Accuracy of a DTM Derived from Full-waveform Laser Scanning Data under Unstructured Eucalypt Forest: A Case Study

Luísa GOMES PEREIRA and Gil GONÇALVES, Portugal

Key words: full-waveform laser scanning, DTM, vertical accuracy, unstructured Eucalypt forest

SUMMARY
A Digital Terrain Model (DTM) is fundamental for extracting several forest canopy structure metrics from data acquired with small-footprint airborne laser scanning (ALS). This modern remote sensing technology is based on laser measurements from a laser system mounted on an aircraft and integrated with a geodetic GNSS receiver and an inertial measurement unit (IMU) or inertia navigation system (INS). In the context of a research project for deriving forest inventory parameters and fuel variables under eucalypt stands in Mediterranean climates, the vertical precision of the DTM obtained by automatic filtering of full-waveform ALS data had to be evaluated. The DTM accuracy estimation on a study area with peculiar characteristics, which are often avoided in related studies, will also allow verifying the performance of full-waveform ALS systems. The accuracy estimation is carried out in a novel way. By novel way, it is meant an exhaustive, well-planned collection of reliable control data in forest environment. The collection of the control data involves the production of DTM on 43 circular plots (radius = 11.28m) using total stations and geodetic GNSS receivers. These DTM, with a total of 3356 points, allowed one to evaluate consistently and reliably the vertical accuracy of the terrain surface produced with ALS under a eucalypt forest. This global accuracy, expressed by the Root Mean Square Error (RMSE) of the vertical differences between the field surveyed surface and the ALS derived DTM surface is 0.15m (mean=0.08m and std=0.09m). This impressive value indicates that, for an ALS point cloud density of 10pts/m² and footprint of 20 cm, the methodology used to extract the DTM from full-waveform ALS data under an unstructured eucalypt forest is very accurate. In this article it is addressed both the strategy adopted to collect the control data and the quality assessment of the DTM produced by means of the ALS data.
1. INTRODUCTION

In inland Portugal, due to economical and technical reasons the most important raw material for hardwood pulps is eucalypt (Eucalyptus globulus Labill.). Because of its high growth rate and its high pulp yield the eucalypt wood is cheap. The resulting pulps have also excellent technical properties. In Europe, the growth of eucalypt trees is limited to Mediterranean and frost-free climates, which are prevalent only in Portugal and in some parts of Spain (Patt et al., 2006). In Portugal, eucalypt stands occupy 22.9% of its 31,700 Km$^2$ of forest. Its management is mainly done by short rotations of about 10 to 12 years plantations and with a topping tree diameter situated between 5 and 10 cm (Tomé et al., 2007).

In the context of a research project for estimating forest inventory parameters and fuel variables under eucalypt stands in Mediterranean climates, the vertical precision of the DTM obtained by automatic filtering of full-waveform ALS data had to be evaluated. In fact, the methods for forest inventory and fuel variables estimation from ALS data rely on the derivation of a DTM, corresponding to the ground surface, and of a DSM corresponding to treetops and shrub heights (Hyppa et al., 2008). Hence, the errors in the DTM will propagate to the DSM and therefore to the derivation of the forest inventory variables, like the tree height, and of the fuel variables, like the shrub height.

The vertical accuracy of a DTM produced with ALS data is influenced by the errors/inaccuracies originated from the (Huising and Gomes Pereira, 1998; Hodgson and Bresnahan, 2004; Su and Bork, 2006; Hyppa et al., 2008):

- ALS system: GPS, INS, and laser scanner;
- strategy for acquisition of ALS data: point density, first/last pulse, several pulses, flight height, scan angle;
- conversion of ALS data (full-waveform modelling, filtering and interpolation methods);
- characteristics of the target surface that is, topography and type(s), density and spatial distribution of vegetation.

While a general understanding of the accuracy of the ALS systems has been achieved, the accuracy of the derived DTM from ALS data in a forest environment has not been thoroughly evaluated (Hodgson and Bresnahan, 2004; Reutebuch et al., 2003). Moreover, the full-waveform data have, in comparison to conventional pulsed ALS data, the advantage of echo detection being done in post processing making the ranging process more robust. This may certainly lead to higher accuracy of the derived distances and thus to more accurate DTM (Ullrich et al., 2008).

