Combining low-density LiDAR and satellite images to discriminate species in mixed Mediterranean forest

Ángela Blázquez-Casado^{1,2} · Rafael Calama^{2,3} · Manuel Valbuena⁴ · Marta Vergarechea^{2,3} · Francisco Rodríguez^{1,2,5}

Received: 9 April 2018 / Accepted: 15 April 2019 / Published online: 17 May 2019
© INRA and Springer-Verlag France SAS, part of Springer Nature 2019

Abstract

• **Key message** Using a combination of remote sensing data, *Pinus pinaster* Ait. and *Pinus pinea* L. were distinguished at individual tree level in mixed Mediterranean stands with over 95% accuracy. This approach is easily applicable over large areas, enhancing the economic value of non-wood forest products, stone pine nuts, and resin, and aiding forest managers to accurately predict this production.

• **Context** The discrimination of tree species at individual level in mixed Mediterranean forest based on remote sensing is a field which has gained greater importance. In these stands, the capacity to predict the quality and quantity of non-wood forest products is particularly important due to the very different goods the two species produce.

• **Aims** To assess the potential of using low-density airborne LiDAR data combined with high-resolution Pleiades images to discriminate two different pine species in mixed Mediterranean forest (*Pinus pinea* L. and *Pinus pinaster* Ait.) at individual tree level.

• **Methods** A Random Forest model was trained using plots from the pure stand dataset, determining which LiDAR and satellite variables allow us to obtain better discrimination between groups. The model constructed was then validated by classifying individuals in an independent set of pure and mixed stands.

• **Results** The model combining LiDAR and Pleiades data provided greater accuracy (83.3% and 63% in pure and mixed validation stands, respectively) than the models which only use one type of covariables.

• **Conclusion** The automatic crown delineation tool developed allows two very similar species in mixed Mediterranean conifer forest to be discriminated using continuous spatial information at the surface: Pleiades images and open source LiDAR data. This approach is easily applicable over large areas, enhancing the economic value of non-wood forest products and aiding forest managers to accurately predict production.

Keywords Crown delineation · Inventory · Modeling · Pleiades · Remote sensing · Stone pine · LiDAR

Handling Editor: Barry Alan Gardiner

Contributions of the co-authors Conceptualization: Francisco Rodríguez; data curation: Rafael Calama, Manuel Valbuena, and Marta Vergarechea; data analysis: Ángela Blázquez-Casado; writing (original draft): Ángela Blázquez-Casado; writing (review and editing): Rafael Calama, Ángela Blázquez-Casado, and Francisco Rodríguez; supervising the work: Francisco Rodríguez.

This article is part of the topical collection on Mediterranean pines

✉ Ángela Blázquez-Casado
ange.blazquez@gmail.com

¹ föra forest technologies, C/Eduardo Saavedra, 38, 42004 Soria, Spain

² iuFOR University Institute for Sustainable Forest Management
INIA-UVA, Palencia, Spain

³ Departamento de Selvicultura y Gestión de Sistemas Forestales,
INIA-CIFOR, Ctra. A Coruña, km 7.5, 28040 Madrid, Spain

⁴ Departamento de Educación Gobierno Vasco, IES Murgía BHI,
Álava, Spain

⁵ Departamento de Producción Vegetal y Recursos Forestales, EIFAB,
Universidad de Valladolid, Campus Duques de Soria,
42005 Soria, Spain

1 Introduction

Forest management is currently undergoing important changes and challenges with the up-to-date development and use of new tools built up based on remote sensing technology. In general, field data have always been gathered manually, but nowadays, the application of new technology in forest inventories has revolutionized this field and has brought down the costs associated with forest inventories (Barrett et al. 2016). These new technologies have the greatest potential for evaluating forest characteristics, given that canopy vegetation height, one of the key data on forest resources evaluation, is a function of species composition, site quality, and age; hence, this data can be used for land cover classification, habitat mapping, and forest management (Dubayah and Drake 2000). Additionally, forest parameters related to tree height and stem volume at both plot and individual tree level can be estimated using the information from remote sensing systems (Popescu et al. 2003; Hollaus et al. 2009; Tonolli et al. 2011; Ruiz et al. 2016; Castaño-Díaz et al. 2017; Arias-Rodil et al. 2018).

An increasing amount of evidences suggest that mixed-species forests, in comparison with monospecific ones, provide a greater number of ecological and socio-economic goods and services, even in Mediterranean ecosystems (Barba et al. 2013; Riofrío et al. 2017). Consequently, the management of these forests should be based on a comprehensive assessment of their wide diversity and multifunctionality, as well as on accurate species classification. Furthermore, when the management of these stands is aimed at the individual tree level rather than stand level, species discrimination is essential in order to provide an inventory for the different resources. The use of remote sensing techniques offers the advantage that such questions can be addressed at large scales in mixed forests, including at landscape level.

The high economic and social importance of non-wood forest products (NWFP) is one of the main factors that distinguishes Mediterranean forests from other temperate forests (Calama et al. 2008). The production of stone pine nuts from *Pinus pinea* L. and resin from *Pinus pinaster* Ait. in the Mediterranean forests of the Northern Plateau of the Iberian Peninsula has made an important contribution to rural activity since ancient times (Lopez-García 1980; Nanos et al. 2000). In many stone pine forests, pine nuts are more profitable than timber, with prices reaching over 100 €/kg, in turn favoring the NWFP transformation industry. In the case of resin, tapping is an activity which is currently undergoing resurgence. It was very important until the end of the 1970s, when the increase in the labor costs and international competitiveness of Chinese natural resins tapped from *Pinus massoniana* Lamb. rendered this traditional activity no longer profitable (Soliño et al. 2018). In the recent years, within the context of the global financial crisis resulting in higher rates of rural unemployment in Spain, together with the drop in Chinese resin exports, and an

increasing demand for high-quality secondary products for cosmetics, pharmacy, and healthcare, resin tapping has become a profitable activity again, and many of the abandoned stands are being tapped again (Rodríguez-García et al. 2015; Soliño et al. 2018). *Pinus pinea* L. and *Pinus pinaster* Ait. share autoecological conditions (Alonso et al. 2012), so pine nuts and resin are both harvested in Mediterranean mixed pine forests. Hence, the accurate assessment of nut and resin yields is particularly important, not only for the forest management decision-making process but also as regards associated industrial process and evaluation of market prices.

Classification of tree species assisted by remote sensing data is motivated by a wide variety of applications for the sustainable forest management of wood resources and ecosystem services (European Environmental Agency 2006; Fassnacht et al. 2016). Thus, to know the tree species distribution in mixed stands is important, due to the fact that different species produce very different NWFP and, therefore, determine the forest management. Several studies have used remote sensing data, Light Detection and Ranging (LiDAR), and multispectral images to predict stand classification (Korpela et al. 2010; Ørka et al. 2012; Ballanti et al. 2016), for both species groups and individual tree species (Suratno et al. 2009a; Deng et al. 2016). Distinguishing between two species of the same genus is especially difficult, since the differences between them are often subtle. Species that show real-world differences in physical and/or spectral properties should be clearly recorded by the sensors (Vauhkonen et al. 2014; Fassnacht et al. 2016). However, when the studied species are similar, slight differences in crown structure (e.g., umbrella and apical crowns in *P. pinea* and *P. pinaster* respectively) and stem slenderness could be the key to discriminate between them, while spectral information concerning chlorophyll contents should help in this work.

