

# Evaluation of the use of low-density LiDAR data to estimate structural attributes and biomass yield in a short-rotation willow coppice: an example in a field trial

María Castaño-Díaz<sup>1</sup> · Pedro Álvarez-Álvarez<sup>1</sup> · Brian Tobin<sup>2</sup> · Maarten Nieuwenhuis<sup>2</sup> · Elías Afif-Khoury<sup>1</sup> · Asunción Cámara-Obregón<sup>1</sup>

Received: 16 February 2017 / Accepted: 28 September 2017 / Published online: 23 October 2017  
© INRA and Springer-Verlag France SAS 2017

## Abstract

• **Key message** LiDAR data (low-density data, 0.5 pulses  $m^{-2}$ ) represent an excellent management resource as they can be used to estimate forest stand characteristics in short-rotation willow coppice (SRWC) with reasonable accuracy. The technology is also a useful, practical tool for carrying out inventories in these types of stands.

• **Context** This study evaluated the use of very low-density airborne LiDAR (light detection and ranging) data (0.5 pulses

$m^{-2}$ ), which can be accessed free of charge, in an SRWC established in degraded mining land.

• **Aims** This work aimed to determine the utility of low-density LiDAR data for estimating main forest structural attributes and biomass productivity and for comparing the estimates with field measurements carried out in an SRWC planted in marginal land.

• **Methods** The SRWC was established following a randomized complete block design with three clones, planted at two densities and with three fertilization levels. Use of parametric (multiple regression) and non-parametric (classification and regression trees, CART) fitting techniques yielded models with good predictive power and reliability. Both fitting methods were used for comprehensive analysis of the data and provide complementary information.

• **Results** The results of multiple regression analysis indicated close relationships ( $R_{fit}^2 = 0.63–0.97$ ) between LiDAR-derived metrics and the field measured data for the variables studied (H,  $D_{20}$ ,  $D_{130}$ , FW, and DW). High  $R^2$  values were obtained for models fitted using the CART technique ( $R^2 = 0.73–0.94$ ).

• **Conclusion** Low-density LiDAR data can be used to model structural attributes and biomass yield in SRWC with reason-

**Handling Editor:** Aaron R. Weiskittel

## Contribution of the co-authors

- María CASTAÑO-DÍAZ = co-writing the paper and analyzing the data.
- Pedro ÁLVAREZ-ÁLVAREZ = co-writing the paper and analyzing the data.
- Brian TOBIN = co-writing the paper and supervising the work.
- Maarten NIEUWENHUIS = co-writing the paper and supervising the work.
- Elías AFIF-KHOURI = supervising the work.
- Asunción CÁMARA-OBREGÓN = supervising the work and coordinating the research project.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13595-017-0665-7>) contains supplementary material, which is available to authorized users.

✉ María Castaño-Díaz  
castanodmaria.uo@uniovi.es

Pedro Álvarez-Álvarez  
alvarezpedro@uniovi.es

Brian Tobin  
brian.tobin@ucd.ie

Maarten Nieuwenhuis  
maarten.nieuwenhuis@ucd.ie

Elías Afif-Khoury  
elias@uniovi.es

Asunción Cámara-Obregón  
camara@uniovi.es

<sup>1</sup> GIS-Forest Research Group, Department of Organism and Systems Biology, University of Oviedo, E-33600 Mieres, Spain

<sup>2</sup> UCD Forestry, Agriculture and Food Science Centre, University College Dublin, Belfield, Dublin 4, Ireland

able accuracy. The models developed can be used to improve and optimize follow-up decisions about the management of these crops.

**Keywords** Willow · Mining land · Energy crops · SRC · Airborne laser scanning · LiDAR

## 1 Introduction

The prospects of successfully achieving and maintaining sustainable energy production worldwide depend on the increased use of renewable resources in general and biomass in particular (Edenhofer et al. 2011). One of the best ways of ensuring the long-term availability of biomass for producing renewable energy is to establish and grow new perennial energy crops, which can also add value to marginal land (Rosso et al. 2013) or can be used for bioremediation purposes.

Biomass plantations are an attractive source of renewable energy (González-Ferreiro et al. 2013) and also have many other advantages. Some of the reasons why crops are grown for bioenergy purposes include the recovery of economic activities in rural areas, provision of a neutral CO<sub>2</sub> balance, and restoration of degraded land.

Depending on the final destination (heat and/or electricity), three types of biomass can be produced: oilseed, alcohol, and lignocellulose (IDAE 2007). Crops that produce lignocellulosic biomass (fiber crops) can be used to produce both heat and electricity and can be grown as short-rotation coppice (SRC). Producing these so-called energy crops is considered one of the most energy-efficient methods of carbon conversion, as growing the crops is considered an efficient means of reducing greenhouse gas emissions (Styles and Jones 2007).

Short-rotation plantations can be established on various types of land, including marginal land (Broeckx et al. 2012). Zurba et al. (2013) recommended planting SRC on marginal land and brownfields, in parallel with other sustainable land management options. Planting SRC on this type of land may also contribute in the long term to improving soil quality and biodiversity, protecting groundwater and preventing soil erosion (Kuzovkina and Quigley 2005; Zurba et al. 2013). The use of Salicaceae (*Salix* and *Populus* spp.) provides several advantages: the ease of propagating plants from cuttings (low production cost and easy to establish), the wide range of improved genetic material available, production of high biomass yields in a short time, and vigorous coppicing regrowth after cutting (Keoleian and Volk 2005). Taking into account the wide adaptability of members of the genus *Salix* to extreme conditions and to nutrient-impooverished and polluted soils (Kuzovkina and Quigley 2005), SRC willow can be established on marginal land or in soils that are not suitable for agricultural exploitation (Jama and Nowak 2012). Indeed, short-rotation

willow coppice (SRWC), together with *Populus* spp., *Eucalyptus* spp., and *Robinia pseudoacacia* L., is one of the most promising bioenergy cropping systems for use in temperate regions of Europe (Venturi et al. 1999) as well as in Canada and the USA (Tahvanainen and Rytko 1999; Weih 2004).

The region of Asturias (north-western Spain) was a major coal-producing region during the past century. Although coal mining continues to be one of the most important sources of employment in the region (Paredes-Sánchez et al. 2016; Suárez-Antuña 2005), the sector is currently in recession, and large areas of mining land have been abandoned. The mining company *Grupo Hunosa* currently owns up to 700 ha of former mining land that is suitable for machine-based establishment of forest energy crops. This is currently considered the best option for use of this land, despite the difficulties in establishing energy crops (unfavorable soil structure/properties in these degraded areas). To date, the only trials involving SRC energy crops in Asturias are those associated with research projects (7 ha). At present, a commercial plantation (for bioenergy purposes) is being established in 20 ha of abandoned mine land of similar characteristics to those considered in the present study.

In 2008, an experimental trial with willow energy crops was established in abandoned mining land in Asturias. The aims of this experimental trial were to obtain information about structural attributes and biomass production in an SRWC crop established in a restored coal mining area and to evaluate the effects of clone, fertilization, and planting density on crop yield. For this purpose, detailed and comprehensive field inventories were conducted in order to obtain as much information as possible about the development of the energy crop. Forest inventories were used to estimate multiple parameters at plot level, including structure and biomass production.

For some decades now, remote sensing has enabled information about forest biomass (particularly in extensive forest areas) to be obtained at a wide range of spatial and temporal scales, thus greatly reducing costs and the amount of fieldwork required (Montealegre Gracia et al. 2015). The correlation between the spectral response of vegetation and structural attributes or biomass production has been investigated in numerous studies in which active sensors were used (Estornell et al. 2011; González-Ferreiro et al. 2012; Næsset 2002; Næsset and Gobakken 2008).

The interest shown by the aforementioned company in developing and applying new procedures and the possibility of obtaining data (forest structure and other forest variables) from airborne sensors provides a valuable opportunity to quantify the resources obtained directly from SRC. This represents a breakthrough in this field, as carrying out the field inventories necessary for adequate planning and monitoring of the forest energy plantations (characterized by high tree densities of 5000–20,000 plants ha<sup>-1</sup>, high

number of shoots per stool, etc.) is tedious and time consuming.

The use of free light detection and ranging (LiDAR)-derived data provides certain advantages such as smaller estimation errors, a reduction in the duration of field inventories and the ability to cover larger areas of land.

LIDAR is one of the most important technologies developed in this field in recent years. This technique is already being used successfully to evaluate total forest area, improving the accuracy of forest inventories and reducing the cost and time spent on these (Eid et al. 2004; González-Ferreiro et al. 2012; Wehr and Lohr 1999). LiDAR is an active remote sensing system based on the use of a laser sensor and the application of various techniques to determine the distance from a laser transmitter to an object Sánchez Martínez et al. (2011). This distance is established by measuring the time delay between emission of a signal and detection of its reflection (Tanarro 2010). The method is therefore a combination of three different technologies: laser telemetry, the global positioning system (GPS), and inertial measurement units (IMUs). The laser beam, emitted at a frequency of thousands of energy pulses per second toward the earth, creates a dense strip of 3D points (Manue 2007). This 3D point cloud is based on accurate measurement by a plane-mounted pulse sensor, which calculates the distance separating it from the earth's surface and objects existing on it (Magdaleno-Mas and Martínez-Romero 2006). As the position and orientation of the sensor are known for each pulse emitted, each return signal has unique three-dimensional coordinates. LiDAR data have been captured for the entire Spanish territory under the National Aerial Orthophotography Plan (PNOA). Data were collected during 2012 in the region of Asturias.

