Recibido: 06/12/2018
Aceptado: 12/12/2018

REVISTA DE TELEDETECCIÓN
Asociación Española de Teledetección
(2018) 52, 27-40
ISSN 1133-0953
EISSN 1988-8740
https://doi.org/10.4995/raet.2018.11106

Assessing the use of discrete, full-waveform LiDAR and TLS to classify Mediterranean forest species composition

Torralba, J.*, Crespo-Peremarch, P., Ruiz, L. A.

Geo-Environmental Cartography and Remote Sensing Group (CGAT), Department of Cartographic Engineering, Geodesy and Photogrammetry, Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain.

Abstract: LiDAR technology –airborne and terrestrial- is becoming more relevant in the development of forest inventories, which are crucial to better understand and manage forest ecosystems. In this study, we assessed a classification of species composition in a Mediterranean forest following the C4.5 decision tree. Different data sets from airborne laser scanner full-waveform (ALS_{FW}), discrete (ALS_{D}) and terrestrial laser scanner (TLS) were combined as input data for the classification. Species composition were divided into five classes: pure *Quercus ilex* plots (QUI); pure *Pinus halepensis* dense regenerated (HALr); pure *P. halepensis* (HAL); pure *P. pinaster* (PIN); and mixed *P. pinaster* and *Q. suber* (mPIN). Furthermore, the class HAL was subdivided in low and dense understory vegetation cover. As a result, combination of ALS_{FW} and TLS reached 85.2% of overall accuracy classifying classes HAL, PIN and mPIN. Combining ALS_{FW} and ALS_{D}, the overall accuracy was 77.0% to discriminate among the five classes. Finally, classification of understory vegetation cover using ALS_{FW} reached an overall accuracy of 90.9%. In general, combination of ALS_{FW} and TLS improved the overall accuracy of classifying among HAL, PIN and mPIN by 7.4% compared to the use of the data sets separately, and by 33.3% with respect to the use of ALS_{D} only. ALS_{FW} metrics, in particular those specifically designed for detection of understory vegetation, increased the overall accuracy 9.1% with respect to ALS_{D} metrics. These analyses show that classification in forest ecosystems with presence of understory vegetation and intermediate canopy strata is improved when ALS_{FW} and/or TLS are used instead of ALS_{D}.

Key words: airborne laser scanning, terrestrial laser scanning, classification, understory vegetation, forestry.

Evaluación del uso de LiDAR discreto, full-waveform y TLS en la clasificación por composición de especies en bosques mediterráneos

Resumen: La tecnología LiDAR, tanto en sus versiones aerotransportada como terrestre, ha adquirido relevancia en los últimos años en la realización de inventarios forestales que permiten entender y adecuar la gestión de los ecosistemas forestales. En este estudio, se evaluó la clasificación por composición de especies en un bosque mediterráneo mediante el árbol de decisión C4.5. Para ello, se emplearon diferentes conjuntos de datos derivados de LiDAR discreto (ALS_{D}), LiDAR de retorno de onda completa (*full-waveform*, ALS_{FW}) y láser escáner terrestre (TLS) como datos de entrada de la clasificación. La composición de especies se dividió en cinco clases: parcelas

To cite this article: Torralba, J., Crespo-Peremarch, P., Ruiz, L. A. 2018. Assessing the use of discrete, full-waveform LiDAR and TLS to classify Mediterranean forest species composition. *Revista de Teledetección,* 52, 27-40. https://doi.org/10.4995/raet.2018.11106

* *Corresponding author: jetorpe@upv.es*

puras de *Quercus ilex* (QUI); puras de *Pinus halepensis* regenerado (HALr); puras de *P. halepensis* (HAL); puras de *P. pinaster* (PIN); y mixta de *P. pinaster* y *Q. suber* (mPIN). Además, se realizó una subdivisión de la clase HAL en cobertura de sotobosque escasa y densa. Como resultado se obtuvo una fiabilidad del 85,2% en la clasificación de las clases HAL, PIN y mPIN combinando ALS_{FW} y TLS. En la clasificación de las cinco composiciones de especies, la fiabilidad alcanzada empleando ALS_{FW} y ALS_{D} fue del 77,0%. Finalmente, en la clasificación de las subclases de cobertura de sotobosque se logró un 90,9% de fiabilidad con ALS_{FW}. En general, la combinación de ALS_{FW} y TLS mejoró los resultados en un 7,4% en la clasificación de las clases HAL, PIN y mPIN en comparación con el uso de los datos de los sensores por separado, y en un 33,3% con respecto al uso de ALS_{D}. Las métricas ALS_{FW}, en particular aquellas diseñadas especialmente para la detección del sotobosque, mejoraron la precisión en un 9,1% con respecto a las métricas derivadas de ALS_{D}. Estos análisis muestran que el uso del ALS_{FW} y TLS mejora la clasificación de los ecosistemas forestales con presencia de sotobosque y diferentes especies arbóreas en los estratos intermedios con respecto al ALS_{D}.

Palabras clave: láser escáner aerotransportado, láser escáner terrestre, clasificación, sotobosque, forestal.

1. Introduction

Forests are complex three-dimensional systems (Shugart *et al.,* 2010), composed of different species, that establish both horizontal and vertical structural relationship, depending on the availability of resources and nutrients. Tree and shrub species diversity, size variations, and their adaptation to the environment contribute to the structural complexity and wealth of forest ecosystems (Pan *et al.,* 2013). Information concerning forest composition and structure is fundamental to understand and manage forest ecosystems (Scarascia-Mugnozza *et al.,* 2000) and produce forest inventories.

The Mediterranean forest is a global biodiversity exponent (Myers *et al.,* 2000), due to the high number of endemic plants and the heterogeneity of species in the Mediterranean Basin, where herbaceous plants, shrubs and trees coexist (Cowling *et al.,* 1996). Performing a sustainable forest management according to each forest type requires knowing the forest canopy structure, the vertical relationship among the different tree species, and the effect of tree canopy on shrubs growth. This task is especially relevant in the Mediterranean forests, one of the most disturbed hotspots by human activity for thousands of years (Geri *et al.,* 2010). Remote sensing technology allows for mapping, understanding and modelling ecosystems on a large scale (Lefsky *et al.,* 2002; Vogeler and Cohen, 2016). Nevertheless, conventional optical sensors have a limited capacity to fully represent the three-dimensional forest structure (Lefsky *et al.,* 2002).