In general, the few published empirical studies assessing the vertical accuracy of DTM created by using ALS data in forested environments were done in boreal forest zones. Table 1 lists some published accuracy values of DTM created by using ALS data in different forest
environments. A limitation of the studies is that they do not allow one to predict the DTM accuracy for they do not address all the four accuracy influencing factors above listed.

Table 1: Published accuracy values of DTM produced with ALS data.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Area extent (km²)</th>
<th>ALS Point density (pts/m²)</th>
<th>Nº of check points</th>
<th>RMSE and bias (cm)</th>
<th>Characteristics of study area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kraus and Pfiefer (1998)</td>
<td>91</td>
<td>0.1</td>
<td>466</td>
<td>57; 20</td>
<td>Beech.</td>
</tr>
<tr>
<td>Hyyppa et al. (2000)</td>
<td>0.014</td>
<td>8-10</td>
<td>740</td>
<td>22 (Standard deviation) 32; 22</td>
<td>Boreal forest. Norway spruce and Scots pine. Coniferous forest (Douglas-fir in western Washington region). ALS data of six land cover categories: pavement, low grass, high grass, brush/low trees, evergreen and deciduous.</td>
</tr>
<tr>
<td>Reutebuch et al. (2003)</td>
<td>5</td>
<td>4</td>
<td>347</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hodgson and Bresnahan (2004)</td>
<td>2000</td>
<td>0.25</td>
<td>654</td>
<td>21.4; not reported</td>
<td></td>
</tr>
<tr>
<td>Takahashi et al. (2005)</td>
<td>two study areas of 2500 and 625</td>
<td>11.2</td>
<td>283</td>
<td>39; 14</td>
<td>Sugi plantations (Cryptomeria japonica D. Don)</td>
</tr>
<tr>
<td>Yu et al. (2005)</td>
<td>8 study areas of 1</td>
<td>10</td>
<td>1474</td>
<td>9 (Standard deviation) 30 (Standard deviation)</td>
<td>Boreal Forest (Kalkkinen test site). Norway spruce and Scots pine.</td>
</tr>
<tr>
<td>Su and Bork (2006)</td>
<td>27</td>
<td>0.75</td>
<td>256</td>
<td>59; 2</td>
<td>Riparian Meadows, Upland Grasslands, Shrublands and Aspen Forest.</td>
</tr>
<tr>
<td>Hollaus et al. (2006)</td>
<td>128</td>
<td>1.8</td>
<td>2200</td>
<td>Flat areas: 10; no reported Sloped areas (&gt;60º) 50; not reported</td>
<td>ALS data along road-sides in a Alpine valley.</td>
</tr>
</tbody>
</table>

In this paper it is assessed the vertical accuracy of a DTM obtained by automatic filtering of full-waveform ALS data. In addition to the peculiar characteristics of the study area, which are often avoided in related studies, the accuracy estimation is carried out in a novel way. By novel way, its meant an exhaustive, well planned collection of reliable control data under an...
eucalypt forest comprising regular as well as irregular spacing plantations. The understory is heterogeneous and is mainly composed of whin, fern, heath, and baccharis. Prior to the vertical accuracy assessment of the DTM, the quality of the delivered laser data was also evaluated. To this end, measurements on horizontal and inclined bare surfaces were used.

2. STUDY AREA AND DATA ACQUISITION

2.1 Study area

The study area was selected nearby the city of Águeda, in the district of Aveiro, situated in the Northern part of Portugal. The selected area measures 900 ha (Figure 1-a) and, while dominated by eucalypt plantations, also includes some pine stands and few built-up areas. The forest stands in the area comprise regular as well as irregular spacing plantations, both even and uneven-aged stands, and stands with as well as without extensive undergrowth. The topography of the study area varies from gentle to steep slopes, with altitudes varying from 27 to 162 m (Figure 1-b).