LiDAR technology has been used successfully to characterize numerous types of forest stands (Hayashi et al. 2014; Lefsky et al. 2002; Means et al. 1999; Means and Acker 2000; Næsset et al. 2004). However, in forest inventories, the technique has been found to underestimate height (Clark et al. 2004; Næsset 1997; Zimble et al. 2003), and some authors (Falkowski et al. 2006) have suggested that higher density data (6–8 pulses  $m^{-2}$ ) are required for forest monitoring. However, other studies have shown that a low pulse density is sufficient for establishing strong correlations with the main attributes measured in forest inventories (Hawbaker et al. 2010; Means and Acker 2000; Thomas et al. 2006). Although studies based on the use of low-density discrete-return LiDAR to determine forest structure have been reported (Coops et al. 2007; Hall et al. 2005), to date only basal area has been estimated in short-rotation coppice (SRC) (Seidel and Ammer 2014). Nonetheless, the structure of SRCs facilitates the application of LiDAR technology as these are dense, rather uniform stands with little or no accompanying vegetation. These features favor good correlations between variables measured in forest inventories and those measured using LiDAR technology.

The main objective of this study was to assess the usefulness of low-resolution discrete return LiDAR (0.5 pulses  $m^{-2}$ ) data to estimate structural attributes and biomass production with the aim of facilitating management of an SRC plantation. For the purposes of this study, we developed statistical models that relate the information provided by the LiDAR to the data obtained in detailed field inventories conducted in the study area (Montealegre Gracia et al. 2015). The methods were evaluated by complementary techniques: parametric multiple linear regression, which enabled us to develop predictive models, evaluated by  $R_{fit}^2$  and RMSE in order to indicate the accuracy of the fits, and non-parametric classification and regression trees (CART), which provided more detailed, descriptive information about the variables. The main aims of the study were (i) to estimate the forest structure and the productivity of SRWC and (ii) to apply and compare the use of different types of model fitting methods (multiple linear regression and CART).

## 2 Material and methods

### 2.1 Study area

#### 2.1.1 Location of the study area

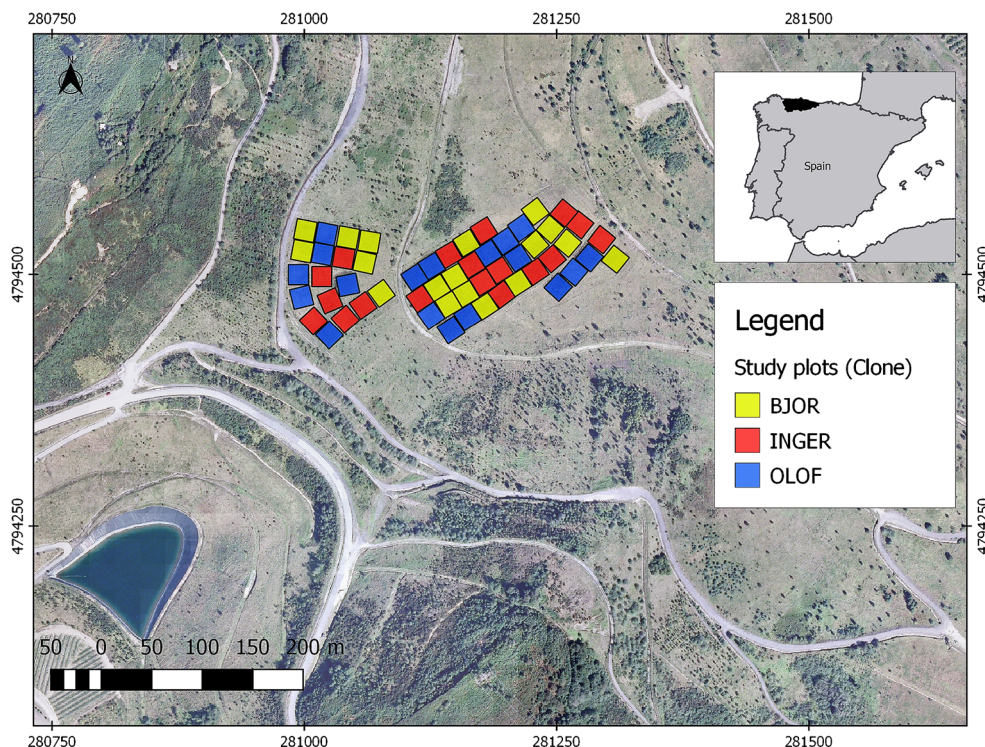
The experimental trial included three commercial willow clones and covered an area of  $\approx 2$  ha in the region of Asturias (north-western Spain) (Fig. 1). The *Salix* energy crop trial was established in May 2008 in restored land surrounding an abandoned opencast coal mine, denominated Mozquita (ETRS89 UTM 30 N, N: 4,794,443, E: 280,981). The study area is characterized by an average annual temperature of 13 °C and an average annual precipitation of 1,115 mm, of which 345 mm falls during the growing season (May–September). The climate is oceanic with high annual precipitation and, although summer precipitation is relatively low in some areas, physiological drought does not occur in any part of the region, which is located entirely within the European Biogeographic Atlantic Region (EEA 2011).

The clay loam substrate (with a high presence of coarse elements, approximately 30%) was dumped and ameliorated in 2003. Soil formation is at an early stage and the soil structure is still unstable. The steep slopes of the terrain minimize groundwater effects. The physiography of the plots was characterized by a mean slope of 19% and an elevation ranging from 508 to 597 m above sea level.

#### 2.1.2 Experimental design

In the winter of 2008, the surface was subsoiled, plowed to a depth of 30–40 cm, and harrowed before the willow cuttings were planted. Three commercially available willow clones

**Fig. 1** Distribution of the sampling plots used for estimation of structural attributes and biomass yield. The photographic insert shows the experimental layout of the three commercial willow clones under study (green, *Bjor* clone; red, *Inger* clone; and blue, *Olof* clone)



were chosen for the study because of their adaptability to extreme soil conditions (e.g., nutrient poor and polluted soils) (Kuzovkina and Quigley 2005) and because they display good structural attributes and yield capacities for biomass production in SRC (Keoleian and Volk 2005). The cuttings were planted according to a double row planting design, leaving a distance of 0.75 m between each set of double rows, a distance of 1.5 m to the next set of double rows, and a distance between plants of 0.9 m (10,000 plants ha<sup>-1</sup>) or 0.6 m (15,000 plants ha<sup>-1</sup>) to provide two stocking levels (Fig. 2).

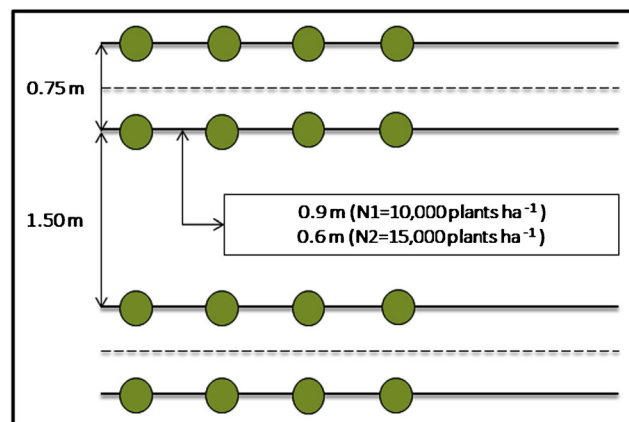
The experiment was established following a randomized complete block design (three blocks), in which three qualitative factors were considered for analysis: clone (three levels), planting density (two levels), and fertilization treatment (three levels), as outlined in Table 1.

Finally, the basic design was repeated in three blocks, with a total of 54 square plots each with an area of 400 m<sup>2</sup> (20 × 20 m) and constituted by 9 double rows with 22 or 33 cuttings per row (depending on the stocking density). Irrigation and pest/disease control were not performed during the cultivation period throughout the study area.

## 2.2 Field data collection

The experimental plots were measured in the autumn of 2012 according to the protocol described by the UK Forest Research (Forest Research 2003) for collecting data in short-rotation willow plantations (first rotation, stand age 5 years). Several variables were measured after the

vegetative period with the aim of assessing the performance of each clone. Measurements were made in rectangular subplots of 27 m<sup>2</sup> (9 × 3 m for density N1) and 18 m<sup>2</sup> (6 × 3 m for density N2) located in the center of each plot, to avoid the edge effect. A total of 40 stools (live or dead) were measured in each of the 54 subplots in this study. Within each of the subplots, shoot diameters were measured 20 and 130 cm aboveground level (D<sub>20</sub> and D<sub>130</sub>) with a digital caliper, and the total mean heights (H) were measured with a Vertex III hypsometer. The survival rate was also recorded at the end of each vegetative season. Before the trial, crops were harvested (in autumn 2012; stand age 5 years) and 5 stools were randomly selected in each of the abovementioned subplots and subsequently cut.



**Fig. 2** Diagram of the planting designs used in the trial

**Table 1** Main characteristics of the experimental design

Characteristic	Trial		
Total area	2.3 ha		
Experimental design	Randomized complete blocks		
Number of replicates	3		
Species	<i>Salix</i> spp.		
Clones tested	<i>Bjor</i> (B), <i>Inger</i> (I) and <i>Olof</i> (O)		
Origin of clones	Sweden		
Progenitor	B: <i>Salix schwerinii</i> × <i>Salix viminalis</i> I: <i>Salix trianta</i> × <i>Salix viminalis</i> O: <i>Salix viminalis</i> × ( <i>Salix schwerinii</i> × <i>Salix viminalis</i> )		
Plant density/spacing between cuttings	N1 = 10,000/0.9 N2 = 15,000/0.6 (plants·ha <sup>-1</sup> /m)		
Treatment	F0	F1	F2
Fertilization	None	300 kg ha <sup>-1</sup> N-P-K 6:20:12	600 kg ha <sup>-1</sup> N-P-K 6:20:12
Herbicide	None	Application of glyphosate (4 l/ha)	Application of glyphosate (4 l/ha)

A total of 270 stools were harvested manually with pruning shears (Electrocoup F3010) or a chainsaw. The fresh weight (FW) of each stool was measured with an electronic balance (precision ± 10 g). A representative subsample (300–500 g) of each stool was taken to the laboratory and weighed immediately with a precision balance (precision ± 1 g). The subsample was subsequently dried to constant weight at 70 °C: the dry weight (DW) was recorded and the dry biomass of the plot was calculated by multiplying the FW by the ratio of dry to fresh weight of the subsample.