Forest inventories have typically been the main instrument to describe forest structure and quantify forest resources (Bauwens *et al.,* 2016). However, carrying out an accurate traditional forest inventory is effort and time consuming, and consequently field data acquisition is limited (Liang *et al.,* 2018). Light Detection and Ranging (LiDAR) systems have been contributing to the estimation of biophysical parameters of forest ecosystems for the last decades (Cao *et al.,* 2014). Many studies have demonstrated the potential of LiDAR to measure and estimate several forest characteristics over a wide range of forest types (Dubayah and Drake, 2000; Wulder *et al.,* 2012; Valbuena *et al.,* 2016).

In particular, discrete Airborne Laser Scanning (ALS_{D}) has become an efficient tool for registering information from height distributions in forest stands (Zaldo *et al.,* 2010; Cao *et al.,* 2014; Ruiz *et al.,* 2016). However, ALS_{D} data have restrictions to register different vegetation layers (Crespo-Peremarch *et al.,* 2016). ALS_{FW} tries to fill this gap given that it is a system that registers the complete signal emitted by the sensor and whose point density is much higher (Heinzel and Koch, 2011), being capable of describing the physical properties of the intercepted objects (Ruiz *et al.,* 2014). ALS_{FW} data have successfully been used in forest applications such as: improving the extraction of the forest height distribution (Duong, 2010), tree species classification (Hollaus *et al.,* 2009; Heinzel and Huber, 2016) and characterizing understory vegetation (Hancock *et al.,* 2017; Crespo-Peremarch *et al.,* 2018). Similar to the ALS technology, Terrestrial Laser Scanning (TLS) has become relevant in the last 15 years

in forestry applications (Wilkes *et al.,* 2017). This scanner allows for a periodic, automatic and accurate assessment of the forest structure and the three-dimensional distribution of plant components (Liang *et al.,* 2016). TLS has shown its potential to measure forest attributes very accurately (Liang *et al.,* 2018), however its use presents some drawbacks. For instance, TLS is restricted to small areas and registration from different scans is required to avoid occlusion, making TLS configuration at each scan position time consuming. In addition, point clouds are much denser, the size of files are very large, and hence data processing is more complex (Liang *et al.,* 2016; Estornell *et al.,* 2017). Some of the forest attributes that TLS can measure accurately are diameter at breast height (DBH), tree height and tree position (Maas *et al.,* 2008; Othmani *et al.,* 2011; Kankare *et al.,* 2015; Cabo *et al.,* 2018), forest cover canopy (García *et al.,* 2011), canopy gap fraction between trees and understory (Cifuentes *et al.,* 2015; Crespo-Peremarch and Ruiz, 2017), and tree species classification (Othmani *et al.,* 2013; Lin and Herold, 2016; Åkerblom *et al.,* 2017).

Integrating these LiDAR techniques (i.e. ALS_D, ALS_{FW} and TLS) can improve the forest structural characterization and determination of species composition. The aim of this paper is to analyze and compare the classification performance in several forest structural types according to species composition and understory vegetation cover combining several data sources such as ALS (discrete and full-waveform), discrete ALS-derived products (nDSM) and TLS.

2. Material and methods

2.1. Study area

The study area (Figure 1) is located in the Natural Park of Sierra de Espadán, in the eastern Spain province of Castellón. This natural park is a Mediterranean forest with soft and rounded hills, presence of abandoned farming with artificial terraces, and mountain peaks up to 1100 meters of altitude. The European Environment Agency report from 2016 (Bastrup-Birk *et al.,* 2016) classified this area as a semi-natural forest with a natural function, composition and structure, but modified by human activities throughout history. Forest type and conditions, and species composition have been influenced by human needs and changes in land use, as well as reforestation of single species policies from the last century.

This area displays a heterogeneous landscape dominated by pure and mixed native coniferous and deciduous forests, with species of the genera *Pinus* and *Quercus*.

The most dominant species in the area is *Pinus halepensis* (Aleppo pine), which mainly forms pure stands with different even-aged and densities. Density of *P. halepensis* stands ranges from overstocked stands with small sapling (10000 to 45000 trees·ha^{-1}) to poorly and medium stocked stands with young and high forest (300 to 2500 trees·ha^{-1}). *P. pinaster* (Maritime pine) is the second most represented species in the area, forming pure stands with densities ranging from 800 to 1250 trees·ha^{-1}, and mixed stands with *Quercus suber* (cork oak) as codominant species at the upper strata, ranging from 500 to 1200 trees·ha^{-1}. *Quercus ilex* (Holm oak) shows up in punctual places forming pure stands and sometimes mixed with other species such as pines or oaks. In some areas, mixed stands are observed, where *P. pinaster* dominates the upper strata, while *Q. suber* and *Q. ilex*, and *Juniperus thurifera* (*Spanish juniper*) are codominant species with densities between 500 and 800 trees·ha^{-1}.

Understory vegetation presence and density are very heterogeneous in this ecosystem, and depend on the tree composition (Crespo-Peremarch *et al.,* 2018). Forest stands dominated by *P. halepensis* have taller and denser understory vegetation than those dominated by *P. pinaster* and *Q. suber.* The most common genera of the understory species are *Erica, Genista, Rhamnus, Pistacia, Juniperus, Rosmarinus, Quercus, Phillyrea, Daphne* and *Thymus*.

2.2. Field inventory

Data were collected in 74 circular plots (706 m^2) distributed throughout the study area in September 2015. Data collected from each plot included DBH from trees with a value above 5 cm, height and canopy base height from the seven dominant trees in each plot, tree species, and percentage of understory vegetation cover.