![Figure 1: (a) Localization of study area within Portugal and its delimitation; (b) DTM of the study area.](image)

2.2 Acquisition of the ALS and image data

The ALS data were acquired on the 14\textsuperscript{th} of July of 2008. The laser system utilized was the Litmapper 5600. This system has, as main hardware components the high-resolution laser scanner from RIEGL, the LMS-Q560 with full-waveform processing, the AEROCtrl GPS/IMU system for the precise determination of position and attitude with an Inertial Measurement Unit (IMU) of 256Hz raw data rate and based on differential GPS, and the DigiCAM, a medium-format Airborne Digital Camera System operated simultaneously with the ALS system. The main system software relates to the AEROffice - the GPS and IMU...
data post processing software - the RiWorld and RiAnalyze software from Riegl for respectively coordinate transformation and processing of the ALS data, and the software bundle from TerraSolid. This comprises the software parts TerraScan, TerraModeler and TerraMatch. TerraScan is the dedicated software solution for manipulation, viewing, and classifying ALS data. TerraModeler is a full featured terrain modelling application for creating digital terrain/surface models (DTM/DSM). Terra Match is used for automated correction of ALS data (http://www.igi-systems.com/products/litemapper/components.htm, 19.01.2010).

Together with the ALS data were acquired stereo-pairs of CIR images. The parameters selected for ALS and image data acquisition are listed in table 2. The flight trajectory and the image expositions are shown in figure 2.

<table>
<thead>
<tr>
<th>ALS</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALS sensor: Riegl LM5600</td>
<td>Camera sensor: Digicam H39 + 50mm focal length</td>
</tr>
<tr>
<td>Wavelength: 1064 nm</td>
<td>Wavelengths in CIR mode (nm):</td>
</tr>
<tr>
<td>Scan angle: 45°</td>
<td>B=500-620; G=580-800; R=800-1000</td>
</tr>
<tr>
<td>Pulse rate: 150 KHz</td>
<td>Image repetition rate: 1.9sec</td>
</tr>
<tr>
<td>Effective Measurement rate: 75kHZ</td>
<td></td>
</tr>
<tr>
<td>Beam divergence: 0.5 mrad</td>
<td>Flying height above terrain: 700m</td>
</tr>
<tr>
<td>Ground speed: 46.26 m/s</td>
<td>Overlap: 60%</td>
</tr>
<tr>
<td>Flying height above terrain: 700m</td>
<td>Sidelap: 30%</td>
</tr>
<tr>
<td>Swath: 497m</td>
<td>Nr. of pixels forward: 7216</td>
</tr>
<tr>
<td>Sidelap: 70%</td>
<td>Nr. of pixels sideward: 5412</td>
</tr>
<tr>
<td>Single run density: 3.3pt/m2</td>
<td>Distance between lines: 150m</td>
</tr>
<tr>
<td>Expected final density: 9.9pt/m2</td>
<td>Ground sampling distance (GSD): 8.2cm</td>
</tr>
<tr>
<td>Distance between lines: 150m</td>
<td></td>
</tr>
</tbody>
</table>
3. METHODS

3.1 Procedures for acquisition and quality assessment of reference data

Reference data are needed to verify, in terms of precision and reliability, the DTM produced by means of the laser data and a filtering algorithm (3.2.2). The strategy for the reference data collection was not straightforward. In forest areas, the collection of these data is time consuming, mainly in plots with a high density of shrubs and trees. Furthermore, because the data were georeferenced geodetic GNSS receivers had to be used.

The planning of the topographical survey was based on that of the forest inventory, which started by selecting 43 plots within the study area (red dots in figures 3-a and 3-b). The DTM was represented by the coordinates of terrain points located aside trees, which give also the locations of the trees, and by the coordinates of prominent terrain points, like those on breaklines (Figure 3-c).

