echoes and also to extract weak echoes which are not found by multi-echo systems. Both cases often occur in vegetated areas.

3.2 GROUND POINTS CLASSIFICATION

The classification process is based on a previous work described in (Bretar *et al.*,2004). From an initial location within the point cloud, the filtering algorithm propagates following the steepest local slope over a grid topology. A neighborhood of lidar points is extracted at each grid location. The neighborhood extension is set so that the overlapping ratio between two adjacent locations should be at least 50 %. An initial estimate of the terrain elevation is performed by calculating an average value of laser point height belonging to a rank filtered subset. The filtering algorithm is based on a bipartite voting process.

Lidar points will be classified as ground or off-ground points depending on their height difference to the local terrain estimate (mean value of lidar points classified as ground points). Considering the overlapping ratio of the neighborhoods, laser points are classified several times either as ground or off-ground points. At the end of the propagation, a label corresponding to the most representative votes is affected to each lidar point.

A post-processing step is performed to detect under-terrain outliers. Such points mainly come from under-ground erroneous echoes that were extracted during waveform processing. The filter is based on a robust local plane estimation of ground points. Points located above a given threshold are removed.

3.3 DTM AND CHM COMPUTATION

DTMs are triangulated from lidar ground points and finally re-sampled on a 0.5 m resolution grid, in agreement with the spatial resolution of the lidar acquisition (4-5 pts/m2).

DSMs are computed from first echoes using the Inverse Distance Weighting (IDW) interpolation technique. CHMs are obtained by subtracting DTM from DSM.

4 RESULTS AND DISCUSSION

4.1 POINT DETECTION

Lidar waveform post-processing allows to improve the density of the final point cloud up to more than 130 % on very dense vegetated areas (see table 1). Table 1 shows that on large ground areas with only sparse trees (like in plot 1), only few additional echoes are detected. The number of detected points increases when the vegetation is getting denser in both overstory and understory vegetation.

Table 1. Statistics on the point extraction over different plots. Plot 1: sparse Black pine stand; plot 2: dense Black pine stand; plot 3: dense Sylvester Pine stand.

Area	Plot 1	Plot 2	Plot 3
Nb multi-echo points	19863	1729	1566
Nb fullwave points	25769	3305	3712
% additional points	30 %	91 %	137 %

Analyzing the differences between the fitted waveforms and the multi-echo point cloud, one can notice that the additional points come from weak and overlapping echoes that are now detected. These points are located near the tree canopy and in the understory. Only few additional points are detected on the ground due to the fact that pine crowns, although thin, are very dense and the laser beam can hardly get through them.

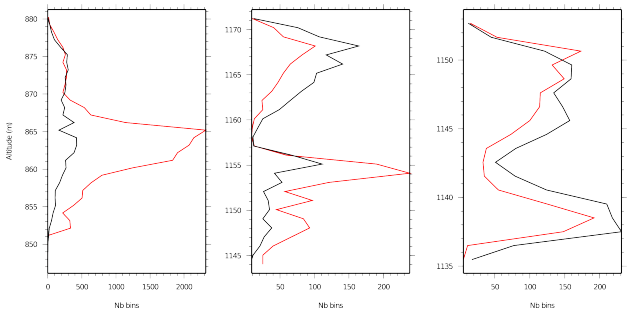

Figure 2. Histograms of 3D point cloud altitudes with 1 m bin size: Plot 1 (left), plot 2 (middle), plot 3 (right). Red lines correspond to multi-echo point cloud and black lines to the additional points extracted from full-waveform data.

Histograms on figure 2 show the altimetric distribution of multi-echo points (red lines) and of additional points detected by processing lidar waveforms (black lines). In plot 1 (left subfigure), the landscape is hilly and as a consequence there is only one wide peak corresponding to ground points and low vegetation. Due to a very low tree density the overstory peak is reduced and hardly distinguishable. In plot 2 (middle subfigure) overstory and understory can be clearly separated. The ground peak is also quite large because the slope of the plot is very high. Most of the additional points are here located in the tree canopy. In the third plot (right subfigure), which is relatively flat, both understory and overstory are very dense and almost continuous. Additional points are here located in the canopy as well as in the low vegetation.