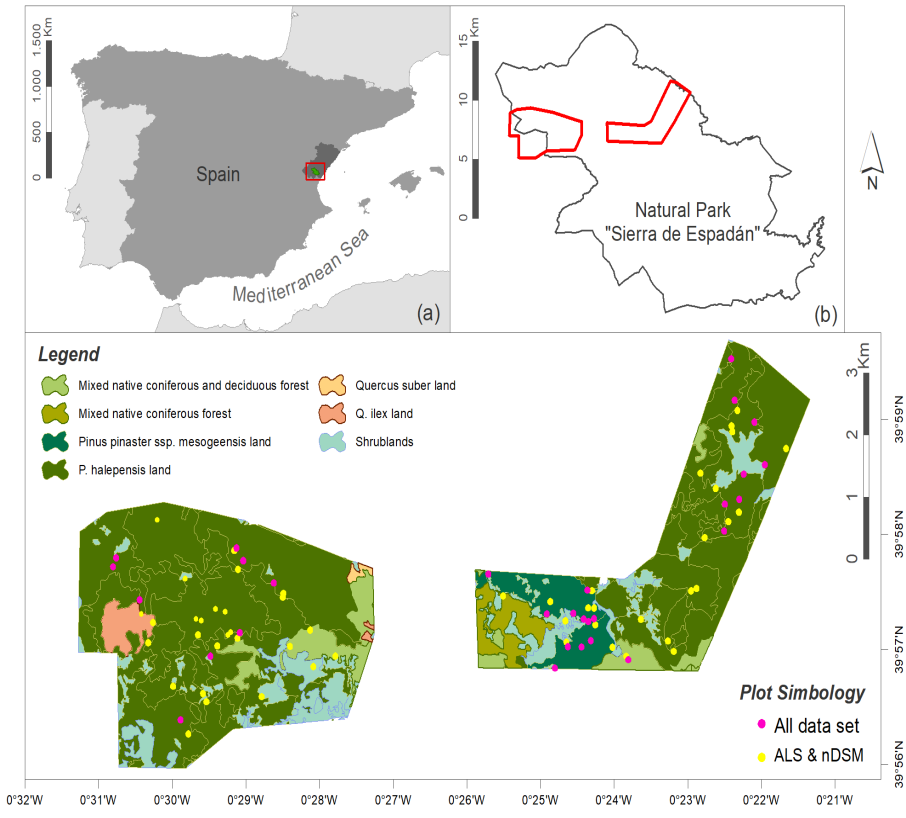

Figure 1. Study area location in (a) Castellón, Spain (South-Western Europe), (b) Natural Park of Sierra de Espadán, and (c) study area delimitation multi-data plots represented in pink, and ALS and nDSM plots in yellow. The different forest types extracted from the Forest Map of Spain (Scale 1:50000) (Magrama, 2006) are also represented.

2.3. ALS data

ALS data were acquired on September 16th 2015 flying over the entire study area (7465.53 ha) using a LiteMapper 6800 sensor with an average pulse density of 14 pulses·m^{-2}, whose characteristics are showed in Table 1. The flight altitude ranged from 600 to 820 m above sea level with a minimum overlap of 55% and a maximum of 77% between flight lines. Data were provided by the flight company in ALS_D and ALS_{FW} formats, being the former used to generate the Digital Terrain Model (DTM). Vertical accuracy of the ALS_D data set was verified using ground control points located in open and flat areas, obtaining a RMSE of 4.3 cm.

Table 1. ALS and TLS specifications. Adapted from Crespo-Peremarch and Ruiz (2017).

	ALS	TLS
Sensor	Lite Mapper 6800	Faro Focus 3D 120
Accuracy	240 mm (H) 150 mm (V)	± 2 mm at 25 m
Range	1600 m (operational altitude)	0.6-120 m
Returns	Up to 7	1
Pulse frequency	300 kHz	97 Hz
Scan angle	± 37°	*Horizontal:* 300° *Vertical:* 360°
Wavelength	1550 nm	905 nm
Bean divergence	≤ 0.50 mrad	0.19 mrad

The normalized Digital Surface Model (nDSM) was generated from the ALS_D point cloud using the FUSION 3.5 software (McGaughey, 2016). First, ground points were obtained from the initial point cloud by means of the filtering algorithm described by Kraus and Pfeifer (1998). Next, the DTM was computed by interpolation of ground points. A Digital Surface Model (DSM) was computed from the initial point cloud, and the difference between the DSM and the DTM was used to obtain the nDSM, also known as canopy height model (see Figure 4).

2.4. Terrestrial laser scanner data

TLS data were registered within the same two-month period as field and ALS data. Point clouds were collected in 27 out of 74 plots with a FARO FOCUS 3D 120 scanner (technical specifications in Table 1) from nine positions within each plot to minimize the occlusion, as follows: one at the plot center, four at each cardinal points (N, S, E, W) 15 m away from the plot center, and four at the secondary cardinal points (NE, SE, SW, NW) 7.5 m from the plot center. Once ground points were identified, point clouds height was normalized, then the nine scans were merged into a single point cloud. TLS pre-processing was done using LAStools software (Isenburg, 2018) (see Figure 4).

2.5. Definition of species composition classes

In order to differentiate the plots according to the percentage of trees from the same species, stand density (trees·ha^{-1}) and stand basal area (m^2·ha^{-1}) of each species were analyzed for each plot. The parameter number of trees per hectare provides information about the complex processes involved in tree competition in a given stand (Zeide, 2004). West (2009) mentioned that stocking density in a plot is an essential variable to describe the stage of development of a stand. Controlling stand density helps to prevent catastrophic fires (Scarascia-Mugnozza *et al.,* 2000) and, as a consequence, maintaining the forest with a correct density level can reduce the frequency and intensity of fires (Valbuena *et al.,* 2008). Basal area is a parameter related to the tree size, providing information about tree stand volume and growth. Since the composition of tree species can be influenced by understory vegetation diversity and composition (Palik and Engstrom, 1999; Barbier *et al.,* 2008), the distribution of the understory vegetation was also used as a criteria to categorize the plots .

Figure 2 shows a flowchart of the rules followed to categorize pure and mixed plots according to the tree density (trees·ha^{-1}) and basal area for each species.

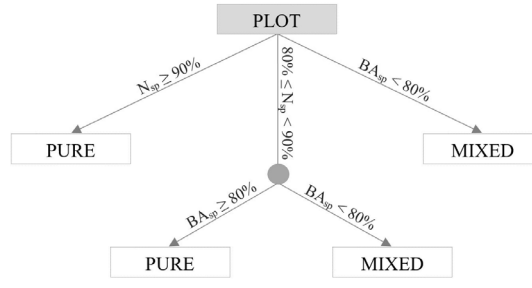

Figure 2. Rules to categorize pure and mixed plots.

Furthermore, two classes were created within pure *P. halepensis* to represent the great difference in density values. One class having overstocked stands with 12000 to 45000 trees·ha^{-1} and DBH below 10 cm, and a second class with a density ranging from 300 to 2500 trees·ha^{-1} and a DBH above 10 cm. Since the latter has a variable presence of understory vegetation, this class was subdivided into two classes: plots below 50% of understory vegetation cover, and above or equal to 50%.

