

Darío Domingo Ruiz

# Characteriation of Mediterranean Aleppo pine forest using low- density ALS data

Departamento  
Geografía y Ordenación del Territorio

Director/es  
De la Riva Fernández, Juan  
Lamela Gracia, María Teresa

<http://zaguan.unizar.es/collection/Tesis>





**Universidad**  
Zaragoza

Tesis Doctoral

CHARACTERIATION OF MEDITERRANEAN  
ALEPPO PINE FOREST USING LOW-DENSITY ALS  
DATA

Autor

Darío Domingo Ruiz

Director/es

De la Riva Fernández, Juan  
Lamela Gracia, María Teresa

**UNIVERSIDAD DE ZARAGOZA**

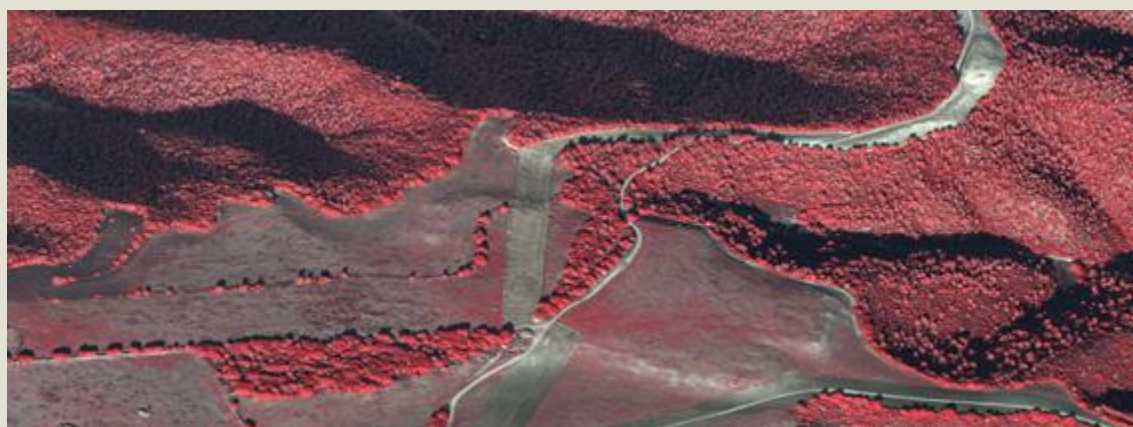
Geografía y Ordenación del Territorio

2019





# Characterization of Mediterranean Aleppo pine forest using low-density ALS data



**Darío Domingo Ruiz**

**Directores: M<sup>a</sup> Teresa Lamelas Gracia y Juan de la Riva Fernández**

PhD Thesis

Zaragoza 2019



Departamento de  
Geografía y  
Ordenación del Territorio  
**Universidad Zaragoza**

The author of this PhD Thesis was supported by Government of Spain, Department of Education Culture and Sports under Grant (FPU Grant BOE, 14/06250). Furthermore, the work was supported by FoResBiomALS research project from the Centro Universitario de la Defensa de Zaragoza (2017-09), SERGISAT research project (CGL2014-57013-C2-2-R) from the Spanish National Plan for Scientific and Technical Research and National Innovation Plan (Ministry of Economy and Competitiveness), and Geoforest-IUCA research group (Group S51\_17R, financed by FEDER 2014-2020 and Government of Aragon, “Construyendo Europa desde Aragón”).

Cover page image: RGB and NIR coloured point clouds (tile 664-4644) from PNOA ©INSTITUTO GEOGRÁFICO NACIONAL DE ESPAÑA - Autonomous Community of Aragón.

This PhD Thesis is developed as compendium of papers according to the Doctoral program in *Ordenación del Territorio y Medio Ambiente* at University of Zaragoza. The PhD student, Darío Domingo, is the first author and responsible for each and every article listed below. The references of the works that constitute the PhD Thesis body are the following:

1. **Domingo, D.**, Lamelas, M.T., Montealegre, A.L., de la Riva, J. 2017. Comparison of regression models to estimate biomass losses and CO<sub>2</sub> emissions using low-density airborne laser scanning data in a burnt Aleppo pine forest. *European Journal of Remote Sensing*, 50 (1), 384-396. doi: 10.1080/22797254.2017.1336067.
2. **Domingo, D.**, Lamelas, M.T., Montealegre, A.L., García-Martín, A., de la Riva, J. 2018. Estimation of total biomass in Aleppo pine forest stands applying parametric and nonparametric methods to low-density airborne laser scanning data. *Forests*, 9, 158-175. doi: 10.3390/f9030158.
3. **Domingo, D.**, Montealegre, A.L., Lamelas, M.T., García-Martín, A., de la Riva, J., Rodríguez, F, Alonso, R. 2019. Quantifying forest residual biomass in *Pinus halepensis* Miller stands using Airborne Laser Scanning data. *GIScience and Remote Sensing*, 56 (8), 1210-1232. doi: 10.1080/15481603.2019.1641653.
4. **Domingo, D.**, Alonso, R, Lamelas, M.T., Montealegre, A.L., Rodríguez, F, de la Riva, J. 2019. Temporal Transferability of Pine Forest Attributes Modeling Using Low-Density Airborne Laser Scanning Data. *Remote Sensing*, 11 (3), 261. doi:10.3390/rs11030261.

La presente tesis doctoral se ha elaborado como compendio de publicaciones siguiendo la modalidad ofrecida por el programa de Doctorado en Ordenación del Territorio y Medio Ambiente de la Universidad de Zaragoza. El doctorando, Darío Domingo, figura como primer autor y responsable de todos y cada uno de los artículos publicados. A continuación, se detallan las referencias de los trabajos que constituyen el cuerpo de la tesis:

1. **Domingo, D.**, Lamelas, M.T., Montealegre, A.L., de la Riva, J. 2017. Comparison of regression models to estimate biomass losses and CO<sub>2</sub> emissions using low-density airborne laser scanning data in a burnt Aleppo pine forest. *European Journal of Remote Sensing*, 50 (1), 384-396. doi: 10.1080/22797254.2017.1336067.
2. **Domingo, D.**, Lamelas, M.T., Montealegre, A.L., García-Martín, A., de la Riva, J. 2018. Estimation of total biomass in Aleppo pine forest stands applying parametric and nonparametric methods to low-density airborne laser scanning data. *Forests*, 9, 158-175. doi: 10.3390/f9030158.
3. **Domingo, D.**, Montealegre, A.L., Lamelas, M.T., García-Martín, A., de la Riva, J., Rodríguez, F, Alonso, R. 2019. Quantifying forest residual biomass in *Pinus halepensis* Miller stands using Airborne Laser Scanning data. *GIScience and Remote Sensing*, 56 (8), 1210-1232. doi: 10.1080/15481603.2019.1641653.
4. **Domingo, D.**, Alonso, R, Lamelas, M.T., Montealegre, A.L., Rodríguez, F, de la Riva, J. 2019. Temporal Transferability of Pine Forest Attributes Modeling Using Low-Density Airborne Laser Scanning Data. *Remote Sensing*, 11 (3), 261. doi: 10.3390/rs11030261.