The mean values and standard deviations for these variables are shown in Table 2. The biomass and structural attributes differed depending on the treatment applied (F0, F1, and F2) (Table 1), thus explaining the high standard deviations. The trial involved a dense plantation (10,000 and 15,000 plants ha<sup>-1</sup>) with a very homogeneous stand structure in each plot and scarcely any companion vegetation. The plot location data (four vertices) were collected with GPS submeter precision (*Trimble Geoexplorer 2008 series*).

## 2.3 LiDAR procedure

### 2.3.1 LiDAR data

The LiDAR data (Table 3) were acquired in July 2012 with ALS60 (Leica) and LMS-Q680 (Riegl) sensors. The beam divergence was 0.3 mrad and the pulsing frequency, 45 kHz; the scan frequency was 70 Hz, and the maximum scan angle, 50°. The first and last return pulses were recorded. Flights were conducted across the whole study area, and one flight was conducted for each strip, yielding an average measurement density of about 0.5 pulses m<sup>-2</sup>. The LiDAR data provided by the PNOA (official web page, 2016) included information about return type (first and last); X, Y, and Z coordinates; and intensity of the returned pulse by the sensor.

### 2.3.2 Extraction of LiDAR variables

For the low-density data acquired (0.5 pulses m<sup>-2</sup>), FUSION software (McGaughey 2010) was used to filter, interpolate,

**Table 2** Mean values and standard deviations of the test variables (H, D<sub>20</sub>, D<sub>130</sub>, FW, and DW) for the three clones (data corresponding to a field inventory carried out in 2012)

Clone	Treatment	H (m)	D <sub>20</sub> (cm)	D <sub>130</sub> (cm)	FW (t ha <sup>-1</sup> )	DW (t ha <sup>-1</sup> )
<i>Bjor</i>	F0	1.4 ± 0.4	1.1 ± 0.3	0.6 ± 0.2	1.8 ± 0.6	0.76 ± 0.4
	F1	2.4 ± 1.2	1.8 ± 0.9	1.2 ± 0.5	8.1 ± 6.6	4.45 ± 3.7
	F2	2.9 ± 0.5	2.2 ± 0.5	1.4 ± 0.4	10.9 ± 3.5	5.88 ± 2.1
<i>Inger</i>	F0	1.7 ± 0.7	1.4 ± 0.6	1.0 ± 0.5	4.9 ± 3.7	2.54 ± 1.9
	F1	2.7 ± 0.8	1.9 ± 0.3	1.2 ± 0.3	17.1 ± 7.7	9.32 ± 4.4
	F2	2.7 ± 0.7	2.0 ± 0.4	1.3 ± 0.3	14.7 ± 8.3	8.23 ± 5.0
<i>Olof</i>	F0	2.7 ± 0.7	1.7 ± 0.3	1.3 ± 0.2	6.4 ± 4.5	3.60 ± 2.6
	F1	5.7 ± 0.5	3.2 ± 0.3	2.4 ± 0.2	34.8 ± 11.9	19.71 ± 6.7
	F2	5.5 ± 1.4	3.0 ± 0.7	2.3 ± 0.6	33.2 ± 10.8	18.88 ± 6.9

**Table 3** Specifications of the LiDAR flight. Source: PNOA

Parameter	Specification
Pixel dimension	0.25 m
Point density	0.5 pulses m <sup>-2</sup>
Geodetic reference system	ETRS89
Map projection	UTM
Pulse rate	45 kHz
Areas covered	1:5000
Sensors	ALS60 (Leica); LMS-Q680 (Riegl)

and generate the digital terrain model (DTM)/digital crown model (DCM) and also to compute the following variables related to the metrics of heights and return intensity distributions within the limits of the 54 field plots: *mean, maximum and minimum values, mode, standard deviation, variance, interquartile distance, coefficients of skewness and kurtosis, average absolute deviation, and percentiles*. The proportion of returns (%) above a specific height threshold was also estimated in order to separate tree crop canopy returns from other vegetation.

The following steps were carried out with several processing programs (algorithms) implemented in the Fusion LiDAR Toolkit (McGaughey 2010).

- **GROUNDFILTER.** Ground returns were extracted from the LiDAR point cloud by using the GroundFilter tool, which implements a filtering algorithm adapted from Kraus and Pfeifer (1998) based on linear prediction (Kraus and Mikhail 1972).
- **GRIDSURFACE CREATE.** These returns were used to generate a DTM (1 m<sup>2</sup> resolution) grid with the GridSurfaceCreate tool, which computes the elevation of each grid cell by using the average elevation of all points within the cell; if the cell does not contain ground return points, its elevation is generated by interpolation from the neighboring cells. The metrics were generated for the exact size of each 20 × 20 m plot.
- **CLIPDATA.** The normalized LiDAR point cloud was obtained by subtraction of the ellipsoidal height of the DTM from the Z coordinate of each LiDAR return with the ClipData tool; this tool was used also to exclude returns below a normalized height of 0.5 m, considered as not belonging to tree crowns (e.g., hits on shrubs, rocks and logs).
- **POLYCLIPDATA.** The normalized LiDAR point cloud was clipped with the limits of each field plot—which were stored as polygons in vector format—by using the PolyClipData tool; one independent file was created per plot.
- **CLOUDMETRICS.** The metrics of heights and return intensity distributions of the 54 clipped and normalized point clouds were computed with the CloudMetrics tool.

Table 4 shows the complete set of metrics and the corresponding abbreviations used in this study.

## 2.4 Statistical analysis

Several statistical methods can be used in remote sensing prediction studies. In this study, parametric and non-parametric fitting methods were used to predict the main structural attributes and biomass production variables from LiDAR data and field measurements. Both fitting methods were used for comprehensive analysis of the data and because they provide important complementary information.

**Table 4** Statistics extracted from the heights and intensities of LiDAR flights and used as independent variables for the regression models

Description	Abbreviation
All returns above 1.00	All returns above 1.0
All returns above mode	All returns above mode
Canopy relief ratio	Canopy relief ratio
Elevation kurtosis	Elev kurt
Elevation L1	Elev L1
Elevation minimum	Elev minimum
Elevation mean	Elev mean
Elevation stddev	Elev stddev
Elevation 30th percentile	Elev P30
Elevation 40th percentile	Elev P40
Elevation 50th percentile	Elev P50
Elevation 60th percentile	Elev P60
Elevation 70th percentile	Elev P70
Elevation 75th percentile	Elev P75
Elevation skewness	Elev skewness
Elevation variance	Elev variance
Int sttdev	Int sttdev
Intensity kurtosis	Int kurt
Intensity L4	Int L4
Intensity of 5th percentile	Int P05
Intensity of 10th percentile	Int P10
Intensity of 25th percentile	Int P25
Intensity of 30th percentile	Int P30
First return above mode	First return above mode
Percentage all returns above 1.00	% all returns above 1.0
Percentage all returns above mean	% all returns above mean
Percentage first returns above 1.00	% first returns above 1.0
Return 1 count above 0.50	Return 1 count above 0.50
Total return count	Total return count
Total first returns	Total first returns
Total returns count above 0.50	Total returns count above 0.50

### 2.4.1 Parametric methods

Multiple linear regression (MLR) was used to model the relationships between field measurements (H,  $D_{20}$ ,  $D_{130}$ , FW, and DW) and the LiDAR variables in order to produce general models (for all clones together) and models classified by clone (Bjor, Inger, and Olof).

Candidate predictor variables were required to have an entering F-statistic with a significance level of 0.05 or less for inclusion in the model, and no predictor was left in the model with a partial F-statistic with a significance level greater than 0.05. Dependent variables derived from field data and predicted in regressions were mean height (H), basal diameter ( $D_{20}$ ), diameter at breast height ( $D_{130}$ ), fresh weight (FW), and dry weight (DW).

Comparison of the model estimates was based on the following two statistics: the adjusted coefficient of determination ( $R_{\text{fit}}^2$ ) and the root mean square error (RMSE). The  $R_{\text{fit}}^2$  compares the descriptive power of regression models that include the diverse numbers of predictors, and RMSE is a quadratic scoring rule that measures the average magnitude of the error (the square root of the average of squared differences between predicted and actual observations), which was calculated to provide additional information. Finally, residual plots were checked in order to validate the model fit. The variance inflation factor (VIF) was also taken into account. This factor quantified the severity of multicollinearity in ordinary least squares regression analysis and also provided an index that measured the extent to which the variation in an estimated regression coefficient increased due to collinearity. Only models in which all parameters were significant at the 5% level and with a VIF < 5 were included, thus ensuring that predictions were not highly correlated (Belsley et al. 2005; Mandeville 2008).

In this study, all of the data were used to construct the general and clone-specific models, as according to Myers (1990, p. 170) and Hirsch (1991), final estimation of model parameters using the entire dataset is more precise than when a model is fitted using only one portion of the data, especially with relatively small sample numbers.