It was also decided to measure grids of approximately 1m x1m points on bare surfaces (roads and fields; Figures 4-a, 4-b and 4-c) to assess the quality of the delivered ALS data.
This information was collected by means of a topographic survey using the irradiation method. The coordinate system in which the ALS and image data were collected is the WGS84 UTM zone 29, for X and Y coordinates, and the WGS84 ellipsoidal height for the Z coordinate (from now on referred to as absolute coordinates X, Y and H). Because this is not a local system, the geographic information collected in the field had to be converted to that system by using the Global Positioning System (GPS). To this end, it was decided to attach to each plot two points, named GPS Base, whose coordinates were measured with two GNSS receivers. These two points were placed as close as possible to the plot and as much as possible in an opened space (Figure 5). This criterion turned out to be difficult to fulfil in the study area.

The method used to measure the coordinates of the two points that constitute the GPS Base was the relative positioning by using a fixed receiver on a geodetic pillar with known coordinates on the WGS84 UTM zone 29 system. This method, in post-processing, is the most precise and may reach levels of precision in the order of the cm.

The GPS observations were collected according to a previous planning dictated by the following specifications: minimum time of observation of 60 minutes, depending on the point surroundings (normally it took 120 minutes), minimum number of 7 satellites and value of the Position Dilution of Precision (PDOP) less than 3. In total, the coordinates of 82 points, i.e., 41 GPS Bases were measured. The number of GPS bases does not coincide with that of plots because, firstly two of the plots were disregarded due to inaccessibility and secondly two GPS Bases were assigned to two plots each.
The quality of these coordinates was assessed by using the differences in X, Y and H between the coordinates of the two points of the GPS Bases. These differences were computed by using their coordinates estimated in a local system (DXYl and DZl) and in the absolute system (DXYA and DZA). The local coordinates are computed by means of topographic methods using a total station. It is assumed that the errors that originate from this source is negligible when compared to those originated by using GPS methodology. In planimetry, the mean error and RMSE are respectively 1.7 cm 2.5 cm whilst in altimetry they have the values of 1.9 mm and 2.6 cm respectively.

3.2 ALS data processing

3.2.1 Full waveform processing

The full-waveform laser data were processed with the RiAnalyze software from Riegl. A maximum of 5 returns were obtained with a minimum vertical separation of 50 cm. The processed ALS data is adjusted (corrected for errors) and stored in the form of LAS blocks (Figure 6). A sample of the laser data and the corresponding digital orthophoto are shown in figure 7. In this figure the colours assigned to each laser point corresponds to its height. A sample of laser data within the study area is shown in figure 8 in 3-D.

Figure 6: Spatial data organization in the study area (approx. 120 GB).

Figure 7: A sample of the ALS data (left) and the corresponding digital orthophoto (right).
3.2.2 Point cloud filtering

The point cloud obtained by processing the full-waveform data contains both laser returns from objects located on the ground (buildings, trees, understory vegetation) as the ground itself. Moreover, in vegetated areas the laser pulse can penetrate the canopy, travelling between the branches and leaves and be reflected by the understory (e.g. sub-canopy trees and shrubs) and more often by the ground.

Since we are interested in the reconstruction of the DTM surface it is necessary to remove from the point cloud the non-ground returns. Removing non-ground measurements from the point cloud is here referred to as filtering. In order to find the most “probable” ground returns from the initial point cloud it will be used, in the filtering process the last return of each pulse.

Points in a point cloud are classified as either ground or non-ground returns using a progressive TIN (Triangular irregular Network) densification algorithm implemented in the TerraScan software v4.006 (www.terrasolid.fi/en, 19.01.2010) and based on the algorithm published in Axelsson (1999 and 2000). This algorithm begins by constructing a sparse TIN with the lowest point that is within a square neighbourhood of a certain size (BldSz). The TIN is densified by adding points of the set of unclassified points and that satisfy certain criteria. The criteria use the distance between that point and the closest triangle plane and the angle between the line that join each of the triangle vertexes and the point and the line that join its projection on the triangle plane and the triangle vertex. These four values are compared against thresholds (IterDst for the distance and IterAng for the angles). The densifying of the TIN stops when no more unclassified points satisfy the criteria or if a certain ground point density is achieved. An additional threshold (MaxAng) can be used to control the maximum terrain slope allowed at each triangle of the TIN surface.