As a result, five classes of plots were defined in the first instance (Table 2). After including the understory vegetation cover criteria, one of the *P. halepensis* classes was subdivided into two subgroups, obtaining a total of six classes. Class QUI represents pure plots of *Q. ilex*; class HALr pure plots of *P. halepensis* with regenerated to small sapling (fully dense); class HAL is composed of pure plots of young *P. halepensis* and high forest with less density; class PIN is composed of pure plots *of P. pinaster*; class mPIN represents mixed plots of *P. pinaster* and *Q. suber*; finally, class HAL was divided into 2 subclasses: class HAL-a below 50% and class HAL-b above or equal to 50% of understory vegetation cover. Figure 3 shows examples of field photographs from the six types of plots.

Table 2. Tree density ranges for each class, number of plots per class and number of plots with data acquired with all the sensors tested.

Name		Type	Density (trees·ha^{-1})	%Plots per class	TLS plot per class	ALS plots per class	
QUI		Pure *Q. ilex*	1000 – 5000	4	0	3	
HALr		Pure *P. halepensis*	10000 – 45000	12	0	9	
HAL	a	Pure *P. halepensis* understory < 50%	300 – 1250	20	6	15	42
	b	Pure *P. halepensis* understory ≥ 50%	300 – 2550	36	9	27	
PIN4		Pure *P. pinaster*	850 – 1250	11	5	8	
mPIN5		Mixed *P. pinaster* and *Q. suber*	450 – 1200	16	7	12	
Total				100	27	74	

2.6. Metrics extraction

Different metrics from the four data sets (i.e. ALS_D, ALS_{FW}, TLS and nDSM) were used to classify plots into the classes previously described. ALS_D and TLS metrics were extracted using FUSION 3.5 (McGaughey, 2016). This tool computes height and intensity statistics from point clouds, (see Table 3). ALS_{FW} metrics were computed using our own specific software, as reported by (Kimes *et al.,* 2006; Duncanson *et al.,* 2010; Zhang *et al.,* 2011; Ruiz *et al.,* 2016; and Crespo-Peremarch *et al.,* 2018), and can be divided into seven categories: height, energy, peaks, understory, percentiles, Gaussian decomposition, and others (see Table 4). Lastly, nDSM-derived canopy texture metrics (Table 5) were extracted using the freely available software Fetex 2.0 (Ruiz *et al.,* 2011) (http://cgat.webs.upv.es/software/).

As a result, a set of metrics from the four data sets was available for the classification into six classes according to species composition, dominance based on stem density and basal area, and understory vegetation cover.

2.7. Classification models

Several classification models were generated for different combinations of data sets, number of plots and classes. Regarding the data sets, all the possible combinations of the four data sets

Figure 3. Examples of field photographs from the six classes: *Quercus ilex* (QUI), *P. halepensis* dense regenerated (DBH < 10 cm) (HALr), *P. halepensis* and < 50% of understory vegetation cover (HAL-a), *P. halepensis* and ≥ 50% of understory vegetation cover (HAL-b), *P. pinaster* (PIN), mixed *P. pinaster* and *Q. suber* (mPIN).

Table 3. Description of ALS_D and TLS metrics (see McGaughey, 2016, for further description).

ALS_D & TLS
Name and Description
Total number of returns
Count of returns by return number (maximum 9 discrete return, only 1 TLS return)
Minimum value of height or intensity
Maximum value of height or intensity
Mean value of height or intensity
Median value of height or intensity
Mode value of height or intensity
Standard deviation value of height or intensity
Interquartile distance value of height or intensity
Skewness value of height or intensity
Kurtosis value of height or intensity
AAD: Average Absolute Deviation value of height or intensity
MADMedian: Median of the absolute deviations from the overall median value of height or intensity
MADMode: Median of the absolute deviations from the overall mode value of height or intensity
L-moments (L1, L2, L3, L4) value of height or intensity
L-moments skewness value of height or intensity
L-moments Kurtosis value of height or intensity
Percentile values of height or intensity
Canopy relief ratio ((mean-min)/(max-min))
Generalized means for the 2nd and 3rd power: Elev quadratic mean and Elev cubic mean
Percentage of first returns above a specified height (canopy cover estimate)
Percentage of first returns above the mean height/elevation
Percentage of first returns above the mode height/elevation
Percentage of all returns above a specified height
Percentage of all returns above the mean height/elevation
Percentage of all returns above the mode height/elevation
Number of returns above a specified height/total first returns × 100
Number of returns above the mean height/total first returns × 100
Number of returns above the mode height/total first returns × 100

were tested. As the number of plots registered by TLS was fewer than those registered by ALS, classification models using 27 samples were generated when TLS metrics were included, and using 74 samples when these metrics were excluded. Moreover, three sets of classes were tested: (1) the five classes described, (2) all the classes plus subclasses HAL-a and HAL-b, and (3) only subclasses HAL-a and HAL-b. For the latter classification, where *P. halepensis* young and high stands are differentiated according to below 50% and above or equal to 50% of understory vegetation cover, a new classification test discarding all the plots with a value between 40-60% was done. Since percentage of understory vegetation values were visually estimated during field work, this intermediate interval was considered to be uncertain to be used as classification samples. In this case, only data sets from ALS were used to generate classification models. Therefore, 42 samples were used to differentiate between class HAL-a and HAL-b, but only 22 samples when plots with understory vegetation cover between 40-60% were excluded.

Given that a classification model was generated for each combination of data sets, number of plots and classes, a metric selection procedure was required for each combination as well. The initial set of metrics was composed of all the metrics extracted from the data sets combined in each test. In order to reduce the number of metrics used for the classification, we used the *GreedyStepwise* algorithm beside the C4.5 classifier (Quinlan, 1993) from WEKA 3.6.12 (Hall *et al.,* 2009) for the selection of metrics. This algorithm performs forward stepwise selection starting from an empty set of metrics, and stopping the process when any

Table 4. Description of ALS_{FW} metrics.