## Agradecimientos

*“aprender a caminar es más fácil si alguien te tiende su brazo”*

En primer lugar, me gustaría expresar mi más sincero agradecimiento a mis directores de tesis, la Dra. María Teresa Lamelas Gracia y el Dr. Juan de la Riva Fernández por su apoyo incondicional y por su buen saber hacer que han permitido que la investigación desarrollada haya llegado a buen puerto. Gracias por haber confiado en mí desde el inicio, por haberme dado ideas y también por haberme dejado abordar las mías propias. Me habéis enseñado a crecer en lo profesional y en lo personal y estoy seguro de que sin vuestro tesón y aliento esto no hubiera sido posible.

También quiero agradecer a los coautores de las distintas publicaciones, el Dr. Antonio Luis Montealegre, el Dr. Alberto García Martín, el Dr. Rafael Alonso y el Dr. Francisco Rodríguez, en especial por sus valiosas aportaciones, por sus ideas y por todo el esfuerzo que han dedicado.

Asimismo quiero agradecer a todos los miembros del Grupo de Investigación GEOFOREST-IUCA, del Departamento de Geografía y Ordenación del Territorio y al Centro Universitario de la Defensa por acogerme y darme la oportunidad de formarme como doctorando, así como por proporcionarme los recursos económicos y materiales para poder desarrollar las investigaciones. En concreto, el Departamento citado es la casa donde inicié mi formación y ha sido un grato placer poder seguir formando parte de esta familia geográfica de inmejorable calidad humana.

Agradezco a los investigadores que me acogieron durante mis estancias en el extranjero por todo el apoyo mostrado, por los conocimientos que me enseñaron y por estar siempre pendientes de mí. Muchas gracias a Warren B. Cohen y a Yang Zhiqiang por acogerme en el Laboratory for Applications of Remote Sensing in Ecology (LARSE) del USDA Forest Service en Corvallis. Así mismo, muchas gracias a Erik Naesset, a Terje Gobakken y a Hans Ole Ørka por el trato magnífico, la dedicación y el buen saber hacer cuando me acogieron en Norwegian University of Life Sciences de Ås. No me puedo olvidar de Antoine y Nemo que hicieron la estancia en Corvallis mucho más amena y gratificante pese a estar lejos de casa. Tampoco puedo olvidarme de Marie-Claude, Ana, Ida, Lennart y Victor por todos los ratos de pingpong y las salidas al campo. También quiero agradecer a los revisores externos de la tesis, a Ole Martin Bollandsås y a Hooman Latifi por su disponibilidad y aportaciones al presente documento.

Mi más sentido agradecimiento a todos los miembros del proyecto SERGISAT con los que he pasado intensas jornadas de campo que te hacen crecer en lo personal y laboral, porque el campo une y de qué manera. Gracias a Maite, a Teresa, a Juan, a Paloma, a Alberto, a Pere, a Raúl a Marcos y a Demetrio.

Agradezco a todos los compañeros y doctorandos que me han alentado y ayudado durante este periodo. Daniel Borini, Adrián Jiménez, Antonio Montealegre, Xavier Garate, Estela Pérez, Olga Rosero, Daniel Ballarín, Daniel Mora, Aldo Arránz, Ricardo Badía, Samuel Esteban y Andrea Urgilez. Muchas gracias por todos y cada uno de esos cafés juntos.

Del mismo modo agradezco a Katalin Varga y a Antonio Montealegre por los buenos momentos que hemos compartido en diversos cursos y congresos. También quiero acordarme de Yago, hace 9 años me dijiste *“apúntate a Geografía que te gusta el campo y seguro que te va bien”*. Muchas gracias, no fallaste.

Quisiera mostrar mi gratitud hacia todos mis profesores del Departamento de Geografía y Ordenación del Territorio por formarme como geógrafo y sembrar en mí la fascinación por esta ciencia. Además, quiero agradecer profundamente a aquellos profesores que me han ayudado y permitido iniciarme en la labor docente dentro del Departamento y también fuera de él, todo un reto o un sueño que ha sido altamente gratificante.

Finalmente, mi más profundo agradecimiento a mi familia, concretamente a mis padres y a Bea. Por apoyarme incondicionalmente, por aguantarme todos los ratos en los que os contaba cosas totalmente *“estrambóticas”* para vosotros, por valorar lo que es una tesis y por sacrificaros con lo que ello conlleva, en especial cuando he estado fuera de casa varios meses. Por entenderme siempre, por vuestro amor y aliento para cumplir este sueño.

Mil gracias a todas y todos.

## Abstract

Forest ecosystems provide environmental and economic services of great importance to the society. The characterization of these environments has been traditionally accomplished with intense field work. In comparison, the application of remote sensing tools provides a greater overview over large spatial and temporal scales while minimizing costs. Although optical data and Synthetic Aperture Radar (SAR) allow estimating forest stand variables, the development of LiDAR sensors such as Airborne Laser Scanner (ALS) have improved three-dimensional characterization of forest structure. The availability of two ALS public data coverages for the Spanish territory, provided by the National Plan for Aerial Orthophotography (PNOA), opens new research opportunities to generate useful information for forest management.

This PhD Thesis used low-density ALS-PNOA data to estimate different forest variables, with support in fieldwork, in the Aleppo pine (*Pinus halepensis* Miller) forests of Aragón region. The addressed research is relevant mainly for two reasons: first, the examination of suitable methodologies and error sources in forest stand variables prediction at local (small area) and regional scales (large area), and second, the application of ALS data to the characterization of forest areas as a socio-economic reservoir.

This PhD Thesis is a compendium of four scientific papers, which sequentially answer the objectives established. Firstly, a comparative analysis of different parametric and non-parametric models was performed to estimate biomass losses and CO<sub>2</sub> emissions using low-density ALS and Landsat 8 data in a burnt Aleppo pine forest. Secondly, we assess the suitability of variable selection methods when estimating total biomass in Aleppo pine forest stands using low-density ALS data. In the third manuscript, the quantification and mapping of forest residual biomass in Aleppo pine forest of Aragón region and the assessment of the effect of ALS and environmental variables in model accuracy were accomplished. Finally, the temporal transferability of seven forest stands attributes modelling using multi-temporal ALS-PNOA data in Aleppo pine forest at regional scale was explored. In this case, the temporal transferability was assessed comparing two methodologies; the direct and indirect approach. The first one fits a model for one point in time and estimates the forest variable for another point in time. The indirect approach adjusts two models in different points in time to estimate the forest variables in two different dates.