In this work, the parameters used for filtering the laser point cloud were determined experimentally and are given in the table 3.
Table 3: Parameters used in the filtering process.

<table>
<thead>
<tr>
<th>BldSz (m)</th>
<th>MaxAng (º)</th>
<th>IterDst (m)</th>
<th>IterAng (º)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>88</td>
<td>1.4</td>
<td>6</td>
</tr>
</tbody>
</table>

3.3 DTM accuracy assessment

The accuracy assessment of the DTM produced after the filtering process of the laser data relates to the estimation of the mean and RMSE of the differences (dz) between the H values of the reference points (3.1) and those at the same planimetric locations of ALS terrain points. These ALS terrain points are interpolated from a TIN computed with the filtered ALS terrain points (3.2.2). It was decided to create a TIN from the filtered ALS data instead of from the reference data because the density of ALS terrain points is higher than that of reference points (3.8 versus 0.2 points/m$^2$). The higher density of ALS terrain points per m$^2$ than that of the reference points implies that points on the terrain are represented in the ALS filtered data but not in the reference data. Therefore, it is more adequate to interpolate the laser data to the planimetric positions of the reference data. The TIN format is also more appropriate than the grid format once the interpolated height is influenced only by the unaltered height values of its neighbours (3 neighbours when linear interpolation is used).

4. RESULTS

The quality of the DTM obtained with the laser data was computed for each plot individually, by adopting the strategy above detailed. The estimated values for the mean, standard deviation and RMSE are plotted in figure 9.

As it may be seen in the figure 9, the values of the mean of residuals vary, in absolute value, between 5 cm and 28 cm, whilst the RMSE varies between 4 cm and 29 cm. When considering all the plots together, i.e, 3356 points the value of the mean of residuals is 8 cm whilst the RMSE is 15 cm. These are very good values considering the characteristic of the plots. Furthermore, there has been no manual editing, i.e, manual removal of gross errors.
A correlation between the quality of the DTM and the vertical structure of the vegetation was derived and expected to be high. Figures 10, 11 and 12 show the linear regression between the values estimated for the RMSE per plot and respectively, the percentages of the number of points classified by the filtering process as terrain, as shrubbery with height varying from 10 cm until 2 m and as higher vegetation cover.

Figure 9: Values per plot of the mean, standard deviation and RMSE of residuals.

Figure 10: Linear regression between the value of RMSE and the process-derived percentage of terrain points per plot.
Contrary to what was expected, the low values obtained for the correlation coefficient \((r_1=0.08 ; r_2=0.14; r_3=0.10)\) indicate that there is almost no correlation between the quality of the DTM and the vertical structure of vegetation. The same tests and results were obtained when using the mean of the residuals instead of the RMSE. This means that the algorithm published in Axelsson (1999 and 2000), together with the high density of the laser points per \(m^2\) (in mean 10), is quite robust and insensitive to the obstacles on the terrain. The spatial distribution of the RMSE values, per plot, within the study area is illustrated in figure 13. As in previous studies (Reutebuch et al., 2003), there seems to be no correlation between the RMSE values and the spatial distribution of the plots.
In order to verify the contribution of the quality of the ALS data alone into the assessment of the quality of the DTM, the quality of the laser data was verified. To this end, three DTM were produced, by using topographic means, on flat bare surfaces (a paved road, a football field and an arable field). The assessment consisted in computing the mean and RMSE of residuals. These were obtained by computing the differences, at the same planimetric locations, between the measured and interpolated H values of the reference and ALS points, respectively. The latter were estimated by using a TIN constructed with the measured ALS points. The obtained mean and RMSE are listed in table 4. These values are quite good and the worst, obtained for the arable land results from the fact that the pole used for measuring sunk into the fresh soil for a few centimetres. This pole is used to transport the prism needed to measure the coordinates of the DTM points by topographic means.