### 2.4.2 Non-parametric method

In a preliminary analysis carried out to determine the most appropriate non-parametric statistical method, random forest (RF) and classification and regression trees (CART) were compared. The CART method was chosen because it provided better fits to the data, with larger  $R^2$  values and lower RMSE, than yielded by RF (see Online Resource 4). It is also a good exploratory technique that aims to determine classification and prediction rules.

The main advantages of the CART method are as follows (Gordon 2013; Timofeev 2004): (i) it does not require

specification of any functional form, (ii) it does not require variables to be selected in advance, (iii) it can easily handle outliers, (iv) it does not require the assumptions of statistical models and is computationally fast, (v) it is flexible and can deal with missing data, (vi) it works better than RF with relatively small databases, and (vii) it is easy to interpret (unlike random forest).

The objective of CART is usually to classify a dataset into several groups by use of a rule that displays the groups in the form of a binary tree (Breiman et al. 1984), which is determined by a procedure known as recursive partitioning. In this study, the CART method was used to classify the variables considered (H,  $D_{20}$ ,  $D_{130}$ , FW, and DW) in relation to the LiDAR data available.

Each tree branch is described by the value of one descriptor, chosen so that all objects in a daughter group have more similar response variable values. The split for continuous variables is defined by  $x_i < a_j$ , where  $x_i$  is the selected descriptor or explanatory variable, and  $a_j$  is its split value. To choose the most appropriate descriptor  $x_i$  and value of  $a_j$ , CART uses an algorithm in which all descriptors and all split values are considered, selecting those giving the best reduction in impurity between the mother group ( $t_p$ ) and the daughter groups ( $t_L$  and  $t_R$ ) (Deconinck et al. 2005). This process is repeated for each daughter group until the maximal tree height is reached. Mathematically, this is expressed as follows:

$$\Delta i(s, t_p) = i(t_p) - p_L i(t_L) - p_R i(t_R) \quad (1)$$

where  $i_t$  is the impurity,  $s$  the candidate split value, and  $p_L$  and  $p_R$  are the fractions of the objects in respectively the left and right daughter groups.

The impurity is defined as the total sum of squares of the deviations of the individual responses from the mean response of the group and is expressed as follows:

$$i(t) = \sum_n \left( y_n - \bar{y}(t) \right)^2 \quad (2)$$

where  $i(t)$  is the impurity of group  $t$ ,  $y_n$  is the value of the response variable for object  $x_n$ , and  $\bar{y}(t)$  is the mean of the response variable in group  $t$ .

CART methods are not required to conform to probability distribution restrictions, and there is no assumption of linearity or any need to pre-specify a probability distribution for the errors (Bell 1999).

Complexity and robustness are competing characteristics that must be considered simultaneously during construction of statistical models. The more complex a model is, the less reliable it will be for purposes of prediction. To prevent this from occurring, stopping rules must be applied during elaboration and the development of a decision tree to prevent the model from becoming overly complex. Common parameters

used in stopping rules include (a) the minimum number of observations in a leaf, (b) the minimum number of observations in a node prior to splitting, and (c) the depth (i.e., number of levels) of any leaf from the root node (Song and Lu 2015). On this occasion, no pruning was necessary because a maximum of 2 levels was considered for tree depth (or a maximum of 6 nodes). The risk estimate, which is a measure of the within-node variance, was used as an indicator of model performance (IBM Corp. Released 2015).

### 3 Results

We used two well-defined and complementary parametric and non-parametric fitting procedures to analyze and estimate the best response. In this case, the structural attributes (H,  $D_{20}$ , and  $D_{130}$ ) and biomass yield variables (FW and DW) were strongly and positively correlated ( $R^2 > 0.87$ , additional data are given in Online Resource 5).

The main results obtained with both fitting methods are shown in Tables 5, 6, and 7. All the dependent variables (H,  $D_{20}$ ,  $D_{130}$ , FW, and DW) were analyzed and interpreted at trial level and separately for each clone (see Table 7 for data on the *Olof* clone and additional data are given in Online Resources 1 and 2 for data on the *Bjor* and *Inger* clones). The parameter estimates and goodness-of-fit statistics for the best parametric model (multiple linear regression) are summarized in Table 5. Finally, the best non-parametric (CART) models are included in Table 6 (trial level) and Table 7 (*Olof* clone). The scatter plots generated by MLR and CART (Online Resource 3) show the generally close relationship between the predicted values and the field measured values.

#### 3.1 Parametric methods

At trial level, the models for the structural attributes (H,  $D_{20}$ , and  $D_{130}$ ) provided good fits, with more than 76% of the variance explained, and the LiDAR variables with greatest influence were those related to *elevation percentile*, *total returns*, and *the first returns on 0.5*. The highest  $R^2$  value was obtained for the mean height variable ( $R_{\text{fit}}^2 = 89\%$ , Table 5). When modeling the biomass variables (FW and DW), the  $R_{\text{fit}}^2$  values were higher than 75% (Table 5), and in both cases, the most influential variables in the model were *elevation percentile* and *percentage of all returns above mean*.

Regarding the fit for the models for each clone, the *Bjor* clone model provided the best results with H and  $D_{20}$  (91 and 97%, respectively; Table 5). However, the *Olof* clone model produced the highest  $R_{\text{fit}}^2$  using  $D_{130}$  (95%; Table 5). In the case of the biomass variables (FW and DW), the amount of variance explained was sometimes lower, with  $R_{\text{fit}}^2$  values above 66%. The *Bjor* clone model produced the highest  $R_{\text{fit}}^2$  values (90%) for both the FW and DW variables (Table 5).

The homogeneity of variance was evaluated using SPSS graphs obtained with ZRESID and ZPRED (standardized predicted values and standardized residuals) commands. Plots of residuals against predicted values showed no evidence of heterogeneous variance and no systematic pattern. The results indicated the absence of atypical or scattered sample data and, furthermore, bell-shaped histograms indicated that all datasets were normally distributed.

#### 3.2 Non-parametric methods

The proportion of variance explained by the mean height variable-dependent model for the whole trial produced an  $R^2$  value of 84.6%. In this case, the most important LiDAR variable was the mean elevation. It is important to highlight the influence of this variable on the other structural attributes and production variables, apart from  $D_{20}$ , with the most significant independent variable related to the elevation of the percentiles, namely the variable *Elev. P60* (Table 6).

The independent variables that best define the decision trees (CART) per clone were more varied. The fitting provided good results for all three clones studied. For the *Olof* clone, which yielded the highest values for structural attributes and biomass production in the field, the fit was good, with  $R^2 > 92\%$  for all study variables (Table 7). However, in plots with lower values for mean height, diameter, and biomass, i.e., plots with the *Bjor* and *Inger* clones, the independent explanatory variables were those related to different percentiles of elevation and % returns (Online Resources 1 and 2). In the CART analysis for each clone, we observed that, with some exceptions, most of the models produced a second level of classification, with variables related to elevation and % returns.

Once the fitting was completed, we were able to verify that the LiDAR variables associated with mean elevation and elevation percentile (tree height) were the most important in the fitting process, as these were included in all models produced by the parametric and non-parametric fitting methods used.

### 4 Discussion

The parametric and non-parametric model fitting conducted in this study has revealed acceptable results that indicate the usefulness of LiDAR data for estimating structural attributes and biomass production in forest energy crops grown on degraded mining land.

The data collected in the field inventory and LiDAR data (available for free) captured in the same period were highly correlated. One of the objectives of the study was to determine differences in the performance of the different models studied (H,  $D_{20}$ ,  $D_{130}$ , FW, and DW) at trial level (without differentiating clones) and separately for each clone.



**Table 5** Results of multiple regression showing the best models obtained for H, D<sub>20</sub>, D<sub>130</sub>, FW, and DW, for each trial and each clone

Dependent variable	Factor	Model type	Independent variable	Parameter estimate	Std. Error	RMSE	R <sup>2</sup>	R <sub>fit</sub> <sup>2</sup> (%)
H (m)	Trial	Multiple	(Constant)	3.496	0.771	0.54	0.90	89.00
			Elev P50	0.810	0.076			
			Total return count	- 0.009	0.002			
	Bjor clone	Multiple	Return 1 count above 0.50	0.006	0.002	0.29	0.93	91.00 <sup>a</sup>
			(Constant)	1.116	0.176			
			% all returns above 1.00	0.324	0.035			
	Inger clone	Multiple	Return 1 count above 0.50	- 0.059	0.009	0.51	0.70	68.00
			Elev skewness	- 1.159	0.207			
			Elev variance	0.712	2.222			
	Olof clone	Multiple	(Constant)	1.360	0.210	0.60	0.88	87.00
			% first returns above 1.00	0.041	0.07			
			(Constant)	- 9.161	3.847			
D <sub>20</sub> (cm)	Trial	Multiple	Elev P50	1.234	0.123	0.40	0.76	75.00
			Int P05	0.900	0.350			
			(Constant)	2.262	0.567			
	Bjor clone	Multiple	Total return count	0.336	0.056	0.12	0.98	97.00 <sup>a</sup>
			Return 1 count above 0.50	- 0.004	0.001			
			(Constant)	0.003	0.001			
	Inger clone	Multiple	Elev P25	4.569	0.343	0.30	0.65	63.00
			Total return count	0.574	0.097			
			Elev P01	- 0.011	0.001			
	Olof clone	Multiple	Canopy relief ratio	- 1.617	0.266	0.24	0.93	91.00
			First returns above 1.00	3.111	0.439			
			Lnt mode	0.010	0.002			
	Trial	Multiple	(Constant)	- 0.046	0.019	0.29	0.82	81.00
			Elev P30	0.247	0.290			
			Return 1 count above 0.50	0.856	0.158			
	Bjor clone	Multiple	Total return count	- 3.436	1.253	0.17	0.91	89.00
			Total first returns	0.521	0.168			
			Elev P50	- 0.006	0.001			
	Inger clone	Multiple	Canopy relief ratio	1.007	0.126	0.19	0.81	77.00
			(Constant)	1.191	0.540			
			Elev mean	1.191	0.540			
	Olof clone	Multiple	Total first returns	2.052	0.561	0.14	0.96	95.00 <sup>a</sup>
			Int skewness	0.449	0.131			
			(Constant)	- 0.006	0.001			
FW (t ha <sup>-1</sup> )	Trial	Multiple	Int P25	0.215	0.094	6.64	0.76	75.00
			(Constant)	- 33.601	12.788			
			Elev P50	1.418	0.329			
	Bjor clone	Multiple	Int P10	0.579	0.104	1.84	0.92	90.00 <sup>a</sup>
			Int P30	- 0.258	0.069			
			Elev P40	- 1.027	0.351			
	Inger clone	Multiple	Int P10	3.077	1.319	4.03	0.79	77.00
			(Constant)	- 4.894	1.954			
			Elev P60%	4.682	0.859			
	Trial	Multiple	all returns above mean	0.596	0.122	6.64	0.76	75.00
			(Constant)	2.777	0.819			
			% all returns above 1.00	2.326	0.366			
Bjor clone	Multiple	First returns above 1.00	- 0.522	0.111	1.84	0.92	90.00 <sup>a</sup>	
		Elev skewness	- 3.992	0.920				
		Elev L4	9.128	3.725				
Inger clone	Multiple	(Constant)	- 1.355	2.170	4.03	0.79	77.00	
		(All returns above mode) / (Total first returns) × 100	0.489	0.066				
		Int kurt	1.092	0.374				