The results derived from this research indicated that Spearman's rank and All Subset Selection are the most appropriate methods in the ALS metrics selection step commonly applied in modelling. The suitability of the regression methods depends on the sample size and complexity. Thus, multivariate linear regression outperformed non-parametric methods with small samples while support vector machine was the most accurate method with larger samples. Model accuracy increased with higher point density and canopy pulse penetration, while decreasing with wider scan angles. Furthermore, the presence of steep slopes and shrub reduced model performance. In the case of forest stand variables prediction using multi-temporal ALS data, although the indirect approach produced generally a higher precision, the direct approach provided similar results,

constituting a suitable alternative to reduce modelling time and fieldwork costs. The fusion of ALS and passive optical data have evidenced the suitability of this information for quantifying wildfire CO<sub>2</sub> emissions to atmosphere, constituting a good alternative when multi-temporal ALS data is not available. The estimation of forest inventory variables as well as different biomass fractions, such as total biomass and forest residual biomass, provided valuable information to characterize Mediterranean Aleppo pine forests and improve forest management.



## Resumen

Los espacios forestales son una fuente de servicios, tanto ambientales como económicos, de gran importancia para la sociedad. La caracterización de estos ambientes ha requerido tradicionalmente de un laborioso trabajo de campo. La aplicación de técnicas de teledetección ha proporcionado una visión más amplia a escala espacial y temporal, a la par que ha generado una reducción de los costes. La utilización de sensores óptico-pasivo multiespectrales y de sensores radar posibilita la estimación de parámetros forestales, si bien el desarrollo de sensores LiDAR, como el caso de los escáneres láser aeroportados (ALS), ha mejorado la caracterización tridimensional de la estructura de los bosques. La disponibilidad pública de dos coberturas LiDAR, generadas en el marco del Plan Nacional de Ortofotografía Aérea (PNOA), ha abierto nuevas líneas de investigación que permiten proporcionar información útil para la gestión forestal.

La presente tesis utiliza datos LiDAR aeroportados de baja densidad para estimar diversas variables forestales, con ayuda de trabajo de campo, en masas forestales de Pino carrasco (*Pinus halepensis* Miller) en Aragón. La investigación aborda dos cuestiones relevantes como son la exploración de las metodologías más adecuadas para estimar variables forestales considerando escalas locales y regionales, teniendo en cuenta las posibles fuentes de error en el modelado; y, además, analiza la potencialidad de los datos LiDAR del PNOA para el desarrollo de aplicaciones forestales que valoricen las áreas forestales como recursos socio-económicos.

La tesis se ha desarrollado según la modalidad de compendio de publicaciones, incluyendo cuatro trabajos que dan respuesta a los objetivos planteados. En primer lugar, se realiza un análisis comparativo de distintos modelos de regresión, paramétricos y no paramétricos, para estimar la pérdida de biomasa y las emisiones de CO<sub>2</sub> en un incendio, mediante la utilización de datos LiDAR-PNOA y datos ópticos del satélite Landsat 8. En segundo lugar, se explora la idoneidad de distintos métodos de selección de variables para estimar biomasa total en masas de Pino carrasco utilizando datos LiDAR de baja densidad. En tercer lugar, se cuantificó y cartografió la biomasa residual forestal en el conjunto de masas de Pino carrasco de Aragón y se evaluó el efecto de diversas características de la tecnología LiDAR y de las variables ambientales en la precisión de los modelos. Finalmente, se analiza la transferibilidad temporal de modelos para estimar a escala regional siete variables forestales, utilizando datos LiDAR-PNOA multi-temporales. A este respecto, se compararon dos enfoques que permiten analizar la transferibilidad temporal: en primer lugar, el método directo ajusta un modelo para un determinado punto en el tiempo y estima las variables forestales para otra fecha; por otra parte, el método indirecto ajusta dos modelos diferentes para cada momento en el tiempo, estimando las variables forestales en dos fechas distintas.

Los resultados obtenidos y las conclusiones derivadas de la investigación indican que la técnica basada en coeficientes de correlación de Spearman y el método de selección por todos los subconjuntos constituyen los métodos de selección de métricas LiDAR más apropiados para la modelización. El análisis de métodos de regresión para la estimación de variables forestales indicó

que su idoneidad variaba de acuerdo con el tamaño y complejidad de la muestra. El método de regresión lineal multivariante arrojó mejores resultados que los métodos no-paramétricos en el caso de muestras pequeñas. Por el contrario, el método Support Vector Machine produjo los mejores resultados con muestras grandes. El incremento de la densidad de puntos y de los valores de penetración de los pulsos LiDAR en el dosel, así como la presencia de ángulos de escaneo pequeños, incrementó la exactitud de los modelos. De forma similar, el incremento de la pendiente y la presencia de arbustos en el sotobosque implican una reducción en la exactitud de los modelos. En la estimación de variables forestales utilizando datos LiDAR multi-temporales, aunque la utilización del enfoque indirecto arrojó generalmente una mayor precisión en los modelos, se obtuvieron resultados similares con el enfoque directo, el cual constituye una alternativa óptima para reducir el tiempo de modelado y los costes de realización de trabajo de campo. La fusión de datos LiDAR y datos óptico-pasivos ha evidenciado la conveniencia de los métodos aplicados para cuantificar las emisiones de CO<sub>2</sub> a la atmósfera generadas por un incendio. Esta metodología constituye una alternativa adecuada cuando no existen datos multi-temporales LiDAR. La estimación de variables de inventario forestal, así como de diversas fracciones de biomasa, como la biomasa total y la biomasa residual forestal, proporciona información valiosa para caracterizar las masas forestales mediterráneas de Pino carrasco y mejorar la gestión forestal.