ALS_{FW}	
Name and Description	Reference
WD: Waveform distance	(Duong, 2010)
ROUGH: Roughness of outermost canopy	
Hn: Height at nth percentile of energy	(Kimes *et al.*, 2006)
RWE: Return waveform energy	(Duong, 2010)
MAX E: Maximum energy	(Duncanson *et al.,* 2010)
VARIANCE: Variance of energy	
SKEWNESS: Skewness of energy	
HQn: Proportion of energy in nth elevation quarter	
EQn: Proportion of energy in nth energy quarter	
N GS: Number of Gaussian curves in the waveform	
N GS STARTPEAK: Number of Gaussian curves between the beginning of the waveform and the position of MAX E	
N GS ENDPEAK: Number of Gaussian curves between the position of MAX E and the end of the waveform	
CE: Canopy return energy extracted from canopy Gaussian curves	(Zhang *et al.,* 2011)
GE: Ground energy extracted from ground Gaussian curve	
GRR: Ground return ration: GE divided by RWE	
CHn: Elevation of nth quarter of energy, excluding ground Gaussian curve	
RN: CHn divided by WD	
AGS: Average Gaussian curve slope	
SGS: Standard deviation Gaussian curve slope	
MSGS: Modified standard deviation Gaussian curve slope	
HFEV: Height at first empty voxel	(Crespo-Peremarch *et al.,* 2018)
HFEVT: Height at first empty voxel from threshold	
EFEV: Energy from beginning of the waveform to first empty voxel	
nEFEV: Energy from beginning of the waveform to first empty voxel divided by RWE	
FVU: Filled voxels at understory	
NFVU: Filled voxels at understory divided by number of voxels	
BC: Bottom of canopy: elevation of the first canopy Gaussian curve	
BCE: Bottom of canopy energy: energy from the beginning of the waveform to BC	
BCD: Bottom of canopy distance: distance from BC to the end of the waveform	

remaining metric does not improve the classification. As a result, each combination of data sets, number of plots and classes had its own set of selected metrics for the classification.

Once metric selection was performed for each data set combination, the C4.5 algorithm from WEKA 3.6.12 was used to classify the same data sets used for the selection of metrics. Models were generated by cross-validation, and evaluated using the overall accuracy and kappa index. Additionally, confusion matrices were used to assess misclassification between classes.

Table 5. Description of nDSM metrics (see *Ruiz et al.,*2018 for further description).

nDSM	
Name and Description	Class
MeanEDG: Mean value of edgeness factor	Texture Features
STDEVEDG: Standard deviation of edgeness factor	
UNIFOR: Grey Level Co-occurrence Matrix (GLCM) uniformity	
ENTROP: GLCM entropy	
CONTRAS: GLCM contrast	
IDM: GLCM inverse difference moment	
COVAR: GLCM covariance	
Variance: GLCM variance	
Correlation: GLCM correlation	
Skewness: Histogram skewness	
Kurtosis: Histogram kurtosis	

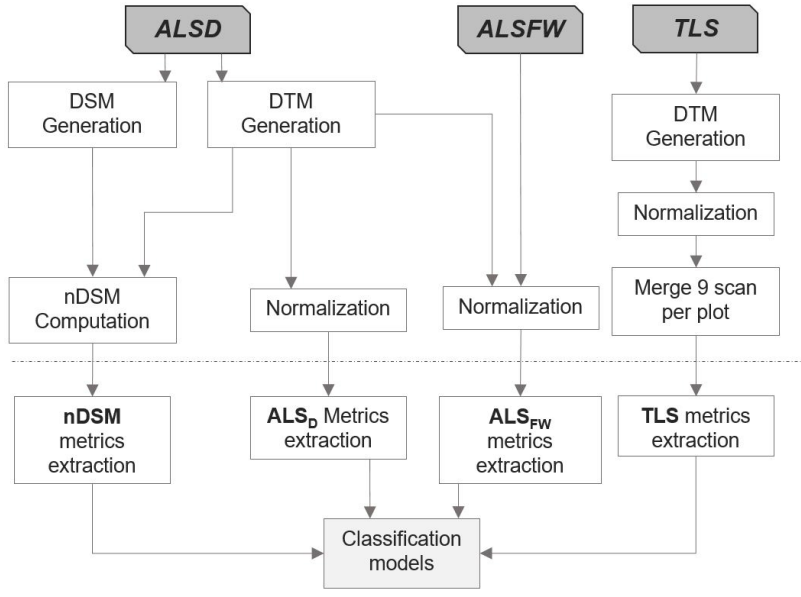

Figure 4. Flowchart of ALS and TLS data processing.

3. Results

A summary of the overall accuracy and kappa index results for all the classification models tested is showed in Table 6. In general, classification by species composition and understory vegetation cover had a higher accuracy when metrics derived from ALS_{FW} and TLS were used. The model combining ALS_{FW} and TLS reached 85.2% of overall accuracy classifying classes HALr, PIN and mPIN, using 27 sample plots. The combination of ALS_{FW} and TLS increased the overall accuracy by 7.4% with respect to only using ALS data set, being the most influential metrics the 25th percentile of the height from TLS data and the mean of HQ2 from ALS_{FW} data. When class HAL was subdivided into subclasses HAL-a and HAL-b, including understory vegetation cover, an overall classification of 74.1% was reached, being 11.1% lower than without considering understory vegetation. Again, ALS_{FW} and TLS data were the best combination tested. The most influential metrics for tree species and understory vegetation classification for this model were the 25th percentile of the height from TLS data, and the standard deviation of the Height at First Empty Voxel from Threshold (HFEVT) and the normalized Energy from the beginning of the waveform to the First Empty Voxel divided by the total waveform energy (nEFEV) from ALS_{FW} data.

Analyzing the behavior of ALS_D and ALS_{FW} in the classification of species composition and structure including 74 plots (classes QUI, HALr, HAL, PIN and mPIN), the overall accuracy did not change considerably. However, when subclasses HAL-a and HAL-b were incorporated, differences in the overall accuracy were 58.1% using ALS_{FW}, 64.9% for ALS_D, and 70.3% using both data sets.

When differences of understory vegetation cover within the *P. halepensis* class (HAL-a and HAL-b) were considered by discarding those plots that have understory vegetation cover between 40% and 60%, ALS_{FW} metrics increased accuracy by 9% with respect to ALS_D metrics. Combining both data sets the overall accuracy was 90.9%, being the 75th percentile of intensity from ALS_D, and the Height at First Empty Voxel (HFEV) and the mean of the maximum energy (MAXEmean) from ALS_{FW} the most relevant metrics. In the case where the 42 plots were classified (i.e. including classes HAL-a $<$ 50% and HAL-b $\geq$ 50% of understory cover), ALS_{FW} metrics increased the overall accuracy above 9.1% compared to ALS_D metrics.

4. Discussion and conclusions

In this research, a comparative analysis of the classification by species composition using ALS_{FW}, ALS_D, TLS and nDSM data was performed.

Table 6. Overall accuracy and kappa index results for several classification models: different combinations of data sets, number of plots, and classes. The color ranges from dark blue with a kappa index = 1 to dark orange with a kappa index = 0. The values in bold are the models with the highest accuracy (%) for each model.