**Table 5** (continued)

Dependent variable	Factor	Model type	Independent variable	Parameter estimate	Std. Error	RMSE	$R^2$	$R_{\text{fit}}^2$ (%)
DW (t ha <sup>-1</sup> )	Olof clone	Multiple	(Constant)	- 67.239	12.760	6.35	0.88	85.00
			Elev P75	100.216	25.310			
			Int P90	1.827	0.496			
			Elev P70	- 89.366	25.076			
			Int L4	- 22.947	8.855			
	Trial	Multiple	(Constant)	- 2.921	1.043	3.67	0.78	77.00
			Elev. P50%	2.817	0.493			
			all returns above mean	0.355	0.067			
	Bjor clone	Multiple	(Constant)	0.210	0.615	1.05	0.92	90.00 <sup>a</sup>
			% all returns above 1.0	1.248	0.217			
			First returns above 1.0	- 0.289	0.064			
			Elev skewness	- 2.615	0.597			
	Inger clone	Multiple	(Constant)	- 1.184	1.190	2.21	0.81	79.00
			(All returns above mode) / (Total first returns) × 100	0.283	0.036			
			Int kurt	0.633	0.205			
Olof clone	Multiple	(Constant)	- 8.093	3.975	5.42	0.68	66.00	
		Elev P75	4.917	0.835				

<sup>a</sup> Best fit at the trial level and per clone for each study variable. The RMSE units are the same as the variable units

The results indicate that free LiDAR data can be used to estimate the variables with acceptable precision ( $R_{\text{fit}}^2 > 63\%$  in multiple regression and  $R^2 > 73.5\%$  for CART analysis) and that satisfactory results were obtained, despite the low density of LiDAR points. At the trial level, although MLR produced higher  $R^2$  values than CART, except for the  $D_{20}$  and FW fits, the values were generally very similar. However, when the fits were carried out taking into account the clone level (i.e., a more uniform sample), the CART method produced larger  $R^2$  values (except for  $D_{130}$  in *Inger* and *Olof* and H and  $D_{20}$  in *Bjor*). Finally, CART and MLR produced very similar RMSE values at trial level, but CART generally yielded smaller RSME values at the clone level (except for *Bjor* and *Olof* clones with  $D_{20}$  and  $D_{130}$ , respectively). The goodness-of-fit levels yielded by the models are comparable to those obtained in other studies using LiDAR to characterize forest attributes (Dalponte et al. 2011; García-Gutierrez et al. 2014; Seidel and Ammer 2014).

The MLR and CART scatter plots (additional data are given in Online Resource 3) show that both methods provided acceptably good fits to the data. However, for the structural variables (H,  $D_{20}$ ,  $D_{130}$ ), the models developed produced better predictions than for biomass variables (FW and DW), in that the data were more widely dispersed.

On the other hand, to verify and compare the models, for both parametric and non-parametric methods, a newly collected dataset (additional inventory) should be used for validation (Hirsch 1991; Kozak and Kozak 2003; Myers 1990) because the only universally acceptable method for validating a model and assessing its goodness of fit after model selection is to use an independent sample (Lever et al. 2016). However, independent validation was not possible in this study, as application of a validation method requires more data and the CART method

thus becomes more unstable (Gordon 2013), especially as regards the models developed for each clone. Nonetheless, the models fitted in this study can be considered sufficiently robust for estimating structural attributes and productivity of SRWC. In this case, the CART analysis of the relationships between field-measured variables and LiDAR data produced reasonable results. Moreover, the errors were acceptable, considering the high degree of variability between the trial plots.

The trial level models produced by the MLR method show that the most important variables, for trial and clone, were those related to elevation and returns, as also shown by the CART method. However, the models generated by MLR for each clone were defined by more diverse LiDAR variables, which also included (apart from those already mentioned) variables related to intensity (see Table 5).

Good fits were obtained for the structural attributes studied, mean height (H) and both diameters ( $D_{20}$  and  $D_{130}$ ), probably because these are closely related to the LiDAR-derived elevation variables (tree height). However, good-fit models were also obtained for FW and particularly DW biomass, as expected, because these variables are strongly correlated with structural variables (see Online Resource 5). Some variables such as mean elevation (*Elev mean*) and elevation percentile (*Elev %*) were included in most of the models. The data for the trial plots planted with the *Olof* clone were uniformly distributed, and the independent variable that provided the best predictions was the mean elevation (*Elev mean*).

The mean height variable was closely correlated with the LiDAR variables associated with the height of the trees (i.e., elevation) and with the returns, as also observed by Montealegre Gracia et al. (2015) in a study of a *Pinus halapensis* Mill plantation in Spain. Another study in plantations of

**Table 6** Results of the non-parametric fitting (CART) of the models obtained for H, D<sub>20</sub>, D<sub>130</sub>, FW, and DW, at the trial level

Variable	R <sup>2</sup> (%)	RMSE	Node	Mean	Standard deviation	Number	Percent	Predicted mean	Parental Node	Independent variable		Value	
										Variable	Improvement		
H (m)	84.6 <sup>a</sup>	0.64	0	3.14	1.65	53	100	3.14	-	-	-	-	
			1	2.41	0.90	42	79.2	2.41	0	Elev mean	1.98	≤ 3.64	
			2	5.89	0.62	11	20.8	5.89	-	-	-	> 3.64	
			3	1.65	0.65	16	30.2	1.65	1	30.2	% first returns above 1.00	0.28	≤ 10.09
D <sub>20</sub> (cm)	76.4	0.38	4	2.88	0.69	26	49.1	2.88	-	-	-	> 10.09	
			0	2.04	0.80	53	100	2.04	-	-	-	-	
			1	1.63	0.49	37	69.8	1.63	0	1.63	Elev P60	0.38	≤ 2.58
			2	2.98	0.52	16	30.2	2.98	-	-	-	-	> 2.58
D <sub>130</sub> (cm)	79.1	0.30	3	1.04	0.29	10	18.9	1.04	1	18.9	% first returns above 1.00	0.09	≤ 5.83
			4	1.85	0.34	27	50.9	1.8	-	-	-	> 5.83	
			0	1.44	0.66	53	100	1.44	-	-	-	-	
			1	1.16	0.42	42	79.2	1.16	0	1.16	Elev mean	0.28	≤ 3.64
FW (t ha <sup>-1</sup> )	78.8	6.47	2	2.48	0.25	11	20.8	2.48	-	-	-	> 3.64	
			3	0.56	0.20	7	13.2	0.56	1	0.56	Total returns above 0.50	0.06	≤ 15.00
			4	1.28	0.34	35	66.0	1.28	-	-	-	> 15.00	
			0	14.90	13.22	53	100	14.94	-	-	-	-	
DW (t ha <sup>-1</sup> )	77.2	3.64	1	9.52	7.32	42	79.2	9.52	0	79.2	Elev mean	110.56	
			2	35.45	10.17	11	20.8	35.45	-	-	-	-	
			3	0.56	0.20	7	13.2	0.56	1	0.56	First returns above 1.00	19.10	≤ 38.00
			4	1.28	0.34	35	66.0	1.28	-	-	-	-	> 38.00
	77.2	3.64	0	8.15	7.63	54	100	8.15	-	-	-	-	
			1	5.05	4.14	43	79.6	5.05	0	5.05	Elev mean	37.54	≤ 3.64
			2	20.26	5.86	11	20.4	20.26	-	-	-	-	> 3.64
			3	3.69	2.89	35	64.8	3.69	1	3.69	% first returns above 1.00	6.43	≤ 25.80
4	11.00	3.53	8	14.8	11.0	-	-	-	-	> 25.80			

<sup>a</sup> Best fit at the trial level for each study variable. The RMSE units are the same as the variable units

**Table 7** Results of non-parametric fitting of the models obtained for H, D<sub>20</sub>, D<sub>130</sub>, FW, and DW, for the *Olof* clone