# Table of contents

---

|  |     |
|--|-----|
| 1. Introduction.....   | 1   |
| 1.1. Background .....  | 3   |
| 1.1.1. Airborne laser scanning .....   | 3   |
| 1.1.2. Forest variable estimation using remote sensing data.....   | 4   |
| 1.2. Importance and justification .....  | 9   |
| 1.3. Hypotheses and objectives.....  | 11  |
| 1.4. Structure .....   | 11  |
| 2. Study area, materials and methods.....  | 17  |
| 2.1. Study area.....   | 19  |
| 2.2. Materials and methods .....   | 22  |
| 2.2.1. Field inventory data.....   | 22  |
| 2.2.2. Estimation of forest stand variables .....  | 24  |
| 2.2.3. Inventories updating .....  | 26  |
| 2.2.4. ALS-PNOA data.....  | 28  |
| 2.2.5. ALS pre-processing and metric computation .....   | 34  |
| 2.2.6. Optical data and spectral indices.....  | 37  |
| 2.2.7. Wildfire biomass loss and CO <sub>2</sub> emission estimation.....  | 39  |
| 2.2.8. Modelling: variable selection, regression methods and validation.....   | 40  |
| 2.2.9. Temporal transferability .....  | 47  |
| 2.2.10. Variable influence assessment.....   | 48  |
| 2.2.11. Mapping of forest variables.....   | 50  |
| 3. Research contributions.....   | 53  |
| 3.1. Comparison of regression models to estimate biomass losses and CO <sub>2</sub> emissions using low density airborne laser scanning data in a burnt Aleppo pine forest ..... | 55  |
| 3.2. Estimation of total biomass in Aleppo pine forest stands applying parametric and nonparametric methods to low-density airborne laser scanning data.....                     | 71  |
| 3.3. Quantifying forest residual biomass in <i>Pinus halepensis</i> Miller stands using Airborne Laser Scanning data.....  | 91  |
| 3.4. Temporal Transferability of Pine Forest Attributes Modelling Using Low-Density Airborne Laser Scanning Data.....  | 117 |
| 4. Conclusions and future working lines .....  | 147 |
| 4.1. Main conclusions.....   | 149 |
| 4.2. Specific conclusions.....   | 150 |
| 4.3. Future working proposals.....   | 163 |
| References.....  | 165 |
| Appendix: paper metrics .....  | 183 |



# 1. Introduction

*This chapter describes the main concepts and the conceptual framework in which this PhD Thesis was developed. Firstly, ALS technology and its use for forestry purposes are described. Secondly, the use of ALS data for estimating forest stand variables, at local and regional scales, using different regression algorithms is presented. Then, the research justification, hypothesis and aims are defined. Finally, the chapter depicts the PhD Thesis structure, including the developed research items and their link to the different objectives that form a thematic unity.*



## 1.1. Background

### 1.1.1. Airborne laser scanning

The characterization and quantification of forest resources started in Europe during the late 18<sup>th</sup> century when society was concerned about wood availability, being the main source of fuel. The estimation of forestry metrics, especially volume, was performed by visual interpretation. The development of forest surveys, inventory tools, sampling and statistical methods and the advance in computers during 19<sup>th</sup> and 20<sup>th</sup> century yield great progress in forestry science. The growth of remote sensing tools such as aerial images, optical passive satellite images, and Synthetic Aperture Radar (SAR) data provided a greater overview of forest resources over large areas (Boyd & Danson, 2005). However, the expansion of Light Detection and Ranging (LiDAR) technology has improved the three-dimensional (3D) characterization of forest ecosystems being a suitable tool for forestry variables estimation (Zhao et al., 2018).

LiDAR technology measures the distance between a laser transmitter and an object or surface using a monochromatic beam of light, coherent and directional (Andersen et al., 2005). The origin of this technology started in the early 1960s when Theodore Harold Maiman developed the first ruby laser that emits powerful pulses of collimated red light. In the 1980s the profile LiDAR was used for forestry application (Aldred & Bonnor, 1985; Maclean & Krabill, 1986). Further growth occurred in the 90s with the generation of Digital Terrain Models (DTM) and forest inventory variables estimation (Lefsky et al., 1999; Means et al., 1999; Næsset, 1997). During the last two decades the use of LiDAR technology have exponentially grown, being developed diverse hardware, software and applications within the forestry topics.

According to the platform used, three main types of LiDAR technologies exist: terrestrial laser scanners (TLS), airborne laser scanners (ALS) and satellite laser scanners (SLS). ALS is one of the most widespread for forestry purposes (Maltamo et al., 2014). Topographic ALS sensors emit their own electromagnetic flux within the infrared wavelengths (900 to 1,064 nm). These wavelength denotes high reflectance values of vegetation and transmissivity of atmosphere (Lefsky et al., 2002). The basic information captured by ALS is denominated point cloud, referring to a dense set of x, y and z coordinates that register the object reflexions reached by the laser light.

ALS technology is usually classified in two main types, according to the way it measures distances between the sensor and the object reached by the laser beam: full-waveform systems and discrete return systems. Full-waveform systems completely register the reflected energy. The distance between transmitter and object is measured using phase difference between emitted signal and scattered radiation. Full-waveform sensors provide richer data than discrete return systems, while the processing is more demanding. Although several processing methodologies has been proposed such as voxelization, sometimes the wavelengths are converted into point clouds similar to the ones provided by discrete sensors but with higher number of returns per pulse. On the other hand, discrete return systems capture one to five returns per emitted laser pulse. The distance between the transmitter and the object is measured as a function of time. The generalized use of discrete

return sensors for forestry and topographic purposes may be explained by the higher implementation in the commercial sector (Shan & Toth, 2008).

ALS discrete-return systems have been widely used for estimating forest variables. Trees are porous objects from the laser pulse point of view. The laser beam can travel through the canopy and the system is able to capture up to five returns. The first return in a forested area may reach the top of the tree or canopy surface, the intermediate or low returns might be scattered by the tree branches and leaves or even the understory, while the last return might reach the terrain. This capability of ALS sensors provides a reliable 3D representation of forest structure. Canopy penetration pulse varies according to canopy closure, determining the points that reach the terrain. According to Chasmer et al. (2006b) only 50% of last returns in forested areas are backscattered by the terrain.

ALS sensors also capture, through a photodiode, the energy reflected by the objects. This information is denominated intensity, registered in 8 or 12 bits. The intensity refers to each laser echo with a footprint varying from 0.2 to 1.0 m in small footprint discrete returns ALS sensors (Andersen et al., 2005; Evans et al. 2009). Intensity values varies according to different parameters such as surface roughness and wetness, flight height, angle of incidence, instrumental characteristics, atmospheric conditions, between others. Consequently, the use of intensity values requires the normalization or calibration of the data in each acquisition. In this sense, although intensity data has been used for some applications, such as species classification (Korpela et al., 2010; Watt et al., 2007), it is still not broadly implemented for forestry purposes.

The main components of an ALS discrete sensor are the platform, the laser sensor, the Global Navigation Satellite System (GNSS), the Inertial Measurement Unit (IMU) and the data manager or computer with specific software. The laser scanner includes the laser pulse transmitter, the scanning mechanism and the receptor to record the distance to the objects. The laser scanner emits laser pulses, with a scanning frequency of up to 300 kHz, directed to the terrain surface. The scanning mechanism (oscillating mirror, rotating polygon, palmer scan, fibre scanner, etc.) draws specific scanning patterns to capture the terrain surface in each flight line (Vosselman & Maas, 2010). The scanning angle or sensor field of view (FOV) and the flight height determine the strip width. Accordingly, the overlap between strips modifies the flight time and accuracy (Evans et al., 2009).