The combined use of LiDAR and satellite images results in a much more complicated analysis than if LiDAR or satellite images are separately used. Besides, when the forest studied contains several species, each of them with their own differential structural characteristics, the challenge is even larger (Vauhkonen et al. 2014). In such stands, where species must be correctly classified, additional complexity is involved in the modeling process (crown individualization) if low-point density (< 1 point per m²) LiDAR data is used (Suratno et al. 2009b; Dalponte et al. 2012). Remote sensing images have been widely used for tree species classification in order to obtain spatially detailed species information (Xie et al. 2008; Dalponte et al. 2012). New satellite images with high spatial resolution are available today, such as the 50-cm images provided by the Pleiades constellation of satellites, which have led to improved analyses. Moreover, Pleiades images have recently been used in studies focusing on forest biomass modeling and forest structure mapping through image texture analysis (Beguet et al. 2014; Rougier and Anne 2014; Maack et al. 2015).

Current estimation and forecasting of cone and resin production are based on individual tree-level models (Nanos et al. 2001; Moreno-Fernández et al. 2013; Calama et al. 2016), thus sound estimates should be based on a previous correct assessment of tree-level features such as species and size. Individual tree delineation from LiDAR data can be used to define individual structure for many wood species (Popescu et al. 2003; Valbuena et al. 2016a). The individual crown delineation allows us to obtain tree-level variables from LiDAR which are included in the algorithms for species classification. Therefore, these tree-level LiDAR variables might also be applied to predict individual tree NWFP production by means of the currently existing models, which depends on the tree as an individual entity (Popescu 2007). Furthermore, species discrimination is often determined satisfactorily using satellite data. Hence, the operational process in pine stands should consist of first carrying out a process of crown delineation using LiDAR data, then applying classification models to identify species and extract tree-level attributes, and subsequently implementing pre-existing production models for each NWFP. Thus, the integration of the methods should permit assessing NWFP over large areas (Lee et al. 2016).

The aim of our study is to analyze the potential of using low-density LiDAR data combined with high-resolution Pleiades images to discriminate different Mediterranean pine species in pure and mixed *Pinus pinaster* Ait. and *Pinus pinea* L. forests, based on an automatic algorithm for crown delineation. The practical motivation is that by distinguishing these two species at a tree scale this will permit us to apply the pre-existing models for different non-wood forest products (resin or pine nuts) which act at a tree-level scale (e.g., Calama et al. 2016; Nanos et al. 2004). For this purpose, (i) in a first phase, a Random Forest model was trained with a dataset collected in pure stands of both species, to determine which LiDAR and satellite variables allow us to obtain the best discrimination between groups, and (ii) the Random Forest constructed was then validated by classifying individuals into an independent set of mixed and pure stands. Our hypotheses are (i) combining LiDAR and Pleiades data to discriminate species in mixed Mediterranean conifer forests should be more efficient than the use of each technique separately, (ii) this combination should provide an efficient tool to discriminate pine species in mixed Mediterranean forests at individual tree level, and (iii) the results for *Pinus pinea* L. should be more accurate than those for *Pinus pinaster* Ait. due to the fact that the former usually grows in open stands.

2 Material and methods

2.1 Study area

Forested land accounts for almost a third of the total land area of Spain (18.27 million ha). Pure forests of *Pinus pinea* L.

cover 401,701 ha, while *Pinus pinaster* Ait. forests make up 1,131,901 ha (MAGRAMA 2012). The study area is situated in the Northern Plateau of the Iberian Peninsula (Fig. 1), a region with 65,275 ha of pure *Pinus pinea* L. forests, 271,000 ha of pure *Pinus pinaster* Ait. forests, and 53,000 ha of mixed *Pinus pinea* L. and *Pinus pinaster* Ait. forests (López et al. 2009). The Duero river basin is on the Northern Plateau, 700–800 m a.s.l., with a homogenous Mediterranean continental climate, characterized by low average annual rainfall (450 mm), strong summer drought (< 50 mm), and low average annual temperatures (12 °C). Soils have a very high sand content (> 90%) and very low water holding capacity (WHC < 100 mm), except in the upper areas where limestone soils with a percentage of clays and limes (> 40–50%) reach WHC values > 250 mm.

The plots within the study area were clustered in three different forest types: pure stands of *Pinus pinea* L., pure stands of *Pinus pinaster* Ait., and mixed stands, where the least represented species accounted for at least 35% of the total number of trees. All considered stands were at the mature development stage, when regeneration is quite poor.

2.2 Field data

In this study, we used a subset from the network of permanent plots installed by the Forest Research Centre of the Spanish National Institute for Agricultural Research (INIA-CIFOR) in the province of Valladolid. Mixed and pure *Pinus pinaster* Ait. plots were installed between the summer of 2015 and spring of 2016. Pure even aged *Pinus pinea* L. plots were installed in 1996 in cooperation with the regional forest service and have been regularly monitored since then (the most recent inventory was carried out in 2016). Plots were selected in order to cover the whole range of site conditions, stand stocking and age identified within the region. The plots were all located in public mature forests, and at installation, the stand should have not been altered (thinnings, final cuttings, or pruning) during at least the previous 5 years.

The plots were circular in shape, with a variable radius. Pure plots included at least 20 trees, and mixed plots included at least 10 trees of the least represented species. Plot center coordinates were recorded using a high-precision submeter positioning Garmin GPS (Garmin International Inc., Missouri, USA), with an accuracy below 1 m. Polar coordinates (x,y) of each tree within the plot were computed after measuring their distance and heading from the center of the plot using a VERTEX IV (Haglöf, Sweden) distance meter and a compass. Trees were then numbered and measured. Tree measurements included species, diameter at breast height (1.3 m) (DBH, cm), total height (Height, m), crown base height (Live_Branch, m), and crown radius (Crown_Radio, m). Afterwards, each tree measured during the fieldwork was plotted on a map. From the whole available network of

Fig. 1 Study area. Locations of training and validation sets

plots, the least represented conditions of species composition were both pure *Pinus pinaster* Ait. plots and mixed *P. pinaster*–*P. pinea* ones. Thus, all the available plots from pure *Pinus pinaster* Ait. stands and mixed stands were included in the sample. In a second phase, we selected all the available pure *Pinus pinea* L. plots located in the vicinity of the

pure *P. pinaster* and mixed plots. In the end, 27 plots were selected for the present study (Fig. 1): nine pure *Pinus pinaster* Ait. plots, six mixed plots, and 12 pure *Pinus pinea* L. plots. Of these, seven plots in pure *Pinus pinea* L. stands and five plots in pure *Pinus pinaster* Ait. stands were randomly selected and used for training the model. The remaining 15

plots (five in pure *Pinus pinea* L. stands, four in pure *Pinus pinaster* Ait. stands, and six in mixed stands) were therefore used to validate the model. All individual trees measured inside each selected plot were included in the sample (Blázquez-Casado et al. 2019).

2.3 Remote sensing data

Low-density LiDAR data for the study area was provided by the National Aerial Orthophoto Program (PNOA). Point clouds were captured in 2010 using a laser scanner, Leica ALS60 sensor (Heerbrugg, Switzerland), with a mean density of 0.5 points per m² and a root mean square error of 20 cm in altitude accuracy. Digital files, classified and colored, were downloaded in .laz format (2 × 2 km).