Variable	R <sup>2</sup> (%)	RMSE	Node	Mean	Standard deviation	Number	Percent	Predicted mean	Parental Node	Independent variable	
										Variable	Value
H (m)	94.2	0.40	0	4.68	1.68	18	100.0	4.68	-	-	-
			1	2.77	0.69	7	38.9	2.77	Elev mean	≤ 3.64	2.30
			2	5.89	0.62	11	61.1	5.89		> 3.64	
			3	6.5	0.34	3	16.7	6.57	Int stdev	≤ 2.52	0.10
			4	5.64	0.51	8	44.4	5.64		> 2.52	
			0	2.62	0.81	18	100	2.62		-	-
D <sub>20</sub> (cm)	93.6	0.20	1	1.70	0.29	7	38.9	1.70	Elev mean	≤ 3.64	0.53
			2	3.20	0.31	11	61.1	3.20		> 3.64	
			3	1.50	0.17	4	22.2	1.50	% first above mode	≤ 12.83	0.02
			4	1.96	0.20	3	16.7	1.96		> 12.83	
			5	3.61	0.24	2	11.1	3.61	% first above 1	≤ 32.11	0.02
			6	3.11	0.25	9	50.0	3.11		> 32.11	
D <sub>130</sub> (cm)	94.8 <sup>a</sup>	0.15	0	2.01	0.65	18	100.0	2.01	-	-	-
			1	1.26	0.19	7	38.9	1.26	Elev mean	≤ 3.64	0.35
			2	2.48	0.25	11	61.1	2.48		> 3.64	
			3	1.09	0.08	3	16.7	1.09	Total return count	≤ 47.50	0.00
			4	1.40	0.13	4	22.2	1.40		> 47.50	
			5	2.85	0.19	2	11.1	2.85	% first returns above 1.0	≤ 32.11	0.01
FW (t ha <sup>-1</sup> )	92.3	4.48	0	24.81	16.20	18	100.0	24.81	-	-	-
			1	8.09	6.12	7	38.9	9.09	% all return above mean	≤ 3.64	177.89
			2	35.45	10.17	11	61.1	35.45		> 3.64	
			3	4.77	2.46	5	27.8	4.44	% first return above mode	≤ 49.00	10.74
			4	16.40	2.76	2	11.1	16.40		> 49.00	
			5	52.43	4.01	2	11.1	51.43	Elev minimum	≤ 39.16	39.16
DW (t ha <sup>-1</sup> )	92.2	2.60	0	14.06	9.36	18	100	14.06	-	-	-
			1	4.31	3.07	7	38.9	4.31	Elev mean	≤ 3.64	60.49
			2	20.2	5.86	11	61.1	20.26		> 3.64	
			3	2.65	1.48	5	27.8	2.65	First return above mode	≤ 49.00	2.65
			4	8.44	0.22	2	11.1	8.44		> 49.00	
			5	29.99	1.19	2	11.1	29.99	Elev minimum	≤ 0.53	12.85
6	18.10	3.37	9	50.0	18.10		> 0.53				

<sup>a</sup> Best fit per *Olof* clone for each study variable. The RMSE units are the same as the variable units

*Pinus radiata* D.Don in northern Spain (González-Ferreiro et al. 2012) indicated that mean height can be accurately modeled from low-density laser data (0.5 pulses  $m^{-2}$ ), yielding an  $R^2$  value of 0.75 (in comparison with  $R^2$  values of 0.70–0.93 obtained by parametric methods and  $R^2$  values of 0.85–0.94 by non-parametric (CART) methods in the present study). In a study of an olive plantation in Spain, with low-density data (0.5 pulses  $m^{-2}$ ), good results were also obtained for height ( $R^2 = 0.67$ ) (Estornell et al. 2014).

The results obtained for the models of the diameter variables ( $D_{20}$  and  $D_{130}$ ) were better than expected, taking into account the reports in the relevant scientific literature. Thus, the models used to estimate diameters performed similarly to those reported by Gonçalves-Seco et al. (2011) for *Eucalyptus globulus* Labill plantations in Spain ( $R^2 = 0.71$ ) and by Dalponte et al. (2011) for a mixed stand ( $R^2 = 0.63$ ). Similar findings were also reported by Graham (2008) for the  $D_{130}$  model and a pine plantation ( $R^2 = 0.82$ ). However, in the same study, a much lower  $R^2$  value was obtained for a natural pine stand ( $R^2 = 0.20$ ). In a study conducted by Estornell et al. (2014) in a plantation of small trees (olives) in the Mediterranean area of Spain, with low-density LiDAR data (0.5 pulses  $m^{-2}$ ), the findings indicated good model performance ( $R^2 = 0.70$ ) for estimating volume, which is directly related to diameter. The findings thus seem similar to those of the present study.

Several studies carried out worldwide with low-point-density data have reported similar results to those obtained in the present study. For example, a study carried out in Europe to evaluate the aboveground biomass in boreal forest zone with an average point density of between 0.7 and 1.2  $m^{-2}$  reported an  $R^2$  value of 0.88 (Næsset and Gobakken 2008). A study carried out in Canada using low-density data (0.5 pulses  $m^{-2}$ ) obtained an excellent fit for the biomass, with an overall  $R^2$  of 0.93 (Treitz et al. 2010). Li et al. (2008) also observed a significant relationship between field-based aboveground biomass estimates based on field and LiDAR measurements for the three study sites, located in the USA and Canada and in which different forest species were used. However, a study by Næsset (2011) in small areas of forest land in Norway where the main tree species were Norway spruce (*P. abies* (L.) Karst.) and Scots pine (*P. sylvestris* L.) showed that these species are subject to substantial inherent canopy height variation, leading to highly variable predictions for estimate aboveground biomass in young forests (Magnussen and Boudewyn 1998).

In the present study, no comparison was made for different LiDAR point densities because of the scarcity of the LiDAR data. Previous studies did not find any evidence indicating that a reduction in point density affects the model accuracy, and Treitz et al. (2010) considered that data captured at 0.5 pulses  $m^{-2}$  may be an excellent source of information for forest management. The same was also concluded by González-Ferreiro et al. (2013), who showed that low-intensity LiDAR data (0.5

pulses  $m^{-2}$ ) can be used without significant loss of information. This suggests that data captured at 0.5 pulses  $m^{-2}$  yields good estimates, without excessive loss of model quality.

According to the findings reported by García et al. (2010), the intensity variables are more strongly related to biomass than to mean height, as also shown in the present study. Likewise, González-Ferreiro et al. (2013) found that the independent variables associated with the return intensity and related canopy measurements can add some valuable information for predicting biomass in eucalyptus plantations in northern Spain ( $R^2 = 0.75$ ; 4 pulses  $m^{-2}$ ). LiDAR studies carried out in small areas have shown that the height percentiles are usually closely correlated with biomass (González-Ferreiro et al. 2012); in the present study, the elevation percentiles (tree height) were found to be the most important LiDAR variables to include in multiple linear regression models for estimating biomass (FW and DW) at trial level. The results of a study by Zhao et al. (2009) showed that the models can accurately predict forest biomass and that the predictive performance was consistent across a range of scales, with  $R^2$  ranging from 0.80 to 0.95 across all fitted models. However, in a study conducted by Van Aardt et al. (2006) in a coniferous plantation, the  $R^2$  values for volume (0.66) and aboveground biomass (0.59) were low, which was attributed to variability in the volume. Condés et al. (2013) noted that better fits can be achieved with multiple linear regression models by inclusion of a larger number of variables. This was also observed in the present study, and the  $R^2$  value increased as the number of independent variables that make up the model increased, e.g., in the multiple linear regression, the best result was obtained for  $D_{20}$  and the *Bjor* clone, with six variables included in the model ( $R^2 = 0.98$ ).

Different studies in conifer plantations worldwide also indicate that height can be accurately modeled from medium- and low-density laser data. In the abovementioned study on *P. radiata*, González-Ferreiro et al. (2012) indicated differences in the model fit for height for different data point densities (0.5 and 4 pulses  $m^{-2}$ ), although the difference in the goodness of fit was only 8%. Thus, although better fits are obtained by using higher point densities, the extra costs involved may not be justified by the final result.

In addition to the importance of reducing the cost of the inventory, the LiDAR information obtained is very useful for forest management purposes. Several studies indicate that the use of LiDAR data generates more accurate inventories than traditional inventory methods based on field measurements (Maltamo et al. 2004; Næsset 2002). The combined use of LiDAR technology and advanced statistical techniques has led to a number of different studies exploring their potential for producing accurate results in forest biomass research (Cho et al. 2012; Gleason and Im 2012; Lefsky et al. 2001, 2005; Weishampel et al. 1996; Wulder 1998). In this respect, selection of a suitable statistical approach is essential considering

that the models will be used for predictive purposes. Besides being easy to use and interpret, the non-parametric method (CART) includes classification rules and is thus a very useful tool for LiDAR-based inventories.

As mentioned earlier, the methods used in this study enable more accurate inventories to be carried out and at a fraction of the cost than possible by traditional methods (assuming the timely availability of suitable data) (Means and Acker 2000). The forest variables estimated from the LiDAR data in this study are of great interest to the timber industry and represent information that is expensive to collect in the field. The results indicate the good relationship between LiDAR data and the various forest variables considered. This is of great value in forest management, providing a tool to determine the best time for harvesting the energy crop, as well as for monitoring and managing plots (Lim et al. 2003).

The results obtained in this study, in terms of the models fitted, indicate that forest energy crops can be accurately modeled from low-intensity laser data and that the models are similar to those reported in the international scientific literature.