The GNSS rover unit, located inside of the plane, collect the position from the GNSS satellites. The enhancement of GNSS position accuracy, performed at real time using reference stations or base differential GNSS at the ground level, provide a centimeter nominal accuracy. The IMU includes the Inertial Navigation System (INS), managing the pitch, roll and yaw of the plane (Baltsavias, 1999). The use of computers to manage the data collected by the GNSS and IMU with specific software allows providing the x, y and z coordinates for each return pulse along the flight acquisition, constituting the ALS point cloud (Baltsavias, 1999).

### **1.1.2. Forest variable estimation using remote sensing data**

ALS data have been proven as a suitable technique for mapping forest vertical and horizontal structure as well as to derive forestry variables (Zhao et al., 2018). The use of structural and



textural information derived from passive optical data have been explored to derive forestry parameters (Dube & Mutanga, 2015; Pfeifer et al., 2012). In this sense, the availability of wide temporal series have provided better results on forestry variables estimation when characterizing disturbance history (Cohen et al., 1996; Coops & Waring, 2001). However, the data captured by passive optical sensors tend to saturate under closed canopy conditions and in dense forests (Lu, 2006). The use of orthophotography to derive forestry parameters such as stand density, height or cover, between others, has also been explored, providing lower accuracies than active sensors (Campbell, 2006) such as SAR or LiDAR.

SAR systems allow characterizing forest structure at global scale. This technology uses different wavelengths from the microwave to provide information about leaves, branches and stems (Periasamy, 2018; Tanase et al., 2014). Although SAR data availability have increased and there exist recent advances in processing software, as for example the tools provided by the Copernicus program, difficulties still arise in estimating forestry variables in heterogeneous and dense forests (Hyde et al., 2006). The improvement of structure from motion (SfM) algorithms and photogrammetric techniques provides new insights in the 3D characterization of forest structure. In this sense, unmanned aircraft vehicles (UAV) have been proposed to estimate forestry variables (Giannetti et al., 2018; Kachamba et al., 2017; Puliti et al., 2017), characterize forest fuels (Fernández-Álvarez et al., 2019), detection of canopy gaps (Bagaram et al., 2018) or tree-stump (Puliti et al., 2018) mainly for small geographical areas (Shin et al., 2018). Furthermore, the derivation of 3D point clouds from orthophotography may increase 3D data availability and improve the subsequent estimation of forestry variables, especially those that describe canopy height (Noordermeer et al., 2019).

ALS have a broad range of applicability within forest management, as for example estimation of forest inventory variables (Guerra-Hernández et al., 2016a; Montealegre et al., 2016), fire-induced change quantification (McCarley et al., 2017) or characterization of forest structural diversity (Kane et al., 2011). ALS point density refers to the number of points per square meter, denoting the spatial resolution. There are two approaches within LiDAR literature to estimate forest variables: the individual tree-based approach (ITB) and the area-based approach (ABA) (Latifi et al., 2015).

The ITB approach generally involves a sequence of tree detection, feature extraction, and estimation of tree variables (Maltamo et al., 2014). Furthermore, it also requires field measurements at tree level. Several algorithms have been proposed for tree segmentation as well as feature extraction (raster-based, point based or multisource-based), providing different accuracies under different forest complexities (Sačkov et al., 2019). The ITB approach generally requires point densities higher than 4-5 points  $\text{m}^{-2}$  (Andersen et al., 2006).

The ABA approach was created by Næsset (1997) and it is also known as the two phase approach inventory or double sampling inventory (Næsset, 2002). The first phase consist on determining the relationships between ALS metrics and the forest stand variables estimated using field data measurements. The second phase fits models that are subsequently extrapolated to the whole study area. The ABA approach has been widely implemented for the estimation of forestry

variables (González-Ferreiro et al., 2013; Latifi et al., 2010; Noordermeer et al., 2018), being a suitable approach for applications using low point density datasets, as the case of the ALS data from the National Plan for Aerial Orthophotography (ALS-PNOA data) in Spain.

The effort made by countries and organizations (i.e.: Finland, Spain, Czech Republic) to provide open low density ALS data at regional and national scales creates new opportunities for forest management. The analysis of the processing methodology of low density ALS-PNOA data in Mediterranean forests has been addressed by Montealegre et al., (2015a and b) who compared several filtering and interpolation routines to assess the most suitable methods to work within Aleppo pine forested areas.

The generation of models requires the selection of the most suitable ALS metrics and regression methods. Variable selection, also known as feature selection, constitutes a relevant step in modelling generation. The recently increase in size of datasets, with tens, hundreds or thousands of available variables, increased the research interest on selection techniques (Guyon & Elisseeff, 2003). This growth in the availability of variables generates a dimensionality problem, typical in many fields of science (Mehmood et al., 2012), which refers to the existence of a higher number of variables than samples. These problem is also known as large  $p$  small  $n$  problem (Martens et al., 1992). The large number of ALS metrics that are derived from the point cloud has substantially increased the number of variables in forestry modelling, being sometimes even higher than the number field plots sampled.

Dealing with large feature sets presents several disadvantages such as technical and model decrease of accuracy. Although hardware and software have improved in the last decades, the use of a large number of variables takes too many computational resources and slows down regression algorithms. Furthermore, in accordance with Kohavi & John (1997), there may be a decrease of model accuracy when the number of variables is significantly higher than the optimal.

Concerning variable selection, a variety of methods exists in order to reduce the dimensionality problem as well as to deal with large feature sets. According to Guyon & Elisseeff (2003) variable selection process have three objectives: improve the prediction performance of the predictors, reduce time and cost when determining the predictors and provide a better understanding of the generated models. A similar definition was proposed by Peña (2002), who established that this selection process should follow the principle of parsimony; filtering or reducing the predictor variables in order to generate models as simple as possible, while maximizing their power.

Variable selection constitutes the first step in model fitting. The regression analysis constitutes the process to estimate or model a relationship between variables. This analysis can be categorized in two main types: parametric and non-parametric regression. The classical regression approach is the parametric one, which assumes the existence of a finite set of parameters. On the contrary, non-parametric methods consider that data distribution cannot be defined by a finite set of parameters.

Parametric regression relies on strong assumptions such as normality, homoscedasticity, linearity, independence, no-collinearity and absence of atypical values. In non-parametric models, meeting

these requirements is not necessary, turning them into a more flexible tool. There exists a wide variety of parametric and non-parametric methods; from simple linear regression models to complex neural networks models. According to Hazelton (2015), non-parametric methods are divided in kernel and local polynomial regression, spline-based regression or regression trees.