	27 Plots				74 Plots				22 Plots		44 Plots	
	Class: HAL / PIN / mPIN		Class: HAL-a / HAL-b / PIN / mPIN		Class: QUI / HALr / HAL / PIN / mPIN		Class: QUI / HALr / HAL-a / HAL-b / PIN mPIN		Class: HAL-a HAL-b		Class: HAL-a HAL-b	
Combinations of data sets	Acc. (%)	Kappa	Acc. (%)	Kappa	Acc. (%)	Kappa	Acc. (%)	Kappa	Acc. (%)	Kappa	Acc. (%)	Kappa
ALS_{FW} ALS_{D} nDSM TLS	85.19	0.75	74.07	0.64								
ALS_{FW} ALS_{D} nDSM	77.78	0.62	55.56	0.38	77.03	0.63	70.27	0.61				
ALS_{FW} ALS_{D} TLS	85.19	0.75	74.07	0.64								
ALS_{FW} nDSM TLS	85.19	0.75	74.07	0.64								
ALS_{D} nDSM TLS	85.19	0.74	74.07	0.64								
ALS_{FW} ALS_{D}	77.78	0.62	55.56	0.38	77.03	0.63	70.27	0.61	90.91	0.79	76.19	0.50
ALS_{FW} nDSM	77.78	0.62	59.26	0.43	74.32	0.57	58.11	0.45				
ALS_{FW} TLS	85.19	0.75	74.07	0.64								
ALS_{D} nDSM	51.85	0.21	59.26	0.44	75.68	0.57	64.86	0.53				
ALS_{D} TLS	85.19	0.74	74.07	0.64								
nDSM TLS	77.78	0.62	66.67	0.54								
ALS_{FW}	77.78	0.62	59.26	0.43	74.32	0.57	58.11	0.45	86.36	0.70	78.57	0.54
ALS_{D}	51.85	0.210	48.15	0.28	71.62	0.50	64.86	0.53	77.27	0.45	66.67	0.29
nDSM	70.37	0.47	51.85	0.34								
TLS	77.78	0.61	66.67	0.54								

Acc.: Accuracy.

Results showed that the species composition types proposed in a Mediterranean landscape can be accurately classified using ALS_{FW} and TLS data, and the understory vegetation cover classes using only ALS_{FW} (in this case, no sufficient plots to test TLS data were available). ALS_{D} and nDSM do not improve differentiation of species composition for classes *P. halepensis* pure plots, *P. pinaster* pure plots, and *P. pinaster* mixed plots (HAL, PIN and mPIN respectively), misclassifying classes *P. pinaster* pure and *P. pinaster* mixed, due to the limitations of ALS_{D} to register intermediate vertical strata. ALS_{D} limitations are evident in mixed plots, where despite the diversity of species there is not variability in canopy heights, since the different species are mixed in the different vertical strata. For instance, *Q. suber* occupies the spaces of light left by *P. pinaster*, having large canopies in the intermediate strata under pine canopies, where ALS_{D} data have limited access.

ALS_{FW} and TLS data are relevant in the discrimination between pure and mixed plots, as well as for determining the understory vegetation cover differences in class HAL (i.e. HAL-a and HAL-b). The highest point density from TLS corresponds to understory vegetation, stem and lower strata of the canopy (Crespo-Peremarch and Ruiz, 2017), therefore TLS data are relevant to analyze and classify species composition in the lower strata. It is also remarkable that ALS_{FW} metrics, in particular those specifically designed for analyzing understory vegetation cover, improve the understory vegetation classification. When ALS_{FW} and TLS data are used instead of nDSM and ALS_{D}, results increase by 14.8%. Analyzing the four data sets separately, classification accuracies range from 48.1% to 66.7%, having all data sets low efficiency discriminating classes *P. pinaster pure* and *P. pinaster mixed*. Discrimination of classes HAL-a, HAL-b, PIN and mPIN improves significantly when ALS_{FW} and TLS are combined,

compared to the results obtained using only ALS_D. The lowest accuracy is obtained using ALS_D data set, whose highest point density is at the top of the trees, since the forest stands studied are even-aged and they present similar canopy heights.

Regarding the classification of the five forest classes, the model obtained using ALS_D misclassifies the different classes except for class *P. halepensis* in state of regenerated (HALr), a forest type characterized by very high density of trees. In contrast, ALS_{FW} only misclassifies classes *P. pinaster mixed* (mPIN) and *P. halepensis pure (HAL*). Accuracy is increased in data set combinations when TLS data are included, since they provide valuable information at intermediate and lower strata, crucial to differentiate these types of species composition categories. Results would probably have differed if plots had been stratified by height ranges per plot, regardless of the type of species, as the ALS_D and nDSM datasets would have increased the accuracy by their ability to record height variability.

The classification of understory vegetation improves considerably when ALS_{FW} metrics are used. This is coherent with previous studies (Crespo-Peremarch *et al.,* 2018), revealing potential of this technique in understory characterization due to its penetration through the forest canopy. However, more detailed field reference data is needed to properly categorize these classes and quantify their classification using ALS.

Compared to similar studies, the accuracy obtained in this study to classify classes HAL, PIN and mPIN is similar to the 91.0% reached by Heinzel and Koch (2011) to classify conifers and broadleaf trees using ALS_{FW}. They showed, beside Cao *et al.* (2014), that pure plots were classified easier than mixed plots using ALS_{FW}. Additionally, our accuracy to differentiate classes QUI, HALr, HAL, PIN and mPIN (77.0%) is similar to that obtained by Hollaus *et al.* (2009) classifying conifers and deciduous trees (83.0%) using ALS_{FW}. Since these results, however, were obtained in a different type of forest ecosystem. Therefore, the comparative results should be considered only as a qualitative reference, enhancing the fact that the use of ALS_{FW} increases the discrimination of understory vegetation.

In practice, the type of forest seems to be crucial in the selection of the data set to be used. If structural types to classify are based on height differences, then sensors that collect information from crown cover, such as ALS_D, may provide sufficient accuracy. However, when forest composition differs mainly in the distribution of vertical strata, sensors that are able to penetrate through the canopy and to register denser distribution of point clouds, such as ALS_{FW} and TLS, are expected to perform better.

Due to its complementarity to register different forest strata, integration of ALS_{FW} and TLS data has demonstrated potential for classifying forest species compositions, in particular understory vegetation cover, that is not always considered in traditional forest inventories, even being a key element for ecosystems, wildlife, soil retention, and fire behavior modeling as a good parameter to quantify ladder fuel. For further work, considering other forest parameters, such as height or diameter distributions of forest landscapes, could potentially improve Mediterranean forest ecosystem characterization.