## 5 Conclusions

The study findings show that low density LiDAR data (0.5 pulses  $m^{-2}$ ) can be used to construct models to estimate the main variables (structural attributes and biomass yield) of interest in the management of short-rotation (willow) energy crops. According to the moderate to highly accurate estimates obtained, the models developed on the basis of LiDAR data can be used to produce good predictions and estimates for an SRC crop, and would serve as a management tool for improving and optimizing follow-up decisions related to a commercial crop. In view of the results obtained, acquisition or purchase of low-density LiDAR information to facilitate the monitoring of the energy plantations can be considered from a commercial point of view. The parametric and non-parametric statistical methods tested in the present study (MLR and CART) provided complementary and robust information for predicting stand variables in an SRC energy crop from low-density LiDAR data (0.5 pulses  $m^{-2}$ ) and therefore can be considered suitable for developing models for accurate estimation of forest variables. The predictive power of both methods was generally high (particularly when limited data were available for fitting), although MLR models are easier to interpret and apply.

**Acknowledgements** This work was funded by the Hunosa Group coal mining company. The authors acknowledge the helpful co-operation of staff from Hunosa Group in this study.

**Funding information** The research was supported by the Hunosa Chair at the University of Oviedo.

## References

- Bell JF (1999) Tree-based methods. The use of classification trees to predict species distributions. In: Fielding A (ed) Machine learning methods for ecological applications. Kluwer, Dordrecht
- Belsley DA, Kuh E, Welsch RE (2005) Regression diagnostics: identifying influential data and sources of collinearity. Wiley, New York
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth and Brooks/Cole, Monterey
- Broeckx LS, Verlinden MS, Ceulemans R (2012) Establishment and two-year growth of a bio-energy plantation with fast-growing *Populus* trees in Flanders (Belgium): effects of genotype and former land use. *Biomass Bioenergy* 42:151–163. <https://doi.org/10.1016/j.biombioe.2012.03.005>
- Cho MA, Mathieu R, Asner GP, Naidoo L, van Aardt J, Ramoelo A, Debba P, Wessels K, Main R, Smit IPJ, Erasmus B (2012) Mapping tree species composition in south African savannas using an integrated airborne spectral and LiDAR system. *Remote Sens Environ* 125:214–226. <https://doi.org/10.1016/j.rse.2012.07.010>
- Clark ML, Clark DB, Roberts DA (2004) Small-footprint LIDAR estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sens Environ* 91:68–89. <https://doi.org/10.1016/j.rse.2004.02.008>
- Condés S, Fernández-Landa A, Rodríguez F (2013) Influencia del inventario de campo en el error de muestreo obtenido en un inventario con tecnología Lidar. 6º Congreso Forestal Español (Sociedad Española de Ciencias Forestales). 6CFE01–432
- Coops NC, Hilker T, Wulder MA, St-Onge B, Newnham G, Siggins A, Trofymow JT (2007) Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees* 21:295. <https://doi.org/10.1007/s00468-006-0119-6>
- Dalponte M, Bruzzone L, Gianelle DA (2011) System for the estimation of single-tree stem diameter and volume using multireturn LIDAR data. *IEEE Trans Geosci Remote Sens* 49:2479–2490. <https://doi.org/10.1109/TGRS.2011.2107744>
- Deconinck E, Hancock T, Coomans D, Massart DJ, Vander Heyden Y (2005) Classification of drugs in absorption classes using the classification and regression trees (CART) methodology. *J Pharm Biomed Anal* 39:91–103. <https://doi.org/10.1016/j.jpba.2005.03.008>
- Edenhofer O, Pichs-Madruga R, Sokona Y, Seyboth K, Arvizu D, Bruckner T, Christensen J, Devernay J-M, Faaij A, Fischer M, Goldstein G, Hansen J, Huckerby A, Jäger-Waldau S, Kadner D, Kammen V, Krey A, Kumar A, Lewis O, Lucon P, Matschoss L, Maurice C, Mitchell W, Moomaw J, Moreira A, Nadai LJ, Nilsson J, Nyboer A, Rahman J, Sathaye J, Sawin R, Schaeffer T, Schei S, Schlömer R, Sims A, Verbruggen C, Von Stechow K, Urama R, Wiser F, Yamba T, Zwickel (2011) Summary for policy makers. In: Edenhofer O, Pichs-Madruga R, Sokona Y, Seyboth K, Matschoss P, Kadner S, Zwickel T, Eickemeier P, Hansen G, Schlöme S, Stechow CV (eds) IPCC special report on renewable energy sources and climate change mitigation. Cambridge University Press, Cambridge
- EEA (2011) Biogeographical regions. European Environment Agency, Copenhagen [online] URL: <http://www.eea.europa.eu/dataand-maps/data/biogeographical-regions-europe-1>
- Eid T, Gobakken T, Næsset E (2004) Comparing stand inventories for large areas based on photo-interpretation and laser scanning by means of cost-plus-loss analyses. *Scand J For Res* 19:512–523
- Estornell J, Ruiz LA, Velázquez-Martí B, Fernández-Sarría A (2011) Estimation of shrub biomass by airborne LiDAR data in small forest stands. *For Ecol Manag* 262:1697–1703. <https://doi.org/10.1016/j.foreco.2011.07.026>
- Estornell J, Velázquez-Martí B, López-Cortés I, Salazar D, Fernández-Sarría A (2014) Estimation of wood volume and height of olive tree

- plantations using airborne discrete-return LiDAR data. *GISci Remote Sens* 51:17–29. <https://doi.org/10.1080/15481603.2014.883209>
- Falkowski MJ, Smith AMS, Hudak AT, Gessler PE, Vierling LA, Crookston NL (2006) Automated estimation of individual conifer tree height and crown diameter via two-dimensional spatial wavelet analysis of LiDAR data. *Can J Remote Sens* 32:153–161. <https://doi.org/10.5589/m06-005>
- Forest Research (2003) Mensurational variables protocol. In: *Yield Models for Energy Coppice of Poplar and Willow*. 1–14
- García M, Riaño D, Chuvieco E, Danson FM (2010) Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sens Environ* 114:816–830. <https://doi.org/10.1016/j.rse.2009.11.021>
- García-Gutiérrez J, González-Ferreiro E, Riquelme-Santos JC, Miranda D, Dieguez-Aranda U, Navarro-Cerrillo RM (2014) Evolutionary feature selection to estimate forest stand variables using LiDAR. *Int J Appl Earth Obs Geoinf* 26:119–131. <https://doi.org/10.1016/j.jag.2013.06.005>
- Gleason CJ, Im J (2012) Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sens Environ* 125: 80–91. <https://doi.org/10.1016/j.rse.2012.07.006>
- Gonçalves-Seco L, González-Ferreiro E, Diéguez-Aranda U, Fraga-Bugallo B, Crecente R, Miranda D (2011) Assessing attributes of high density *Eucalyptus globulus* stands using airborne laser scanner data. *Int J Remote Sens* 32:9821–9841. <https://doi.org/10.1080/01431161.2011.593583>
- González-Ferreiro E, Diéguez-Aranda U, Miranda D (2012) Estimation of stand variables in *Pinus radiata* D. Don plantations using different LiDAR pulse densities. *Forestry* 85:281–292. <https://doi.org/10.1093/forestry/cps002>
- González-Ferreiro E, Miranda D, Barreiro-Fernández L, Buján S, García-Gutiérrez J, Dieguez-Aranda U (2013) Modelling stand biomass fractions in Galician *Eucalyptus globulus* plantations by use of different LiDAR pulse densities. *For Syst* 22:510–525. <https://doi.org/10.5424/fs/2013223-03878>
- Gordon L (2013) Using classification and regression trees (CART) in SAS for applications in public health. *SAS Glob Forum* 1–8
- Graham B (2008) Using low density, small footprint LIDAR in forest inventory applications in the Southeastern U.S
- Hall SA, Burke IC, Box DO, Kaufmann MR, Stoker JM (2005) Estimating stand structure using discrete-return Lidar: an example from low density, fire prone ponderosa pine forests. *For Ecol Manag* 208:189–209. <https://doi.org/10.1016/j.foreco.2004.12.001>
- Hawbaker TJ, Gobakken T, Lesak A, Trømborg E, Contrucci K, Radeloff V (2010) Light detection and ranging-based measures of mixed hardwood forest structure. *For Sci* 56:313–326
- Hayashi R, Weiskittel A, Sader S (2014) Assessing the feasibility of low-density LiDAR for stand inventory attribute predictions in complex and managed forests of northern Maine, USA. *Forests* 5:363–383. <https://doi.org/10.3390/f5020363>
- Hirsch RP (1991) Validation samples. *Biometrics* 47:1193–1194
- IBM Corp. Released (2015) IBM SPSS Statistics for Windows, Version 23.0. IBM Corp, Armonk
- IDAE (2007) Biomasa: cultivos energéticos. Instituto para la Diversificación y el Ahorro de la Energía (IDAE), 49 pp, Madrid
- Jama A, Nowak W (2012) Willow (*Salix viminalis* L.) in purifying sewage sludge treated soils. *Polish J Agron* 9:3–6
- Keoleian GA, Volk TA (2005) Renewable energy from willow biomass crops: life cycle energy, environmental and economic performance. *CRC Crit Rev Plant Sci* 24:385–406. <https://doi.org/10.1080/07352680500316334>
- Kozak A, Kozak R (2003) Does cross validation provide additional information in the evaluation of regression models? *Can J For Res* 33: 976–987. <https://doi.org/10.1139/X03-022>
- Kraus K, Mikhail EM (1972) Linear least squares interpolation. *Photogramm Eng* 38:1016–1029
- Kraus K, Pfeifer N (1998) Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS J Photogramm Remote Sens* 53:193–203. [https://doi.org/10.1016/S0924-2716\(98\)00009-4](https://doi.org/10.1016/S0924-2716(98)00009-4)
- Kuzovkina YA, Quigley MF (2005) Willows beyond wetlands: uses of *Salix* L. species for environmental projects. *Water Air Soil Pollut* 162:183–204. <https://doi.org/10.1007/s11270-005-6272-5>
- Lefsky MA, Cohen W, Spies T (2001) An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Can J For Res* 31:78–87. <https://doi.org/10.1139/cjfr-31-1-78>
- Lefsky MA, Cohen WB, Parker GG, Harding DJ (2002) LiDAR remote sensing for ecosystem studies: LiDAR, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *Bioscience* 52:19–30. <https://doi.org/10.1641/0006-3568>
- Lefsky MA, Hudak AT, Cohen WB, Acker SA (2005) Geographic variability in LiDAR predictions of forest stand structure in the Pacific Northwest. *Remote Sens Environ* 95(4):532–548. <https://doi.org/10.1016/j.rse.2004.12.022>
- Lever J, Krzywinski M, Altman N (2016) Points of significance: model selection and overfitting. *Nat Methods* 13:703–704. <https://doi.org/10.1038/nmeth.3968>
- Li Y, Andersen H-E, McGaughey R (2008) A comparison of statistical methods for estimating forest biomass from light detection and ranging data. *J Appl For* 23:223–231
- Lim K, Treitz P, Wulder M (2003) LiDAR remote sensing of forest structure. *Prog Phys Geogr* 1:1–38. <https://doi.org/10.1191/0309133303pp360ra>
- Magdaleno-Mas F, Martínez-Romero R (2006) Aplicaciones de la teledetección láser (LiDAR) en la caracterización y gestión del medio fluvial. *Ingeniería Civil* 142:1–15
- 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
- Maltamo M, Mustonen K, Hyyppä J, Pitkänen J, Yu X (2004) The accuracy of estimating individual tree variables with airborne laser scanning in a boreal nature reserve. *Can J For Res* 34:1791–1801. <https://doi.org/10.1139/X04-055>
- Mandeville PB (2008) Tema 18: ¿Por qué se deben centrar las covariables en regresión lineal? *Ciencia Uanl* 3:300–305
- Maune D (2007) Digital Elevation Model Technologies and Applications: The DEM User's Manual, 2nd edn. American Society for Photogrammetry and Remote Sensing, Bethesda
- McGaughey R (2010) Fusion/LDV: software for LiDAR data analysis and visualization. USDA Forest Service. Pacific Northwest Research Station. FUSION Manual. FUSION Version 2.80
- Means J, Acker S (2000) Predicting forest stand characteristics with airborne scanning LiDAR. *Photogramm Eng Remote Sens* 66:1367–1371. [https://doi.org/10.1016/S0034-4257\(01\)00290-5](https://doi.org/10.1016/S0034-4257(01)00290-5)
- Means JE, Acker SA, Harding DJ, Blair JB, Lefsky MA, Cohen WB, Harmon ME, McKee WA (1999) Use of large-footprint scanning airborne LiDAR to estimate forest stand characteristics in the western cascades of Oregon. *Remote Sens Environ* 67:298–308. [https://doi.org/10.1016/S0034-4257\(98\)00091-1](https://doi.org/10.1016/S0034-4257(98)00091-1)
- Montealegre Gracia AL, Lamelas Gracia MT, De la Riva Fernández J, García Martín A, Escribano Bernal F (2015) Cartografía de la biomasa aérea total en masas de *Pinus halepensis* Mill. en el entorno de Zaragoza mediante datos LiDAR-PNOA y trabajo de campo. In: *Análisis espacial y representación geográfica: innovación y aplicación*. pp 769–776