Pleiades is a constellation of satellites of the National Centre for Space Studies (CNES, French Government) which provides very-high-resolution images. It was designed by the Optical and Radar Federated Earth Observation (ORFEO) program. It was initialized in 2004, and the first year of the satellite life was in 2011 (Tinel et al. 2012). Pleiades 1A and Pleiades 1B operate as a constellation in the same orbit, phased 180° apart. The identical twin satellites capture images which are transferred, orthorectified, and geometrically corrected by the Spanish Remote Sensing National Plan (PNT). Images were captured in June–July 2014, and each one covers an area of 400 km². The Pleiades images comprise four multispectral bands with a spatial resolution of 2 m. The spectral ranges of the four bands are 430–550 nm (blue), 490–610 nm (green), 600–720 nm (red), and 750–950 nm (infrared). Additionally, there is a panchromatic band (480–830 nm) with a spatial resolution of 0.5 m.

The time delay in the field validation of the remote sensing data does not increase the level of uncertainty since in these mature stands large changes in such short periods are not expected, particularly if there is no disturbance. The low productivity in these inner Mediterranean stands (mean annual volume increment about 1 m³ ha⁻¹) confirms this issue. In addition, as field data came from a net of permanent plots, we can confirm little change on time in the structure (number of stems per ha, stand composition) in the last few years.

2.4 Remotely sensed data processing

2.4.1 LiDAR data processing

LiDAR point files (in .laz format), which include the field plots, were compiled and normalized, subtracting their corresponding ground class points. Points below 3 m were eliminated as it was considered that this was the lowest level at which a distinction between bushes and trees could be drawn. Points over 50 m were considered “mistakes”, as there are no trees above this height in the study area; hence, they were also removed. The

algorithm for tree delineation is the one by Valbuena et al. (2016a), written as a function in the SQL language for use in PostgreSQL 9.2 with the POSTGIS 2.0 add-on, which gave the database manager geometrical and other GIS functions. It was applied to each normalized point cloud in .shp format. The algorithm is based on cloud point LiDAR analysis, searching directly for relative maximums. Once a relative maximum has been located, the next lower LiDAR return is found, and the distance between them is calculated. The algorithm determines whether this distance implies that the point is a new apex or whether it forms part of the first crown, thus allowing individual tree crowns to be delineated. To reach each crown, the algorithm divides the point cloud into layers half a meter deep and analyzes the distance from each point to the crown polygons already delineated in each layer, in a downward direction (Valbuena et al. 2016a). This process delineates the horizontal projection of each crown as a polygon that takes in all the points which impact on each tree, identifying individual tree crowns and drawing them in vector format (Fig. 2). The result takes the form of a .shp layer that contains each crown polygon for each tree in graphic form and an associated table with the following data for each crown: maximum height of tree crown (Max_Height, m), minimum height of tree crown (Min_Height, m), crown area of individual tree (Crown_Area, m²), and crown height of individual tree (Crown_Height, m), equal to the difference between Max_Height and Min_Height. Density (Density, trees per ha) was calculated from the centroid of the polygon using the six-tree sampling method (Prodan 1968).

In order to characterize the structural differences between *Pinus pinea* L. and *Pinus pinaster* Ait., three relations were derived from the previous variables: heights relation (HR), calculated by the relationship between maximum height and minimum height ((Max_Height – Min_Height)/Max_Height); area and height relation (AHR), estimated as the quotient between crown area and maximum tree height (Crown_Area/Max_Height), revealing the relationship between *Pinus pinea* L. and *Pinus pinaster* Ait. outline crowns, something each species is usually characterized by; and crown and height relation (CHR), calculated as the relationship between maximum height and crown base height ((Max_Height – Crown_height)/Max_Height), it expresses the relationship between height tree and natural pruning height, which is generally higher in *Pinus pinea* L. than in *Pinus pinaster* Ait. Thus, six LiDAR variables got into modeling: Crown_Area, Crown_Height, Density, HR, AHR, and CHR (Tables 1 and 2).

2.4.2 Tree crown selection

Following the process of individual tree crown delineation, a subset of trees with clearly identified crowns and which had been matched with trees in the field inventory of the plot was selected for the analysis. Selected training data in pure stands is more reliable compared with mixed ones, due to the extreme

Fig. 2 Crown delineation in a mixed stand using the Valbuena et al. (2016a) algorithm

importance of high-precision tree location at the individual tree level. Once the delineation algorithm generated the crowns, both shapes were matched one to one automatically through the QGIS intersect tool (Quantum GIS Development Team 2017), throwing away crowns when more than one measured tree coincided with one delineated crown to guarantee that each crown coincided with one measured tree. Therefore, 95.5% and 81.0% of trees measured have respectively matched with crown delineation data in *P. pinea* and *P. pinaster* pure stands. On the other hand, matching between field data measured and delineated crowns in mixed stands was 126.0% due to an overestimation of delineated crowns. This resulted in 42 *Pinus pinea* L. trees and 76 *Pinus pinaster* Ait. trees from the pure forest plots to be used for training the model (Table 1), while 82 *Pinus pinea* L. (36 from pure plots and 46 from mixed plots) and 71 *Pinus pinaster* Ait. (36 from pure plots and 35 from mixed stands) were selected for the validation process (Table 2).

2.4.3 Pleiades data processing

Firstly, 0.5-m high-resolution multispectral images were created from the pansharpening algorithm of high-resolution panchromatic and lower resolution multispectral imagery. Based on the previous individual crown delineation, variables derived from

Pleiades images were estimated for each object. Pleiades reflectance information was assigned to crown delineation with the mean values of all pixels within each crown, estimating blue band value, green band value, red band value, and infrared band value. Data from each band was assigned to the generated objects from the crown delineation using the QGIS Zonal Statistics Plugin.

In addition, different image bands were used to calculate two spectral indices: NDVI Normalized Difference Vegetation Index (NDVI) (Rouse et al. 1973) and EVI Enhanced Vegetation Index (EVI) (Liu and Huete 1995). NDVI was selected as it is the most commonly used vegetation indicator, capable of assessing the quantity, quality, and developmental stage of vegetation. It was calculated using Pleiades bands following the equation: $[(\text{Infrared} - \text{Red})/(\text{Infrared} + \text{Red})]$, where infrared and red refers to the recorded Pleiades values in the infrared and red bands, respectively. EVI is an optimized vegetation index designed to enhance the vegetation signal with improved sensitivity which reduces the adverse effects of environmental factors such as atmospheric conditions and soil background. This was calculated using the equation: $[2.5 \times (\text{Infrared} - \text{Red})/(\text{Infrared} + 2.4 \times \text{Red} + 1)]$. Statistics for the training and validation dataset are also shown in Tables 1 and 2. The six variables from satellite data that were finally available for modeling were blue band, green band, red band, infrared band, NDVI index, and EVI index.