The use of machine learning algorithms, showed good performance in several research fields as data mining. Machine learning algorithms, defined as algorithms and statistical models that do not require specific instructions and whose form of the function is unknown, are not directly associated to the two established types of regressions. Parametric regression have been traditionally used in forestry for stand variable prediction with ALS data (Penner et al., 2013). Recently, the application of non-parametric and machine learning methods to this topic has increased popularity (Bollandsås et al., 2013b; Chirici et al., 2008; Liaw & Wiener, 2002).

Model accuracy varies according to several factors such as field and ALS data characteristics, forest complexity, variable selection and regression method used, between others. In this sense, ALS flight configuration determines the final data characteristics, as the quality of the point cloud, affecting model performance. ALS sensors are configured with different scanning patterns, pulse frequencies, scanning frequencies, scanning angles and beam divergence. This configuration, summed up to the flight altitude and speed produce different footprint sizes and point densities. All these settings are normally different from one flight to another.

Several authors have explored the effect of some of these settings on the prediction error in forest attributes modelling. Yu et al. (2004) concluded that an increase in flight altitude decreases accuracy in trees detection and tree height prediction, affecting more to deciduous species than coniferous ones. The effect of footprint size has received little attention in the literature, however several authors agreed that, in small footprint applications, the largest footprints generate a higher bias in the prediction of tree heights (Andersen et al., 2006; Hirata, 2004; Roussel et al., 2017). An increase of pulse frequency, defined as the number of pulses per second and expressed in kilohertz (kHz), generates a lower penetration of the pulses through the canopy (Chasmer et al., 2006b; Næsset, 2009), implying a decrease in tree height prediction accuracy (Chasmer et al., 2006a). The effect of scan angle is relatively low up to  $\sim 20^\circ$ , producing considerably higher prediction errors in forestry metrics with higher values (Liu et al., 2018; Montaghi, 2013). According to Andersen et al. (2006), the beam divergence also affect tree height predictions; a decrease in the angle increases accuracy. Furthermore, the analysis of the effect of point density determines that a decrease of this factor generally does not produce a decrease in accuracy (Garcia et al., 2017; Roussel et al., 2017). Finally, some environmental variables such as slope have been considered a source of error in ALS processing in forested areas. The presence of steep slopes generally increase the errors in point cloud filtering and interpolation processes (Montealegre et al., 2015b).

Forest structure refers to the size, shape, and horizontal and vertical distribution of leaves, branches and stems. These characteristics vary along time, being affected by natural or anthropic disturbances. Fire is caused by natural factors such as volcanic eruptions or lightning and has historically transformed the landscape. Traditional human activities used fire to manage different

land uses, providing an anthropogenic dimension of wildfires. These disturbances constitute some of the most important socio-environmental hazards in Mediterranean forest ecosystems, that might be enhanced by climate change, since extreme meteorological conditions or long droughts increase the fire risk (González-de Vega et al., 2016; Sebastián-López et al., 2008). Although statistic registers showed a reduction in the number of fire events during the last decade (2001-2010), the occurrence of large fires (>500 ha) in Spain has increased (Rodrigues et al., 2014).

Aleppo pine (*Pinus halepensis* Miller) is a flammable species, frequently affected by wildfires, characterized by a high stand density and a continuous presence of branches along the stem (Pausas et al., 2008). Pine forests have a high resilience to fire, but their regeneration process might fail when fire recurrence is high. Forest fires have important effects in vegetation dynamic and atmosphere, as may emit large quantities of greenhouse gases (GHGs) and represent an important carbon sink (Akagi et al., 2013; van der Werf et al., 2010; Wiedinmyer et al., 2011). The quantification of wildfire carbon dioxide (CO<sub>2</sub>) emissions is vital for climate regulation policies (Mieville et al., 2010) as well as for highlighting the service that forests provides to societies (Lal, 2008). The estimation of fire GHGs emissions requires pre-fire biomass estimation, the assessment of the fraction of biomass consumed by fire, usually related to fire severity, and, subsequently, the use of conversion factor to estimate GHG emissions (De Santis et al., 2010). The use of passive remote sensing to estimate fire severity have been broadly analysed in the literature (García-Llamas et al., 2019). Thus, different indexes have been proposed to account for fire damage in vegetation as Normalized Burn Ratio (Key & Benson, 2005), Relative delta Normalized Burn Ratio (Miller & Thode, 2007), between others. Furthermore, the use of ALS data to quantify biomass have been tested for different ecosystems (García et al., 2010; Næsset, 2011).

The estimation of forestry inventory variables is one of the most relevant application for forestry purposes (Latifi et al., 2010; Montealegre et al., 2016). As mentioned above, forest ecosystems constitute important carbon sinks and play a major role in managing GHGs emissions. In this sense, the estimation of biomass and carbon content has growing interest. Several studies have explored the estimation of aboveground tree biomass using ALS data in different ecosystems (Guerra-Hernández et al., 2016b; Mauya et al., 2015). However, the estimation of some biomass fractions, such as shrub biomass or forest residual biomass, have been less studied (Estornell et al., 2012; Hauglin et al., 2014). The presence of understory in forested areas and the existence of shrubland areas are very common in the Mediterranean basin land cover. Thus, the quantification of these biomass fractions may improve the account of carbon reservoirs. Furthermore, the use of some biomass fractions for bioenergy purposes may reduce the CO<sub>2</sub> emissions to the atmosphere produced by other fuels and might help to reach the climate and energy targets of the European Energy Roadmap (Hamelin et al., 2019). Finally, the management of these fractions have several benefits for rural development, as the reduction of wildfire risk or the emergence of new business opportunities for forestland owners (Hauglin et al., 2012).

ALS data provides a wide range of applications when multi-temporal data is available. Forest managers could use multi-temporal ALS data for applications such as: characterizing natural or

anthropic changes, determining under sampled areas and selectively add plots for future inventories, applying existing ALS-based models to subsequent acquisitions in forests with similar characteristics, reducing field work (Fekety et al., 2015) and improving the accuracy of long period forestry trend analysis using fusion of active and passive optical data. Despite the great potential of multi-temporal analysis, its application is still limited by the acquisition costs as well as the need of temporal-concomitant field data (Cao et al., 2016; Dubayah et al., 2010; Ferraz et al., 2018).