Acknowledgments

This research has been funded by the Spanish Ministerio de Economía y Competitividad and FEDER, in the framework of the project CGL2016-80705-R.

References

Åkerblom, M., Raumonen, P., Mäkipää, R., Kaasalainen, M. 2017. Automatic tree species recognition with quantitative structure models. *Remote Sensing of Environment*, *191*, 1-12. https://doi.org/10.1016/j.rse.2016.12.002

Barbier, S., Gosselin, F., Balandier, P. 2008. Influence of tree species on understory vegetation diversity and mechanisms involved—A critical review for temperate and boreal forests. *Forest Ecology and Management*, *254*(1), 1-15. https://doi.org/10.1016/j.foreco.2007.09.038

Bastrup-Birk, A., Reker, J., Zal, N. 2016. *European forest ecosystems: State and trends*. *EEA Report n° 5/2016*. Copenhagen. https://doi.org/10.2800/964893

Bauwens, S., Bartholomeus, H., Calders, K., Lejeune, P. 2016. Forest Inventory with Terrestrial LiDAR: A Comparison of Static and Hand-Held Mobile Laser Scanning. *Forests*, *7*(12), 127. https://doi.org/10.3390/f7060127

Cabo, C., Ordóñez, C., López-Sánchez, C. A., Armesto, J. 2018. Automatic dendrometry: Tree detection, tree height and diameter estimation using terrestrial laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, *69*(November 2017), 164-174. https://doi.org/10.1016/j.jag.2018.01.011

Cao, L., Coops, N., Hermosilla, T., Innes, J., Dai, J., She, G. 2014. Using Small-Footprint Discrete and Full-Waveform Airborne LiDAR Metrics to Estimate Total Biomass and Biomass Components in Subtropical Forests. *Remote Sensing*, *6*(8), 7110-7135. https://doi.org/10.3390/rs6087110

Cifuentes, R., Zande, D. Van Der, Farifteh, J., Salas, C., Coppin, P. 2015. Effects of voxel size and sampling setup on the estimation of forest canopy gap fraction from terrestrial laser scanning data. *Agricultural and Forest Meteorology*, *201*(August), 416. https://doi.org/10.1016/j.agrformet.2015.08.226

Cowling, R. M., Rundel, P. W., Lamont, B. B., Kalin Arroyo, M., Arianoutsou, M. 1996. Plant diversity in mediterranean-climate regions. *Trends in Ecology & Evolution*, *11*(9), 362-366. https://doi.org/10.1016/0169-5347(96)10044-6

Crespo-Peremarch, P., Ruiz, L. A., Balaguer-Beser, A. 2016. A comparative study of regression methods to predict forest structure and canopy fuel variables from LiDAR full-waveform data. *Revista de Teledetección*, *45*, 27-40. https://doi.org/10.4995/raet.2016.4066

Crespo-Peremarch, P., Ruiz, L. Á. 2017. Análisis comparativo del potencial del ALS y TLS en la caracterización estructural de la masa forestal basado en voxelización. In *Nuevas plataformas y sensores de teledetcción. XVII Congreso de la Asociación Española de Teledetección* (pp. 131-135). Murcia: Asociación Española de Teledetección.

Crespo-Peremarch, P., Tompalski, P., Coops, N. C., Ruiz, L. Á. 2018. Characterizing understory vegetation in Mediterranean forests using full-waveform airborne laser scanning data. *Remote Sensing of Environment*, *217*(August), 400-413. https://doi.org/10.1016/j.rse.2018.08.033

Dubayah, R. O., Drake, J. B. 2000. Lidar Remote Sensing for Forestry Applications. *Journal of Forestry*, *98*(6), 44-46. https://doi.org/10.1093/jof/98.6.44

Duncanson, L. I., Niemann, K. O., Wulder, M. A. 2010. Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2009.08.018

Duong, V. H. 2010. *Processing and Application of ICESat Large Footprint Full Waveform Laser Range Data*. Delft University of Technology, Delft, The Netherlands.

Estornell, J., Velázquez-Martí, A., Fernández-Sarría, A., López-Cortés, I., Martí-Gavilá, J., Salazar, D. 2017. Estimación de parámetros de estructura de nogales utilizando láser escáner terrestre. *Revista de Teledetección*, *48*, 67. https://doi.org/10.4995/raet.2017.7429

García, M., Danson, F. M., Riaño, D., Chuvieco, E., Ramirez, F. A., Bandugula, V. 2011. Terrestrial laser scanning to estimate plot-level forest canopy fuel properties. *International Journal of Applied Earth Observation and Geoinformation*, *13*(4), 636-645. https://doi.org/10.1016/j.jag.2011.03.006

Geri, F., Amici, V., Rocchini, D. 2010. Human activity impact on the heterogeneity of a Mediterranean landscape. *Applied Geography*, *30*(3), 370-379. https://doi.org/10.1016/J.APGEOG.2009.10.006

Hancock, S., Anderson, K., Disney, M., Gaston, K. J. 2017. Measurement of fine-spatial-resolution 3D vegetation structure with airborne waveform lidar: Calibration and validation with voxelised terrestrial lidar. *Remote Sensing of Environment*, *188*, 37-50. https://doi.org/10.1016/J.RSE.2016.10.041

Heinzel, J., Koch, B. 2011. Exploring full-waveform LiDAR parameters for tree species classification. *International Journal of Applied Earth Observation and Geoinformation*, *13*(1), 152-160. https://doi.org/10.1016/J.JAG.2010.09.010

Heinzel, J., Huber, M. 2016. Detecting tree stems from volumetric TLS data in forest environments with rich understory. *Remote Sensing*, *9*(1), 9. https://doi.org/10.3390/rs9010009

Hollaus, M., Mücke, W., Höfle, B., Dorigo, W., Pfeifer, N., Wagner, W., … Regner, B. 2009. Tree species classification based on full-waveform airborne laser scanning data. In *Silvilaser 2009* (Vol. 54). Texas, USA.

Isenburg, M. 2018. LAStools - Efficient tools for LiDAR processing. (Version 180409) obtained from http://rapidlasso.com/LAStools. Alemania: Rapidlasso GmbH.