Table 1 Descriptive statistics of the training dataset. Pp_p: *Pinus pinea* L. individual trees in pure stand plots, Pt_p: *Pinus pinaster* Ait. individual trees in pure stand plots

	Minimum		Maximum		Mean		Standard deviation	
	Pp_p	Pt_p	Pp_p	Pt_p	Pp_p	Pt_p	Pp_p	Pt_p
Field data								
DBH (cm)	17.7	28.3	63.7	48.0	35.9	36.6	10.7	5.2
Height (m)	6.7	11.7	15.4	18.8	10.7	15.0	2.5	1.4
Live_Branch (m)	2.0	0.0	9.0	12.7	4.9	7.8	2.0	2.5
Crown_Radius (m)	1.7	0.9	5.8	3.9	3.3	2.5	0.9	0.6
LiDAR data								
Max_Height (m)	6.4	9.3	14.4	16.1	9.9	13.5	2.4	1.5
Min_Height (m)	3.0	3.0	11.5	12.0	6.3	8.0	2.1	2.2
Crown_Area (m ²)	16.9	19.5	80.1	100.8	34.9	30.0	17.7	12.7
Crown_Height (m)	0.8	0.9	10.2	11.2	3.7	5.5	1.7	2.1
Density (trees/ha)	55.0	63.0	237.0	216.0	114.3	107.5	42.7	24.1
HR	0.12	0.08	0.77	0.78	0.37	0.41	0.13	0.15
AHR	1.69	1.34	7.20	7.04	3.46	2.23	1.34	0.88
CHR	−0.86	−0.86	−0.41	−0.26	−0.67	−0.51	0.10	0.13
Satellite data								
Blue	96.4	93.2	123.2	129.4	107.6	107.6	7.0	6.9
Green	120.6	116.2	146.0	141.0	130.4	127.2	5.2	5.0
Red	135.4	133.6	155.1	156.7	143.6	143.2	4.0	4.3
Infrared	246.6	220.3	306.4	278.1	269.2	249.1	13.3	13.7
NDVI	0.26	0.21	0.35	0.31	0.30	0.27	0.02	0.02
EVI	0.42	0.34	0.59	0.53	0.51	0.45	0.04	0.05

2.5 Classification technique

Random Forest classifier (Breiman et al. 2001) is a machine learning methodology based on independent decision trees providing diverse ways to explore numerically and graphically complex relationships, improving the accuracy of the model prediction (Valbuena et al. 2016b; Vega Isuhuaylas et al. 2018). It averages several decision trees, so there is a significantly lower risk of overfitting and it also reduces the chance of stumbling across a classifier that does not perform well. A large number of classification trees are produced from a random subset of training data, with permutations introduced at each node, selecting the most common classification result. In this study, Random Forest was used to classify species based on LiDAR and satellite data at individual tree level.

R software (R Core Team 2016) was used to generate the model, using the Random Forest library as is common practice in forest science (Maschler et al. 2018; Shi et al. 2018). Three parameters are needed to optimize the model: (i) *mtry*, number of variables randomly sampled as candidates at each split; (ii) *ntree*, number of bootstrap replicates; and (iii) *nodesize*, the minimum size of terminal nodes. To identify those parameter values leading to the best species discrimination, the model was optimized based on out-of-bag error estimates (OOB), defined as the rate of classification error estimated for the different subsampling sets from the training set used to train

the model classifier. In addition, the Random Forest algorithm can estimate the importance of the variable in every model, showing the mean decrease of accuracy, which is defined as the loss of accuracy measured by the OOB-error when leaving out a variable (Breiman et al. 2001). Higher values indicate higher importance of each variable for the classification.

Three training models were developed. First, the “LiDAR model” was constructed only using LiDAR variables, then the “Spectral model” was built including only satellite variables, and finally, all available variables were incorporated into the so-called “Complete model”. Each of them was designed using the same approach, attempting to identify the best combination of the three Random Forest parameters and they were evaluated based on out-of-bag error estimation. The training model which provided the greatest predictive capacity was then validated in terms of classification accuracy on the three independent validation datasets: (i) individual trees in pure *Pinus pinea* L. stand plots, (ii) individual trees in pure *Pinus pinaster* Ait. stand plots, and (iii) individual trees in *Pinus pinea* L. and *Pinus pinaster* Ait. mixed stand plots. Validation was done by (i) building confusion matrices which show the relationship between the false positive fraction and true negative fraction (or vice versa), (ii) estimating overall accuracy (OA) as $(\text{number of correctly classified } Pinus \text{ pinea L.} + \text{number of correctly classified } Pinus \text{ pinaster Ait.}) / \text{total number of trees used in the validation}$, and (iii) calculating

Table 2 Descriptive statistics of validation dataset. Pp_p: *Pinus pinea* L. individual trees in pure stand plots, Pt_p: *Pinus pinaster* Ait. individual trees in pure stand plots, Pp_m: *Pinus pinea* L. individual trees in mixed stand plots, Pt_m: *Pinus pinaster* Ait. individual trees in mixed stand plots

	Minimum				Maximum				Mean				Standard deviation			
	Pp_p	Pt_p	Pp_m	Pt_m	Pp_p	Pt_p	Pp_m	Pt_m	Pp_p	Pt_p	Pp_m	Pt_m	Pp_p	Pt_p	Pp_m	Pt_m
Field data																
DBH (cm)	34.6	27.2	26.9	22.5	71.0	58.6	57.2	52.6	44.6	41.0	40.0	32.7	7.1	6.9	7.5	6.2
Height (m)	12.4	11.5	9.6	11.0	19.9	21.6	16.7	16.9	15.0	16.2	12.6	13.7	1.7	2.8	1.6	1.6
Live_Branch (m)	3.8	2.6	2.4	2.7	9.8	14.2	9.1	11.3	6.4	8.6	5.4	7.9	1.7	3.6	1.5	1.9
Crown_Radius (m)	2.6	1.8	2.1	1.4	6.0	3.7	5.9	3.7	3.9	2.7	3.8	2.5	0.7	0.5	0.8	0.6
LiDAR data																
Max_Height (m)	11.1	7.7	6.8	10.0	18.4	18.0	16.0	15.6	14.0	13.7	12.0	12.7	1.7	2.8	1.6	1.3
Min_Height (m)	2.1	2.2	3.1	4.2	13.3	13.8	12.2	10.8	8.4	7.8	7.4	7.4	2.6	2.5	1.7	1.7
Crown_Area (m ²)	23.5	12.5	19.5	19.5	87.7	75.9	105.2	107.0	42.7	26.0	38.6	36.2	16.9	14.9	18.5	17.7
Crown_Height (m)	2.3	1.2	0.5	1.6	11.9	14.9	11.2	10.3	5.5	5.9	4.6	5.3	2.1	3.1	1.7	1.8
Density (trees/ha)	41.0	60.0	48.0	68.0	227.0	179.0	206.0	197.0	106.6	113.8	108.5	122.7	44.6	28.9	36.4	42.0
HR	0.2	0.1	0.08	0.15	0.8	0.9	0.78	0.70	0.4	0.4	0.38	0.41	0.2	0.2	0.13	0.13
AHR	1.7	0.9	1.53	1.53	6.1	5.8	9.01	9.62	3.0	1.9	3.23	2.87	1.0	1.0	1.53	1.51
CHR	-5.1	-4.8	0.22	0.30	-0.7	0.1	0.92	0.85	-2.0	-0.9	0.62	0.59	1.0	1.0	0.13	0.13
Satellite data																
Blue	65.6	54.5	90.6	92.0	87.9	99.4	120.5	113.6	74.7	72.3	102.2	103.0	5.1	11.6	18.7	6.7
Green	86.3	74.4	113.5	117.0	107.3	115.2	133.5	131.5	96.8	93.4	123.7	124.8	4.7	10.5	17.4	4.1
Red	102.1	96.7	132.1	132.4	121.5	135.5	148.4	147.1	111.2	109.7	139.8	140.3	4.0	9.8	17.0	3.8
Infrared	199.3	136.9	201.1	228.9	270.4	269.9	287.5	282.7	238.1	208.1	253.0	252.5	17.2	34.7	29.6	15.7
NDVI	0.27	0.17	0.20	0.22	0.43	0.38	0.36	0.34	0.36	0.30	0.29	0.28	0.03	0.05	0.04	0.03
EVI	0.45	0.27	0.32	0.36	0.76	0.65	0.61	0.59	0.63	0.51	0.48	0.47	0.06	0.09	0.08	0.06

receiver operating characteristic (ROC) area under the curve (AUC) in order to quantify the uncertainty in models' prediction (Zipkin et al. 2012). ROC curve is a plot of the sensitivity of a diagnostic test against one minus its specificity, as the cut-off criterion for indicating a positive test is varied (Everitt 2006). As discussed in Hosmer and Lemeshow (2000), the quantitative-qualitative relationship between the AUC and prediction accuracy can be classified as follows: less than 50% suggests no model discrimination, 50–70% is considered as poor discrimination, 70–80% is considered as acceptable discrimination, and more than 80% suggests excellent discrimination.