The recent effort made by countries and organizations to provide multi-temporal datasets creates new opportunities for upgrading forest inventories at local and regional scales. Local scale refers to small areas whose stand characteristics are similar while regional scale determine large areas with higher stand and structural variability. In addition, several authors have estimated height growth (Gatziolis et al., 2010; Socha et al., 2017) as well as biomass and carbon dynamics (Hudak et al., 2012; Poudel et al., 2018). The estimation of volume (Næsset & Gobakken, 2005; Poudel et al., 2018; Yu et al., 2008), basal area (Næsset & Gobakken, 2005) and site index (Noordermeer et al., 2018) as well as the quantification of wildfire changes (McCarley et al., 2017) and gaps presence (Vepakomma et al., 2008) or the analysis of defoliation effect (Solberg et al., 2006) have also been performed. Two approaches, direct and indirect, have been proposed to model forest attributes using multi-temporal ALS data over time (Bollandsås et al., 2013). The direct approach fits one model for one point in time and temporally transfers the model to other point in time. The indirect one fits two different models for each point in time. The exploration of these approaches provides useful information to forest managers for the reduction of field data acquisitions (Noordermeer et al., 2018).

## 1.2. Importance and justification

LiDAR technology, and specifically airborne laser scanners, has become a valuable source of 3D information to characterize forest ecosystems. Different products can be derived using ALS data, from the generation of precise DTM and digital surface models (DSM) to the characterization of forest stands variables. Thus, the combination of ALS data with field work as well as with data from other remote sensing sources provides accurate information to quantify and evaluate forest resources.

Wildfires constitute a socio-environmental hazard in Mediterranean ecosystems. These events are caused by natural or anthropogenic factors and constitute a relevant source of greenhouse gases emission to the atmosphere. Aleppo pine, being a pyrophyte species, is one of the most affected by fires in Spain. The estimation of fire emissions requires the quantification of pre-fire biomass and the fraction of biomass consumed by the fire. The traditional estimation of pre-fire biomass was performed with field data campaigns, while the estimation of post-fire biomass was based in visual examination or field-based weighting. In this sense, the use of remote sensing tools to determine fire severity, and subsequently, burn efficiency has been widely analysed. The capabilities of ALS data to describe forest structure and quantify forest biomass and the advantages provided by its fusion with optical passive data, might enhance quantification of greenhouse gases emissions sourced from wildfires.

One of the most widespread applications of LiDAR data within forestry is the estimation of stand variables. Traditionally, the estimation of these variables required a systematic sampling of a high number of field plots. The use of ALS data allows applying a stratified random sampling, reducing fieldwork costs, while providing structural information, generally, for the whole study area. Apart from the estimation of traditional inventory variables, as for example dominant height and number of stems or volume, the estimation and mapping of biomass and carbon content has growing interest. In this sense, the estimation of different biomass fractions, such as shrubs and forest residual biomass, has been less analysed. The estimation of these variables are required to better quantify biomass and carbon stocks in Mediterranean Aleppo pine forests at local and regional scales.

The recent effort made by some countries, organizations and companies to acquire multi-temporal datasets opens new opportunities for upgrading forest inventories. Two approaches, the direct and indirect one, exist in ALS literature for temporally transfer models using multi-temporal ALS data. In this sense, different results are found in previous research when comparing both methodologies. The exploration of temporal transferability for estimating forestry attributes at regional scale might improve forest management while reducing modelling time and fieldwork costs.

The exploration of variable selection methods in forest stand variables modelling, using ALS data, might help users to increase model efficiency while reducing modelling time. Variable selection constitutes a relevant step in modelling, especially when the number of independent variables is greater than the sample plots. Furthermore, several studies concluded that the use of large feature sets in modelling presents several disadvantages, such as the use of too many computational resources and the decrease in model accuracy.

The selection of a type of regression method is very important when generating forest stand predictions. Different studies have compared regression algorithms to estimate forest variables determining that the method and parametrization influence model accuracy. Furthermore, the number of ALS metrics included and the number of field plots used to generate the models may also have an important effect. Thus, exploring parametric and non-parametric methods to estimate forest stand variables in Mediterranean pine forest is relevant, as the selection of one method determines model accuracy and consequently forest resources quantification.

The effect of several ALS sensor and flight parameters, as well as of some environmental characteristics on accuracy of predictions has been confirmed by several forest attributes studies. The reduction of point density, the increase of flight altitude, the use of large footprints, the reduction of pulse frequencies or the increase of scan angles, decrease model accuracy. Furthermore, the effect of slope on DTM quality and on the prediction of some stand variables has been also analysed.

Summarizing, this research tried to enrich the knowledge of Mediterranean Aleppo pine forests, which are characterized by a heterogeneous structure and a rugged topography. Furthermore, the

development of forestry applications using ALS-PNOA data enhanced the utility of these public data and may improve forest management at local and regional scales.

### 1.3. Hypotheses and objectives

The estimation of forest stand variables at local and regional scales using ALS and field data constitutes a relevant research topic that provides applicable information for forest managers. ALS data is a suitable technique for mapping 3D structure by relating point clouds to field data. Thus, ALS data has improved traditional forest inventory accuracy while reducing costs.

However, the characterization of Aleppo pine forest as carbon sink, including the different tree and shrub biomass fractions, has been less analysed. The availability of public ALS-PNOA data allows developing studies at regional scale, while previous research performed within Mediterranean ecosystems mainly focus on local scales. Furthermore, although the accessibility to multi-temporal ALS data is still limited by the acquisition costs and by the need of temporal-concomitant field data, the availability of this information provides an opportunity for several applications as forest change characterization. Additionally, the exploration of regression methods when considering different number of field plots and the effect of ALS characteristics and environmental conditions in the prediction error can improve the accuracy of ALS derived products. In this sense, we hypothesized that the use of ALS-PNOA and field data is useful for characterizing Mediterranean Aleppo pine forests with a heterogeneous structure and a hilly topography as well as developing forest applications still not explored in those forests environments.

The main aim of this PhD Thesis was the estimation of forest variables, at local and regional scales, using ALS-PNOA and field data in Aleppo pine forests of Aragón. This objective encompassed several specific aims, which are linked to the papers that constitute the PhD Thesis body (Table 1):