4.2 DIGITAL TERRAIN MODELS

Table 2 summarizes the results of the comparison between 0.5 m resolution raster DTMs derived from multi-echo and full-waveform lidar data, and from terrain measurements. Means and RMSs were computed on the difference images. Results are homogeneous for all comparisons: less than 0.2 m in RMS, except for the comparison between field measurements and lidar point cloud in the first plot. In these cases (RMS = 0.57 m) errors are mainly due to an insufficient field measurement sample for describing the hilly topography of the first plot.

Table 2. Statistics on DTM (in m).

Area	Mean (m)	RMS (m)
Plot 1: difference multi-echo – fullwave	-0,02	0,15
Plot 1: difference field – multi-echo	0,08	0,57
Plot 1: difference field – fullwave	0,05	0,57
Plot 2: difference multi-echo – fullwave	0,06	0,20
Plot 2: difference field – multi-echo	-0,04	0,16
Plot 2: difference field – fullwave	0,01	0,20

Waveform processing did not improve the DTM on these two plots, because very few additional points were detected on the ground (around 6 %) for the first plot and because the classification between ground and low vegetation is still an issue in dense and near-ground vegetation (0.3 to 1 m) for the second plot.

4.3 CANOPY HEIGHT MODELS

Figures 3 and 4 show results of the comparison between multi-echo and fullwave CHMs for plots 1 and 2. On the left the CHM is computed from multi-echo point cloud and on the right the difference between CHM is computed from multi-echo and full-waveform point clouds. The range of differences in height are similar in both plots: from -3 m to about 7 to 9 m. The few negative values are located around the trees and correspond to additional points detected in the low part of the canopy that are not in the multi-echo point cloud.

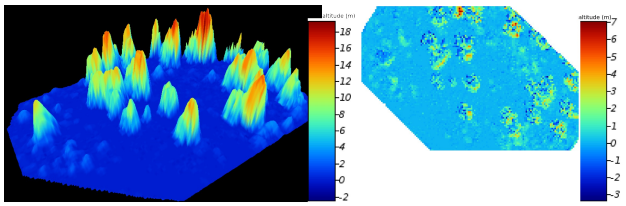

Figure 3. Plot 1: (left) CHM computed from multi-echo point cloud; (right) difference between CHM computed from fullwave and from multi-echo point clouds.

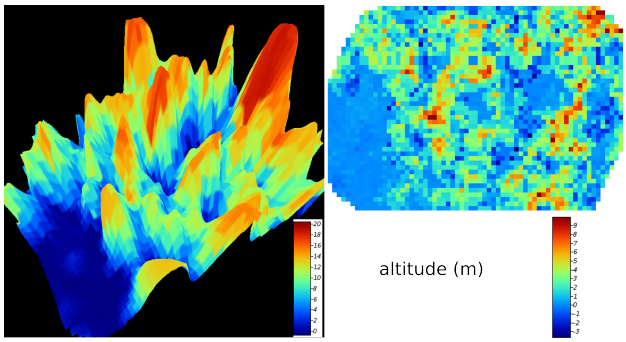

Figure 4. Plot 2: (left) CHM computed from multi-echo point cloud; (right) difference between CHM computed from fullwave and from multi-echo point clouds.

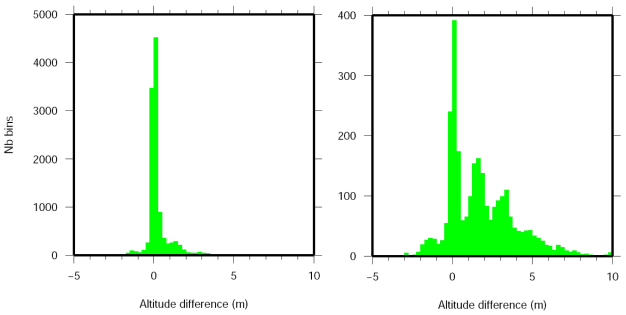

Figure 5. Histograms of CHM differences between. Plot 1 (left, 0.22 m mean difference; Plot 2 (right, 1.7 m mean difference).

Histograms of CHM differences are plotted on figure 5. These histograms are linked to the vegetation density and cannot be directly compared. Nevertheless, what is noticeable is that on a dense forest area processing waveforms significantly changes the description of the canopy: volume,

height and 3D structure are expected to be improved. Detailed validation of the canopy shape with field measurements is in progress.