3 Results

3.1 Training model selection

To achieve the best discrimination between the two pine species, *Pinus pinea* L. and *Pinus pinaster* Ait., three training models were developed based on data from individual trees in pure stand plots. The “Spectral model” is less accurate than the “LiDAR model” (Table 3), so species discrimination based on the latter is better. However, the “Complete model”,

which combines LiDAR and satellite variables, provided the lowest out-of-bag error estimates.

Thus, the model selected was the one combining LiDAR and Pleiades data since it provides the greatest accuracy. The “Complete model” possesses a high degree of predictive and explanatory power, correctly classifying 93.22% of cases (i.e., OOB estimate of error rate was 6.78%). In other words, a high proportion of *Pinus pinea* L. and *Pinus pinaster* Ait. trees were correctly identified. Figure 3 shows the relative importance of each variable in the selected model, as provided by Random Forest, revealing which variables from LiDAR and satellite data are the most significant in discriminating between species. The most important variables defining the characterization of the composition are Crown_Height, AHR (area and height relation), NDVI index, EVI index, infrared band,

Table 3 Accuracy of the three training models in classifying *Pinus pinea* L. and *Pinus pinaster* Ait. individual trees in pure stands

Training models	No. of variables	OOB %
LiDAR model	6	13.56
Spectral model	6	20.34
Complete model	12	6.78

Fig. 3 Mean estimated variable importance in the Random Forest “Complete model”. The maximum value corresponds to the main discriminating variable and the rest are presented in relation to this score

Table 4 Confusion matrix obtained from the validation of pure datasets

		Predicted		Total	User's accuracy (%)
Observed		<i>Pinus pinea</i> L.	<i>Pinus pinaster</i> Ait.		
	<i>Pinus pinea</i> L.	26	10	36	72.2
	<i>Pinus pinaster</i> Ait.	2	34	36	94.4
	Total	28	44	72	–
Consumer's accuracy (%)		92.9	77.3	–	83.3

green band, Crown_Area, blue band, red band, Density, HR (heights relation), and CHR (crown and height relation). The first three variables are highly related to differences in crown and stem shape between the two species.

3.2 Validation model

The validation process was carried out on an independent sample of pure and mixed plots. Table 4 shows the confusion matrix obtained for the selected “Complete model” for the validation on pure datasets. In the case of *Pinus pinea* L. trees, the model correctly classified 72.2% of individuals. In contrast, the level of efficiency obtained in the case of *Pinus pinaster* Ait. was 94.4%. The overall accuracy of the model when used for plots in pure stands was 83.3%, a mid-precision value between the results observed in the training model and those from the validation performed on mixed datasets (Table 5). The results indicate that *Pinus pinea* L. is well discriminated although sub-estimated, because *Pinus pinea* L. are sometimes classified as *Pinus pinaster* Ait., whereas the opposite does not occur. Hence, the “Complete model” achieves a high-confidence classification of both *Pinus pinea* L. and *Pinus pinaster* Ait. trees. Also, the area under the curve (AUC) value also indicates that the model classifier is discriminating between both species with an average rate of 89.6% as shown in Fig. 4a. Thus, the model applied in pure stands showed excellent discrimination in its prediction.

Accuracy results for the selected model on the mixed dataset validation are presented in Table 5. The efficiency of the model in classifying the two species when compared with field data showed that 63.0% of individual trees were correctly classified. The model can be classified as of average accuracy. In the case of *Pinus pinaster* Ait., 74.1% were successfully classified while in *Pinus pinea* L. the percentage was 54.3%. The AUC value from mixed stands was 63.4% (Fig. 4b). The

model applied in mixed stands had lower discrimination power than in pure stands.

4 Discussion

LiDAR data provide a widely accepted tool for characterizing forest structure and attributes such as total biomass or volume (Lim et al. 2003; Næsset 2004; Miura and Jones 2010; Nguyen et al. 2016). In addition, LiDAR information has been extensively used to identify individual tree species (Holmgren and Persson 2004; Zhang et al. 2016; Dechesne et al. 2017). Similarly, satellite images have also been used for tree crown delineation (Wang 2010; Lin et al. 2011) and species classification (Leckie et al. 2005; Ballanti et al. 2016). Finally, the combined use of multispectral and LiDAR data has been evaluated as a tool for classifying individual trees (Holmgren et al. 2008; Suratno et al. 2009a).

Statistical analyses have been carried out applying the Random Forest algorithm, but it could have been done applying other statistical methods, such as linear discriminant analysis. In fact, linear discriminant analysis was carried out, but its results were similar to the ones obtained with the Random Forest algorithm. Because of this, it was decided to work with the Random Forest algorithm since in using this method several decision trees are averaged, resulting in a more robust analysis with lower overfitting risk.

As hypothesized, the Random Forest model which only includes LiDAR variables has greater discriminative power than the model with only spectral variables because the structural differences between the two pine species are more evident than their spectral singularities. However, the model combining information from both sources is even more precise. This finding is in accordance with those of Dalponte et al.

Table 5 Confusion matrix obtained from the validation of individual trees in mixed stand datasets

		Predicted		Total	User's accuracy (%)
Observed		<i>Pinus pinea</i> L.	<i>Pinus pinaster</i> Ait.		
	<i>Pinus pinea</i> L.	25	21	46	54.3
	<i>Pinus pinaster</i> Ait.	9	26	35	74.3
	Total	34	47	81	–
Consumer's accuracy (%)		73.5	55.3	–	63.0

Fig. 4 ROC curve of the pure and mixed dataset. **a** Pure stand dataset. **b** Mixed stand dataset. AUC area under the curve

(2012), who obtained higher accuracy using models with mixed spectral and LiDAR information than with models based on separate information sources. These authors performed the crown delineation manually and used spectral images and low as well as high-density LiDAR data to classify tree species in the Alps. They discriminated between six species, including evergreen species (*Abies alba* Mill., *Picea abies* (L.) H. Karst., *Pinus mugo* Turra, and *Pinus sylvestris* L.) and deciduous species (*Fagus sylvatica* L. and *Larix decidua* Mill.), and found that evergreen species displayed similar structural and spectral responses, with the exception of *Pinus mugo* Turra, the structure of which is clearly different. The response of deciduous species, however, is well discriminated, taking into account the absence of leaves during the winter season and the spectral intensity at the beginning of the vegetative development. The mean accuracy of the classification between evergreen species was 58%, slightly lower than the results for mixed stands obtained in this study due to the fact that the differences in structural variables between *Pinus pinea* L. and *Pinus pinaster* Ait. are greater than those of their spectral variables. Hence, structural variables have greater discriminative power than spectral variables in the classifier model.