<table>
<thead>
<tr>
<th>Surface</th>
<th>Mean (cm)</th>
<th>RMSE (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paved road</td>
<td>2.8</td>
<td>3.2</td>
</tr>
<tr>
<td>Football field</td>
<td>1.8</td>
<td>3.4</td>
</tr>
<tr>
<td>Arable field</td>
<td>6.8</td>
<td>8.4</td>
</tr>
</tbody>
</table>

In addition to this evaluation of the accuracy of the ALS data the mean and standard deviation of the residuals between the H values of the GPS point bases and those derived from the laser data at the same location were computed. A mean of 0.3 cm and a RMSR of 6.8 cm were obtained. The difference in the mean of the residuals between that computed with the altitudes of the GPS bases and those computed with the altitudes in flat surfaces is due to the fact that whilst the altitudes of the GPS bases are spread all over the study area, and thus involving all the 30 laser strips, the other use only 7 laser strips, 5 of which are common.
5. DISCUSSION AND CONCLUSIONS

A study as the one here presented is quite demanding. To properly assess the quality of full-waveform ALS data and of a derived DTM in an area of 900 ha with unstructured eucalypt forest requires an exhaustive and well-planned collection of reliable control data. The georeferencing of these data makes use of GNSS techniques, which are very time consuming in environments were a high density of high vegetation predominates. Furthermore, topographic measurements in plots covered with high shrubs are also difficult to execute. The accuracy of the produced DTM is very good and suitable for a great number of applications. The value of 15 cm in RMSE is small than that published in several studies carried out in less chaotic forests (table 1). It also compares to the RMSE value claimed 11 years ago in Huising and Gomes Pereira (1998), but for flat surfaces, and still referred in several studies. For similar surfaces, a value of 3 cm in RMSE was instead obtained which shows, therefore, an improvement in the accuracy of the data collected with the new ALS technique.

The filtering algorithm based on the filter of Axelsson (1999 and 2000) also proved to be very robust to terrain with the aforesaid characteristics. The low values obtained for the correlation coefficients between the RMSE values per plot and the percentages of the number of points classified by the filtering process as terrain, as shrubbery with height varying from 10 cm until 2 m and as higher vegetation cover indicate that there is almost no correlation between the quality of the DTM and the vertical structure of vegetation. This means that the Axelsson algorithm, together with a high density of the laser points per m$^2$ (in mean 10), is quite robust and insensitive to the obstacles on the terrain. The spatial distribution of the RMSE values per plot shows also no correlation between the RMSE values and the spatial distribution of the plots.

To conclude, the full-waveform ALS system proved to be superior to the traditional laser pulsed system, in terms of data accuracy, for the production of DTM even in areas with an unstructured high-density forest.

ACKNOWLEDGEMENTS

The present study was funded by the Foundation for Science and Technology (FCT) of Portugal in the framework of the project PTDC/AGR-CFL/72380/2006.

REFERENCES


BIOGRAPHICAL NOTES

Dr. Luísa Pereira is a survey engineer with a PhD. in Digital Photogrammetry. She is involved in education and research for more than 25 years. During the last 10 years she has been working on Aerial Laser Scanning mainly on error analysis and DTM production for several applications. She is a Professor at Aveiro University.

Dr. Gil Gonçalves is a survey engineer with a PhD. on Geographic Information Science. He is involved in education and research for more than 20 years. Currently, he is Professor at the Faculty of Sciences and Technology of the University of Coimbra. Dr. Gonçalves’s main research interest is in the area of terrain modelling with special emphasis on the integration of Airborne Laser Scanning data and aerial images.

CONTACTS

Dr. Luísa Gomes Pereira
Aveiro University/ESTGA
Apartado 473
3750-170 Águeda
Portugal
Tel.: +351 234 611 500/501
Fax: +352 235 611540
Email: luisapereira@ua.pt
Web site: www.estga.ua.pt

Dr. Gil Gonçalves
University of Coimbra, Faculty of Sciences and Technology
Apartado 3008,
3001-454 Coimbra
PORTUGAL
Tel.: +351 239 791150
Fax: +351 239 793069
Email: gil@mat.uc.pt
Web site: http://www.mat.uc.pt/~gil/