5 CONCLUSIONS AND FUTURE WORK

Processing lidar waveforms has been investigated in this paper in order to extract more echoes than equivalent multi-echo data. We studied the altimetric distribution of additional points and evaluated the potential of processing waveforms to improve DTM and CHM. DTMs were finally compared to field measurements.

Improving peak detection was shown in this paper to be very successful to extract additional targets in the return waveforms. Depending on vegetation density, we detected from 30 % to 130 % additional points. These points are located mainly within the canopy and in highly dense understory. Very few additional points were detected on the ground, which explains why the DTMs were not significantly improved. On the contrary, CHM really benefited from waveform processing as the number of echoes were doubled in the overstory and inside the canopy. The 3D structure of the vegetation is thus expected to be significantly improved, and detailed field measurements are in progress to confirm this result.

Modeling raw lidar signal also enables to extract, beyond target position, additional parameters which are of interest to study the geometry and the radiometry of the targets: both echo intensity and width, and also shape parameters when complex models, such as generalized Gaussian model, are used to decompose lidar waveforms into a sum of target contributions. This information is promising to improve the classification of ground and low vegetation points, and also to identify tree species.

ACKNOWLEDGMENTS

We thank Laurent Albrech for his help during the field data measurement campaigns and Nicolas David for his help to georeference the lidar data. We also thank the GIS Draix and the Cemagref of Grenoble for providing the full-waveform lidar data. Field measurements were realized thanks to the financial support of CNES.

REFERENCES

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

Persson, A., Holmgren, J. and Soderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogram. Eng. Rem. Sensing*, 68(9): 925-932.

Nilsson M., Brandtberg T., Hagner O., Holmgren J., Persson A., Steinvall O., Stemer H., Söderman U. and H., O., 2003. Laser scanning of forest resources. The swedish experience. *ScandLaser Scientific Workshop on Airborne Laser scanning of Forests,* pp. 43-51.

Hyyppä, J., Hyyppä, H., Litkey, P., Yu, X., Haggrén, H., Rönnholm, P., Pyysalo, U., Juho Pitkänen, J. and Maltamo, M., 2004. Algorithms and Methods of Airborne Laser-Scanning for Forest Measurements. *In International conference "Laser-scanners for Forest and Landscape Assessment"*, XXXVI, part 8/W2: pp. 82-89.

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

Naesset, E. and Bjerknes, K.-O., 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Rem. Sens. Environ*., 78(3): pp. 328-340.

R. Dubayah, J. Blair, Lidar Remote Sensing for Forestry Applications, Journal of Forestry 98 (6) (2000) 44–46.

Chauve, A., Mallet, C., Bretar, F., Durrieu, S., Pierrot Deseilligny, M. and Puech, W., 2007. Processing full-waveform LiDAR data : modelling raw signals. *In Laser Scanner et Sylvilaser 2007*, 12-14 sept. 2007.

Wagner, W., Ullrich, A., Ducic, V., Melzer, T. and Studnicka, N., 2006. Gaussian decomposition and calibration of a novel smallfootprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing* 66, 100–112.

Hofton, M., Minster, J. and Blair, J., 2000. Decomposition of Laser Altimeter Waveforms. *EEE Trans. Geosci. Rem. Sens.* 38(4), 1989–1996.

Reitberger, J., Krzystek, P. and Stilla, U., 2006. Analysis of full-waveform lidar data for tree species classification. *In ISPRS Photogrammetric Computer Vision (PCV)*, Bonn, Germany, pp. 228–233.

Persson, A., Söderman, U., Töpel, J. and Alhberg, S., 2005. Visualization and Analysis of Full-Waveform Airborne Laser Scanner Data. *In ISPRS workshop 'Laserscanning 2005'*, Enschede, The Netherlands, pp. 103–108.

Bretar, F., Chesnier, M., Pierrot-Deseilligny, M. and Roux, M., 2004. Terrain modeling and airborne laser data classification using multiple pass filtering. *In Proc. of the XXth ISPRS Congress*, IAPRS, Vol. XXXV part B, ISPRS, Istanbul, Turkey, pp. 314–319.