Including all variables in the complete model is maybe sub-optimal due to some of them probably being redundant. In addition, machine learning algorithms like Random Forest are able to manage correlations between variables without any normalization process and the relationship between them need not be linear. Then, the importance of structural variables in discriminating between species is related to the wide differences in individual allometry observed between them. *Pinus pinea* L. usually shows a flattened, short and wide umbrella-like crown, with a long clean stem. On the contrary, *Pinus pinaster* Ait. in pure stands usually presents an apical crown-shape, with lower crown base height (Live_Branch), covering the ground from sight. Differences in crown height

(Max_Height – Min_Height) reflect the importance of this structural variable that the model considers as the most important in these mature stands. Crown height (Max_Height – Min_Height) is usually greater in *Pinus pinaster* Ait. than in *Pinus pinea* L. The second variable in importance is area and height relation (Crown_Area/Max_Height) which reflects the weight of the structural characteristics involved in its formulation for each species. Regarding the relationship between the crown area (Crown_Area) and area and height relation (Crown_Area/Max_Height), high values are indicative of *Pinus pinea* L. The third LiDAR variable of importance is crown area (Crown_Area), although it is moderately important. It is worth mentioning that this variable is not density-dependent because this type of stands shows very low canopy cover with very little crown overlapping. In addition, the spectral index variables—NDVI, EVI, and infrared band—selected for the model give higher values in *Pinus pinea* L. than *Pinus pinaster* Ait. stands. This may be linked with both the greater plasticity of *Pinus pinea* L., often displacing *Pinus pinaster* Ait. (Bravo-Oviedo et al. 2010), and the generalized decay (associated with decoloration and defoliation) observed in *Pinus pinaster* Ait. forests within the region (Prieto-Recio et al. 2015).

The existence of mixed stands of *Pinus pinaster* Ait. and *Pinus pinea* L. trees indicates that both species have a similar ecological strategy and species mixture may be desirable in terms of increasing and diversifying productivity. However, the conditions currently present in the stands are likely to lead to future dominance of *Pinus pinea* L. over *Pinus pinaster* Ait. given the greater competition tolerance of the former (Ledo et al. 2014). Both species' spectral response in mixed stands could be more similar than in pure stands. Thus, this uniformity might reflect the fact that living together, both pines become more alike, indicating their ecological competition. Furthermore, tree allometry depends on whether a tree grows in a pure or a mixed stand (Forrester et al. 2017). However,

structure crown changes in mixed stands could not be attributed only to species composition. Differences of density and canopy structures, which in part can be due to interspecific interactions, might modify structural characteristics (Riofrío et al. 2017).

As expected, the training model selected in this study achieves greater accuracy in the pure stand validation set due to the fact that it was only trained with pure stands of a single species, as shown in Tables 1 and 2. Therefore, the average accuracy when validating on pure stands is higher than on mixed stands. In addition, the degree of accuracy is greater for *Pinus pinaster* Ait. classification. Nevertheless, in pure stands, when the model classifies a tree as *Pinus pinea* L. it is usually correct (only 5.6% of *Pinus pinaster* Ait. trees are classified as *Pinus pinea* L.). In the case of *Pinus pinaster* Ait., the failure rate is higher, with of trees 27.8% being misclassified. However, in mixed stands, *Pinus pinea* L. trees are underestimated while *Pinus pinaster* Ait. trees are overestimated by around 34%. This may be due to the wider range of variability in the allometric variables of *Pinus pinea* L. individuals growing in mixed stands while *Pinus pinaster* Ait. individuals tend to maintain their pyramidal crown structure. A decrease in accuracy between training and validation models is logical, since the validation sample is an independent data set from the one used to train the model, and this validation set includes plots from mixed stands, where the reduction in accuracy is expected to be higher, since tree allometry attributes are different in these types of stands. These differences may be improved in future research by using hyperspectral images and higher resolution LiDAR data (Holmgren et al. 2008; Dalponte et al. 2012). However, this data is not often available. Aerial photographs could be useful, but they are not always taken in the infrared channel, and they are also extremely dependent on the orography. From a NWFP management planning point of view, the maximum interest is to know the percentage of each species correctly classified at the stand level. Thus, in mixed stands, classification errors for one species could be compensated with reversed classification errors for the other species.

The main strength of this study is the application of an automatic crown delineation tool together with satellite-based continuous spatial information and open source low-density LiDAR data for distinguishing two very similar species at tree level in a mixed conifer forest. The low density of the point cloud LiDAR data is not a limiting issue in these open stands involving wide tree crowns, thus it is highly likely that the LiDAR pulse hit them a few times. Also, considering that training data was collected in pure stands where no classification mistake is expected and then validated in pure and mixed stands, the results are still good. With all these, the tool could help avoid doing stem by stem traditional inventory. The main limitation may be the Pleiades images, which currently cannot easily be replaced by open source multispectral satellite images, given the lower spatial resolution of the latter

in comparison to Pleiades images. Working with a high level of detail, such as individual tree level, requires a great degree of spatial assessment between the different sources of information, in order to prevent combining information from inside and outside the crown delineation.

5 Conclusion

The species classification tool developed, together with an algorithm for automatic crown delineation, allows the discrimination of two very similar species in mixed Mediterranean conifer forests using continuous spatial information at the earth's surface: Pleiades images and open source LiDAR data. The method presented here should be valid for regional or landscape-scale assessment of resources computed at tree-level (such as NWFP), enhancing the economic value of the forest products and providing support for forest management decision-making.

Acknowledgments The authors wish to thank the Forest Service of Valladolid province for their continuous help in plot installation, maintenance, and data collection. Similarly, the authors wish to express their gratitude to the PNT program for providing us with Pleiades images and to Dr. Fernando Pérez-Cabello for his advice with regard to interpreting the Pleiades images. Also, the authors wish to thank Dr. Iñigo Lizarralde, Dr. Rafael Alonso, and Dra. Beatriz Águeda for their general assistance and to Adam Collins for the English language advice.

Funding Research of Ángela Blázquez-Casado was funded by a contract of Ministerio de Economía, Industria y Competitividad, Spanish Government (DI-14-06953). This study has also been financed through the project AGL2014-51964 FORMIXING (Ministerio de Economía, Industria y Competitividad, Spanish Government) and the CC16-095 PROPINEA agreement between INIA, ITACYL, and the Diputation of Valladolid.

Data availability A dataset generated with the coordinates and the species of each tree measured is available in FigShare repository (Blázquez-Casado et al. 2019) at <https://doi.org/10.6084/m9.figshare.7951166.v2>

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alonso FJG, González GM, Sánchez LG (2012) La problemática de la regeneración natural de los pinares en los arenales de la Meseta Castellana. In La regeneración natural de los pinares en los arenales de la Meseta Castellana. Instituto Universitario de Investigación en Gestión Forestal Sostenible, pp 11–16
- Arias-Rodil M, Diéguez-Aranda U, Álvarez-González JG, Pérez-Cruzado C, Castedo-Dorado F, González-Ferreiro E (2018) Modeling diameter distributions in radiata pine plantations in