- Myers RH (1990) Classical and modern regression with applications, 2nd edn. Duxbury Press, Belmont
- 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 (2002) Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens Environ* 80:88–99. [https://doi.org/10.1016/S0034-4257\(01\)00290-5](https://doi.org/10.1016/S0034-4257(01)00290-5)
- Næsset E (2011) Estimating above-ground biomass in young forests with airborne laser scanning. *Remote Sens* 32:473–501. <https://doi.org/10.1080/01431160903474970>
- Næsset E, Gobakken T (2008) Estimation of above and below ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sens Environ* 112:3079–3090. <https://doi.org/10.1016/j.rse.2008.03.004>
- Næsset E, Gobakken T, Holmgren J, Hyypää H, Hyypää J, Maltamo M, Nilsson M, Olsson H, Persson A, Söderman U (2004) Laser scanning of forest resources: the Nordic experience. *Scand J For Res* 19: 482–499. <https://doi.org/10.1080/02827580410019553>
- Paredes-Sánchez JP, García-Elcoro VE, Rosillo-Calle F, Xiberta-Bernat J (2016) Assessment of forest bioenergy potential in a coal-producing area in Asturias (Spain) and recommendations for setting up a Biomass Logistic Centre (BLC). *Appl Energy* 171:133–141. <https://doi.org/10.1016/j.apenergy.2016.03.009>
- Rosso L, Facciotto G, Bergante S, Vietto L, Nervo G (2013) Selection and testing of *Populus alba* and *Salix* spp. as bioenergy feedstock: preliminary results. *Appl Energy* 102:87–92. <https://doi.org/10.1016/j.apenergy.2012.07.042>
- Sánchez Martínez FJ, Andrés-Moreno C, García-Oliva-González M (2011) Introducción a la tecnología LiDAR. In: Marquínez-García J, Yagüe-Córdova J (eds) Guía para la aplicación del sistema nacional de cartografía de zonas inundables. Dirección General del Agua. Ministerio de Medio Ambiente y Medio Rural y Marino. Madrid, pp 17–32
- Seidel D, Ammer C (2014) Efficient measurements of basal area in short rotation forests based on terrestrial laser scanning under special consideration of shadowing. *IForest* 7:227–232. <https://doi.org/10.3832/ifor1084-007>
- Song Y, Lu Y (2015) Decision tree methods : applications for classification and prediction. *Shanghai Arch Psychiatry* 27:130–135. [10.11919/j.issn.1002-0829.215044](https://doi.org/10.11919/j.issn.1002-0829.215044)
- Styles D, Jones M (2007) Energy crops in Ireland: quantifying the potential life-cycle greenhouse gas reductions of energy-crop electricity. *Biomass Bioenergy* 31:759–772. <https://doi.org/10.1016/j.biombioe.2007.05.003>
- Suárez-Antuña F (2005) La organización de los espacios mineros de la hulla en Asturias. *Scripta Nova: Revista Electrónica de Geografía y Ciencias Sociales* 9:203
- Tahvanainen L, Rytko V (1999) Biomass production of *Salix viminalis* in southern Finland and the effect of soil properties and climate conditions on its production and survival. *Biomass Bioenergy* 16:103–117. [https://doi.org/10.1016/S0961-9534\(98\)00074-9](https://doi.org/10.1016/S0961-9534(98)00074-9)
- Tanarro I (2010) Spectroscopic Techniques. In: Escribano R, Tanarro I (eds) Spectroscopic of the Atmospheres. Consejo Superior de Investigaciones Científicas (CSIC), Madrid
- Thomas V, Treitz P, McCaughey JH, Morrison I (2006) Mapping stand-level forest biophysical variables for a mixedwood boreal forest using Lidar: an examination of scanning density. *Can J For Res* 36:34–47. <https://doi.org/10.1139/x05-230>
- Timofeev R (2004) Classification and regression trees (CART) theory and applications. Doctoral dissertation. Humboldt University, Berlin
- Treitz P, Kevin L, Murray W, Doug P, Nesbitt D, Etheridge D, (2010) LiDAR data acquisition and processing protocols for forest resource inventories in Ontario, Canada. In: *Silvilaser 2010: The 10th International Conference on LiDAR Applications for Assessing Forest Ecosystems*, Freiburg, Germany
- Van Aardt JA, Wynne RH, Oderwald RG (2006) Forest volume and biomass estimation using small-footprint LiDAR-distributional parameters on a per-segment basis. *For Sci* 52:636–649
- Venturi P, Giegler JK, Huisman W (1999) Economical and technical comparison between herbaceous (*Miscanthus x giganteus*) and woody energy crops (*Salix viminalis*). *Renew Energy* 16:1023–1026. [https://doi.org/10.1016/S0960-1481\(98\)00363-2](https://doi.org/10.1016/S0960-1481(98)00363-2)
- Wehr A, Lohr U (1999) Airborne laser scanning an introduction and overview. *ISPRS J Photogramm Remote Sens* 54:68–82
- Weih M (2004) Intensive short rotation forestry in boreal climates: present and future perspectives. *Can J For Res* 34:1369–1378. <https://doi.org/10.1139/x04-090>
- Weishampel JF, Ranson KJ, Harding DJ (1996) Remote sensing of forest canopies. *Selbyana* 17:6–14
- Wulder M (1998) Optical remote-sensing technique for the assessment of forest inventory and biophysical parameters. *Prog Phys Geogr* 22: 449–476. <https://doi.org/10.1177/030913339802200402>
- Zhao K, Popescu S, Nelson R (2009) LiDAR remote sensing of forest biomass: a scale-invariant estimation approach using airborne lasers. *Remote Sens Environ* 113:182–196. <https://doi.org/10.1016/j.rse.2008.09.009>
- Zimble DA, Evans DL, Carlson GC, Parker RC, Grado SC, Gerard PD (2003) Characterizing vertical forest structure using small-footprint airborne LiDAR. *Remote Sens Environ* 87:171–182. [https://doi.org/10.1016/S0034-4257\(03\)00139-1](https://doi.org/10.1016/S0034-4257(03)00139-1)
- Zurba K, Oertel C, Matschullat J (2013) CO<sub>2</sub> emissions from willow and poplar short rotation forestry (SRF) on a derelict mining soil. 1–4