Kankare, V., Liang, X., Vastaranta, M., Yu, X., Holopainen, M., Hyyppä, J. 2015. Diameter distribution estimation with laser scanning based multisource single tree inventory. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 161-171. https://doi.org/10.1016/j.isprsjprs.2015.07.007

Kimes, D. S., Ranson, K. J., Sun, G., Blair, J. B. 2006. Predicting lidar measured forest vertical structure from multi-angle spectral data. *Remote Sensing of Environment*, *100*(4), 503-511. https://doi.org/10.1016/j.rse.2005.11.004

Kraus, K., Pfeifer, N. 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *53*(4), 193-203. https://doi.org/10.1016/S0924-2716(98)00009-4

Lefsky, M. A., Cohen, W. B., Parker, G. G., Harding, D. J. 2002. Lidar Remote Sensing for Ecosystem Studies. *BioScience*, *52*(1), 19-30. https://doi.org/10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., … Vastaranta, M. 2016. Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 63-77. https://doi.org/10.1016/j.isprsjprs.2016.01.006

Liang, X., Hyyppä, J., Kaartinen, H., Lehtomäki, M., Pyörälä, J., Pfeifer, N., … Wang, Y. 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, *144*(October 2018), 137-179. https://doi.org/10.1016/j.isprsjprs.2018.06.021

Lin, Y., Herold, M. 2016. Tree species classification based on explicit tree structure feature parameters derived from static terrestrial laser scanning data. *Agricultural and Forest Meteorology*, *216*, 105-114. https://doi.org/10.1016/j.agrformet.2015.10.008

Maas, H. G., Bienert, A., Scheller, S., Keane, E. 2008. Automatic forest inventory parameter determination from terrestrial laser scanner data. *International Journal of Remote Sensing*, *29*(5), 1579-1593. https://doi.org/10.1080/01431160701736406

Magrama. 2006. Mapa Forestal de España. Escala 1:50.000. Ministerio de Agricultura, Alimentación y Medio Ambiente. Dirección General de Desarrollo Rural y Política Forestal.

McGaughey, R. J. 2016. FUSION/LDV: Software for LIDAR Data Analysis and Visualization. *Seattle (WA): USDS Forest Service, Pacific Northwest Research Station*. https://doi.org/10.1097/BRS.0b013e3182a439cc

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853-858. https://doi.org/10.1038/35002501

Othmani, A., Piboule, A., Krebs, M., Stolz, C. 2011. Towards automated and operational forest inventories with T-Lidar. *SilviLaser*, 1-9.

Othmani, A., Lew Yan Voon, L. F. C., Stolz, C., Piboule, A. 2013. Single tree species classification from Terrestrial Laser Scanning data for forest inventory. *Pattern Recognition Letters*, *34*(16), 2144-2150. https://doi.org/10.1016/j.patrec.2013.08.004

Palik, B., Engstrom, R. T. 1999. Species composition. In M. L. Hunter (Ed.), *Maintaining Biodiversity in Forest Ecosystems* (pp. 65-94). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511613029.005

Pan, Y., Birdsey, R. A., Phillips, O. L., Jackson, R. B. 2013. The Structure, Distribution, and Biomass of the World's Forests. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 593-622. https://doi.org/10.1146/annurev-ecolsys-110512-135914

Ruiz, L. A., Hermosilla, T., Mauro, F., Godino, M. 2014. Analysis of the influence of plot size and LiDAR density on forest structure attribute estimates. *Forests*, *5*(5), 936-951. https://doi.org/10.3390/f5050936

Ruiz, L. Á., Recio, J. A., Crespo-Peremarch, P., Sapena, M. 2018. An object-based approach for mapping forest structural types based on low-density LiDAR and multispectral imagery. *Geocarto International*, *33*(5), 443-457. https://doi.org/10.1080/10106049.2016.1265595

Scarascia-Mugnozza, G., Oswald, H., Piussi, P., Radoglou, K. 2000. Forests of the Mediterranean region: gaps in knowledge and research needs. *Forest Ecology and Management*, *132*(1), 97-109. https://doi.org/10.1016/S0378-1127(00)00383-2

Shugart, H. H., Saatchi, S., Hall, F. G. 2010. Importance of structure and its measurement in quantifying function of forest ecosystems. *Journal of Geophysical Research: Biogeosciences*, *115*(G2), n/a-n/a. https://doi.org/10.1029/2009JG000993

Valbuena, P., Del Peso, C., Bravo, F. 2008. Stand Density Management Diagrams for two Mediterranean pine species in Eastern Spain. *Investigación Agraria: Sistemas y Recursos Forestales*, *17*(2), 97. https://doi.org/10.5424/srf/2008172-01026

Valbuena, R., Maltamo, M., Packalen, P. 2016. Classification of forest development stages from national low-density lidar datasets: a comparison of machine learning methods. *Revista de Teledetección*, *45*, 15-25. https://doi.org/10.4995/raet.2016.4029

Vogeler, J. C., Cohen, W. B. 2016. A review of the role of active remote sensing and data fusion for characterizing forest in wildlife habitat models. *Revista de Teledetección*, *45*, 1-14. https://doi.org/10.4995/raet.2016.3981

West, P. W. 2009. *Tree and Forest Measurement*. *Springer-Verlag Berlin Heidelberg* (2nd ed.). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-540-95966-3

Wilkes, P., Lau, A., Disney, M., Calders, K., Burt, A., Gonzalez de Tanago, J., … Herold, M. 2017. Data acquisition considerations for Terrestrial Laser Scanning of forest plots. *Remote Sensing of Environment*, *196*, 140-153. https://doi.org/10.1016/j.rse.2017.04.030

Wulder, M. A., White, J. C., Nelson, R. F., Næsset, E., Ørka, H. O., Coops, N. C., … Gobakken, T. 2012. Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment*, *121*, 196-209. https://doi.org/10.1016/J.RSE.2012.02.001

Zaldo, V., Moré, G., Pons, X. 2010. Estimación y cartografía de parámetros ecológicos y forestales en tres especies (*Quercus ilex* L. subsp *ilex*, *Fagus sylvatica* L. y *Pinus halepensis* L.) con datos LiDAR. *Revista de Teledetección*, *34*, 55-68.

Zeide, B. 2004. Stand Density and Canopy Gaps. In K. F. Connor (Ed.), *Gen. Tech. Rep. SRS 71. US Department of Agriculture, Forest Service, Southern Research Station* (pp. 79-183). Biloxi, Mississippi: USDA Forest Service Southern Research Station, Asheville, North Carolina.

Zhang, J., de Gier, A., Xing, Y., Sohn, G. 2011. Full Waveform-based Analysis for Forest Type Information Derivation from Large Footprint Spaceborne Lidar Data. *Photogrammetric Engineering & Remote Sensing*, *77*(3), 281-290. https://doi.org/10.14358/PERS.77.3.281