- Spain with existing countrywide LiDAR data. *Ann For Sci* 75:36. <https://doi.org/10.1007/s13595-018-0712-z>
- Ballanti L, Blesius L, Hines E, Kruse B (2016) Tree species classification using hyperspectral imagery: a comparison of two classifiers. *Remote Sens* 8:1–18. <https://doi.org/10.3390/rs8060445>
- Barba J, Curiel Yuste J, Martínez-Vilalta J, Lloret F (2013) Drought-induced tree species replacement is reflected in the spatial variability of soil respiration in a mixed Mediterranean forest. *For Ecol Manag* 306:79–87. <https://doi.org/10.1016/j.foreco.2013.06.025>
- Barrett F, McRoberts RE, Tomppo E et al (2016) A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. *Remote Sens Environ* 174:279–289. <https://doi.org/10.1016/j.rse.2015.08.029>
- Beguet B, Chehata N, Boukir S, et al (2014) Classification of forest structure using very high resolution Pleiades image texture. *Geosci Remote Sens Symp (IGARSS)*, 2014 IEEE Int 2324–2327. doi: <https://doi.org/10.1109/IGARSS.2014.6946936>
- Blázquez-Casado A, Calama R, Valbuena M, Rodríguez F (2019) Trees.V2. FigShare. [Dataset]. <https://doi.org/10.6084/m9.figshare.7951166.v2>
- Bravo-Oviedo A, Gallardo-Andrés C, del Río M, Montero G (2010) Regional changes of *Pinus pinaster* site index in Spain using a climate-based dominant height model. *Can J For Res* 40:2036–2048. <https://doi.org/10.1139/X10-143>
- Breiman L, Friedman J, Olshen R, Stone C (2001) Classification and regression trees. Chapman Hall, New York
- Calama R, Gordo FJ, Mutke S, Montero G (2008) An empirical ecological-type model for predicting stone pine (*Pinus pinea* L.) cone production in the Northern Plateau (Spain). *For Ecol Manag* 255:660–673. <https://doi.org/10.1016/j.foreco.2007.09.079>
- Calama R, Gordo J, Madrigal G, Mutke S, Conde M, Montero G, Pardos M (2016) Enhanced tools for predicting annual stone pine (*Pinus pinea* L.) cone production at tree and forest scale in inner Spain. *For Syst* 25. <https://doi.org/10.5424/fs/2016253-09671>
- Castaño-Díaz M, Álvarez-Álvarez P, Tobin B, Nieuwenhuis M, Afif-Khouri E, Cámara-Obregón A (2017) 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. *Ann For Sci* 74:69. <https://doi.org/10.1007/s13595-017-0665-7>
- Dalponte M, Bruzzone L, Gianelle D (2012) Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sens Environ* 123:258–270. <https://doi.org/10.1016/j.rse.2012.03.013>
- Dechesne C, Mallet C, Le Bris A, Gouet-Brunet V (2017) Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery. *ISPRS J Photogramm Remote Sens* 126:129–145. <https://doi.org/10.1016/j.isprsjprs.2017.02.011>
- Deng S, Katoh M, Yu X, Hyyppä J, Gao T (2016) Comparison of tree species classifications at the individual tree level by combining ALS data and RGB images using different algorithms. *Remote Sens* 8. <https://doi.org/10.3390/rs8121034>
- Dubayah RO, Drake JB (2000) Lidar remote sensing for forestry applications. *J For* 98:44–46
- European Environmental Agency (2006) European forest types: categories and types for sustainable forest management reporting policy, p 114
- Everitt BS (2006) The Cambridge dictionary of statistics (3rd ed). Cambridge University Press, Cambridge
- Fassnacht FE, Latifi H, Stereńczak K, Modzelewska A, Lefsky M, Waser LT, Straub C, Ghosh A (2016) Review of studies on tree species classification from remotely sensed data. *Remote Sens Environ* 186: 64–87. <https://doi.org/10.1016/j.rse.2016.08.013>
- Forrester DI, Benneter A, Bouriaud O, Bauhus J (2017) Diversity and competition influence tree allometric relationships – developing functions for mixed-species forests. *J Ecol* 105:761–774. <https://doi.org/10.1111/1365-2745.12704>
- Hollaus M, Dorigo W, Wagner W, Schadauer K, Höfle B, Maier B (2009) Operational wide-area stem volume estimation based on airborne laser scanning and national forest inventory data. *Int J Remote Sens* 30:5159–5175. <https://doi.org/10.1080/01431160903022894>
- Holmgren J, Persson Å (2004) Identifying species of individual trees using airborne laser scanner. *Remote Sens Environ* 90:415–423. [https://doi.org/10.1016/S0034-4257\(03\)00140-8](https://doi.org/10.1016/S0034-4257(03)00140-8)
- Holmgren J, Persson Å, Söderman U (2008) Species identification of individual trees by combining high resolution LiDAR data with multi-spectral images. *Int J Remote Sens* 29:1537–1552. <https://doi.org/10.1080/01431160701736471>
- Hosmer DW, Lemeshow S (2000) Applied logistic regression, 2nd edn. John Wiley & Sons, Inc, New York
- Korpela I, Ole Ørka H, Maltamo M et al (2010) Tree species classification using airborne LiDAR - effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fenn* 44:319–339. <https://doi.org/10.14214/sf.156>
- Leckie DG, Tinis S, Nelson T, Burnett C, Gougeon FA, Cloney E, Paradine D (2005) Issues in species classification of trees in old growth conifer stands. *Can J Remote Sens* 31:175–190. <https://doi.org/10.5589/m05-004>
- Ledo A, Cañellas I, Barbeito I, Gordo FJ, Calama RA, Gea-Izquierdo G (2014) Species coexistence in a mixed Mediterranean pine forest: spatio-temporal variability in trade-offs between facilitation and competition. *For Ecol Manag* 322:89–97. <https://doi.org/10.1016/j.foreco.2014.02.038>
- Lee J, Cai X, Lellmann J, Dalponte M, Malhi Y, Butt N et al (2016) Individual tree species classification from airborne multi-sensor imagery using robust PCA. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(6):2554–2567. <https://doi.org/10.1109/JSTARS.2016.2569408>
- Lim K, Treitz P, Wulder M, St-Onge B, Flood M (2003) LiDAR remote sensing of forest structure. *Prog Phys Geogr* 27:88–106. <https://doi.org/10.1191/0309133303pp360ra>
- Lin C, Lo CS, Thomson G (2011) A textural modification of the MMAC algorithm for individual tree delineation in forest stand using aerial bitmap images. *Proc - 4th Int Congr Image Signal Process CISP* 2011 3:1604–1608. <https://doi.org/10.1109/CISP.2011.6100501>
- Liu HQ, Huete A (1995) A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans Geosci Remote Sens* 33:457–465
- López C, Espinosa J, Bengoa J (2009) Mapa de Vegetación de Castilla y León. Síntesis, 1(400.000)
- Lopez-García P (1980) Estudio de semillas prehistóricas en algunos yacimientos españoles. *Trab Prehist* 37:419–432
- Maack J, Kattenbom T, Fassnacht FE, Enßle F, Hernández J, Corvalán P, Koch B (2015) Modeling forest biomass using very-high-resolution data - combining textural, spectral and photogrammetric predictors derived from spaceborne stereo images. *Eur J Remote Sens* 48:245–261. <https://doi.org/10.5721/EuJRS20154814>
- MAGRAMA (2012) Criteria and indicators for sustainable forest Management, in Spanish forests 2012. Ministerio de Agricultura, Alimentación y Medio Ambiente. NIPO:280-12-230-1, Madrid, España
- Maschler J, Atzberger C, Immitzer M (2018) Individual tree crown segmentation and classification of 13 tree species using airborne hyperspectral data. *Remote Sens* 10:1218. <https://doi.org/10.3390/rs10081218>
- Miura N, Jones SD (2010) Characterizing forest ecological structure using pulse types and heights of airborne laser scanning. *Remote Sens Environ* 114:1069–1076. <https://doi.org/10.1016/j.rse.2009.12.017>
- Moreno-Fernández D, Cañellas I, Calama R, Gordo J, Sánchez-González M (2013) Thinning increases cone production of stone pine (*Pinus*

- pineae* L.) stands in the Northern Plateau (Spain). *Ann For Sci* 70: 761–768. <https://doi.org/10.1007/s13595-013-0319-3>
- Næsset E (2004) Effects of different flying altitudes on biophysical stand properties estimated from canopy height and density measured with a small-footprint airborne scanning laser. *Remote Sens Environ* 91: 243–255. <https://doi.org/10.1016/j.rse.2004.03.009>
- Nanos N, Tadesse W, Montero G, Gil L, Alia R (2000) Modelling resin production distributions for *Pinus Pinaster* Ait. using two probability functions. *Ann For Sci* 57:379–377. <https://doi.org/10.1051/forest:2000128>
- Nanos N, Tadesse W, Montero G, Gil L, Alia R (2001) Spatial stochastic modeling of resin yield from pine stands. *Can J For Res* 31:1140–1147. <https://doi.org/10.1139/cjfr-31-7-1140>
- Nanos N, Calama R, Montero G, Gil L (2004) Geostatistical prediction of height/diameter models. *For Ecol Manag* 195:221–235. <https://doi.org/10.1016/j.foreco.2004.02.031>
- Nguyen HT, Hutya LR, Hardiman BS, Raciti SM (2016) Characterizing forest structure variations across an intact tropical peat dome using field samplings and airborne LiDAR. *Ecol Appl* 26:587–601. <https://doi.org/10.1890/15-0017>
- Ørka HO, Gobakken T, Næsset E, Ene L, Lien V (2012) Simultaneously acquired airborne laser scanning and multispectral imagery for individual tree species identification. *Can J Remote Sens* 38:125–138. <https://doi.org/10.5589/m12-021>
- Popescu SC (2007) Estimating biomass of individual pine trees using airborne lidar. *Biomass Bioenergy* 31:646–655. <https://doi.org/10.1016/j.biombioe.2007.06.022>
- Popescu SC, Wynne RH, Nelson RF (2003) Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Can J Remote Sens* 29:564–577. <https://doi.org/10.5589/m03-027>
- Prieto-Recio C, Martín-García J, Bravo F, Díez JJ (2015) Unravelling the associations between climate, soil properties and forest management in *Pinus pinaster* decline in the Iberian Peninsula. *For Ecol Manag* 356:74–83. <https://doi.org/10.1016/j.foreco.2015.07.033>
- Prodan M (1968) *Forest bimetries*. Elsevier, Pergamon Press, Oxford
- Quantum GIS Development Team (2017) QGIS geographic information system. Open Source Geospatial Foundation Project
- R Development Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Riofrio J, del Río M, Pretzsch H, Bravo F (2017) Changes in structural heterogeneity and stand productivity by mixing Scot pine and maritime pine. *For Ecol Manag InProcess* 40:38–48. <https://doi.org/10.1016/j.ibmb.2009.12.007>
- Rodríguez-García A, Martín JA, López R, Mutke S, Pinillos F, Gil L (2015) Influence of climate variables on resin yield and secretory structures in tapped *Pinus pinaster* Ait. in central Spain. *Agric For Meteorol* 202:83–93. <https://doi.org/10.1016/j.agrformet.2014.11.023>
- Rougier S, Anne P (2014) Improvements of urban vegetation segmentation and classification using multi - temporal Pleiades images. *South-Eastern Eur J Earth Obs Geomatics* 3:409–414
- Rouse JW, Hass RH, Schell JA, Deering DW (1973) Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resour Technol Satell Symp* 1:309–317 **doi: citeulike-article-id:12009708**
- Ruiz LÁ, Recio JA, Crespo-Peremarch P, Sapena M (2016) An object-based approach for mapping forest structural types based on low-density LiDAR and multispectral imagery. *Geocarto Int* 6049:1–15. <https://doi.org/10.1080/10106049.2016.1265595>
- Shi Y, Wang T, Skidmore AK, Heurich M (2018) Important LiDAR metrics for discriminating forest tree species in Central Europe. *ISPRS J Photogramm Remote Sens* 137:163–174. <https://doi.org/10.1016/j.isprsjprs.2018.02.002>
- Soliño M, Yu T, Alía R, Auñón F, Bravo-Oviedo A, Chambel MR, de Miguel J, del Río M, Justes A, Martínez-Jauregui M, Montero G, Mutke S, Ruiz-Peinado R, García del Barrio JM (2018) Resin-tapped pine forests in Spain: ecological diversity and economic valuation. *Sci Total Environ* 625:1146–1155. <https://doi.org/10.1016/j.scitotenv.2018.01.027>
- Suratno A, Seielstad C, Queen L (2009a) Tree species identification in mixed coniferous forest using airborne laser scanning. *ISPRS J Photogramm Remote Sens* 64:683–693. <https://doi.org/10.1016/j.isprsjprs.2009.07.001>
- Suratno A, Seielstad C, Queen L (2009b) Mapping tree species using LIDAR in mixed-coniferous forests. *Proc Silvilarer* 2009:1–10
- Tinel C, Grizonnet M, Fontannaz D, de Boissezon H, Giros A (2012) ORFEO, the Pleiades accompaniment program and its users thematic commissioning. In XXII ISPRS congress. Vol. 39, p. B3-572. <https://doi.org/10.5194/isprarchives-XXXIX-B3-569-2012>
- Tonolli S, Dalponte M, Neteler M, Rodeghiero M, Vescovo L, Gianelle D (2011) Fusion of airborne LiDAR and satellite multispectral data for the estimation of timber volume in the Southern Alps. *Remote Sens Environ* 115:2486–2498. <https://doi.org/10.1016/j.rse.2011.05.009>
- Valbuena M, Santamaría J, Sanz F (2016a) Estimation of diameter and height of individual trees for *Pinus sylvestris* L. based on the individualising of crowns using airborne LiDAR and the national forest inventory data. *For Syst* 25:1–11. <https://doi.org/10.5424/fs/2016251-05790>
- Valbuena R, Maltamo M, Packalen P (2016b) Classification of forest development stages from national low-density lidar datasets: a comparison of machine learning methods. *Rev Teledetección* (45):15–25. <https://doi.org/10.4995/raet.2016.4029>
- Vauhkonen J, Ørka HO, Holmgren J, Dalponte M, Heinzel J, Koch B (2014) Tree species recognition based on airborne laser scanning and complementary data source. In *Forestry applications of airborne laserscanning*. Springer, Dordrecht, pp 135–156. <https://doi.org/10.1007/978-94-017-8663-8>
- Vega I, Izuuaylas LA, Hirata Y, Santos LCV, Torobeo NS (2018) Natural forest mapping in the Andes (Peru): a comparison of the performance of machine-learning algorithms. *Remote Sens* 10. <https://doi.org/10.3390/rs10050782>
- Wang L (2010) A multi-scale approach for delineating individual tree crowns with very high resolution imagery. *Photogramm Eng Remote Sens* 76:371–378. <https://doi.org/10.14358/pers.76.4.371>
- Xie Y, Sha Z, Yu M (2008) Remote sensing imagery in vegetation mapping: a review. *J Plant Ecol* 1:9–23. <https://doi.org/10.1093/jpe/rtn005>
- Zhang Z, Kazakova A, Moskal LM, Styers DM (2016) Object-based tree species classification in urban ecosystems using LiDAR and hyperspectral data. *Forests* 7:1–16. <https://doi.org/10.3390/f7060122>
- Zipkin EF, Grant EHC, Fagan WF (2012) Evaluating the predictive abilities of community occupancy models using AUC while accounting for imperfect detection. *Ecol Appl* 22:1962–1972

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.