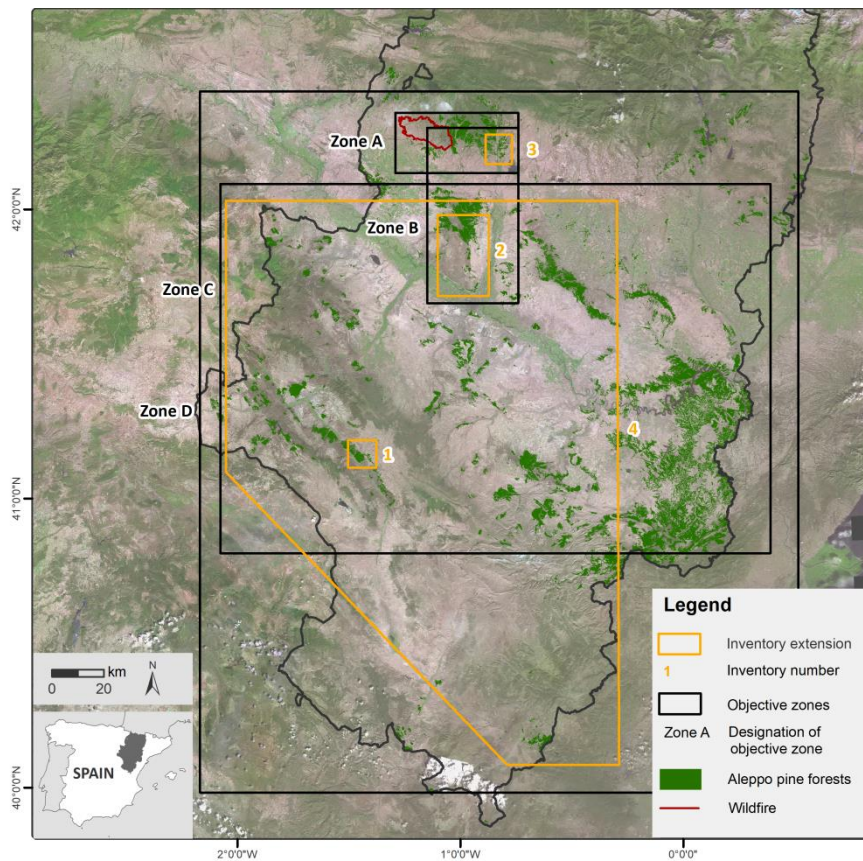
1. Explore the usefulness of low-density ALS data to new applications: estimation of biomass losses and CO<sub>2</sub> emissions to atmosphere, quantification and mapping of forest residual biomass and estimation of different tree fractions and shrub fraction of biomass at stand level.
2. Explore the temporal transferability of models for estimating forest stands attributes at regional scale using multi-temporal ALS-PNOA data.
3. Compare different parametric and non-parametric methods in forest variable modelling.
4. Assessing the suitability of different variable selection methods in order to improve accuracy following the principle of parsimony.
5. Assessing the effect of some ALS parameters and environmental conditions in model performance.

### 1.4. Structure

According to the selected modality by paper compendium, this PhD Thesis includes four chapters and an appendix. In addition to the structure here presented, the first chapter includes the

conceptual framework, the research justification and the objectives. The second chapter describes the study area, data and materials used in the research, as well as the methods and techniques applied to achieve the different objectives. The third chapter contains the published papers that constitute the PhD Thesis body. Finally, the fourth chapter includes the main conclusions as well as possible future research lines. Additionally, the appendix summarizes the metrics of the papers that constitute the PhD Thesis body.

Table 1 includes the specific objectives, methods and utilized tools, linked with the published papers. Figure 2 shows a diagram with the PhD Thesis structure and the relationship between studies, considering the applications, field inventory campaigns, materials and methods. In general, the study area is limited to Aleppo pine forest in Aragón region, but part of the research was performed in smaller zones, according to data availability and complexity. Consequently, the following four zones were selected (Figure 1), being described in more detail in Chapter 2:



**Figure 1.** Location of study zones and inventories. High spatial resolution orthophotography from PNOA Spatial Data Infrastructure (SDI) is included as backdrop.

- Zone A is located in “Las Cinco Villas” region, northwest of Aragón.
- Zone B included two test areas located in the middle Ebro Basin.
- Zone C included a broad part of Aleppo pine forest in Aragón region and cover part of the Ebro Basin and Iberian Range.

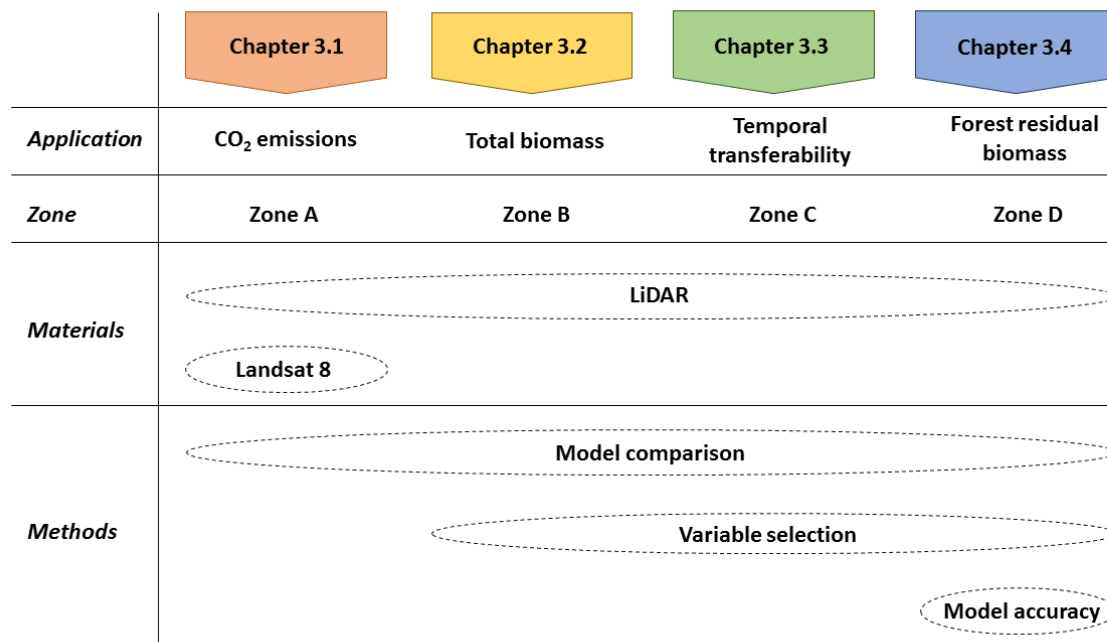


- Zone D represents 197,951.24 ha of the whole Aleppo pine forested area of Aragón. It included all the species distribution range; the forested areas in Zone C, as well as the ones located in the pre-Pyrenees and in the south of Teruel city.

**Table 1.** Summary of PhD Thesis specific aims, methods, tools, and its contributing papers.

| Specific objectives   | Methods and techniques   |
|---|--|
| 1. Estimate biomass losses and CO <sub>2</sub> emissions using low-density ALS data in a burnt Aleppo pine forest.<br><br>3. Compare different parametric and non-parametric models.  | <ul style="list-style-type: none"> <li>- Computation of above ground biomass per plot using allometric equations.</li> <li>- Comparison of nine regression methods to estimate above ground biomass.</li> <li>- Assessment of statistically significant differences between models using Friedman and Nemenyi post-hoc statistical tests.</li> <li>- Extrapolation of the model to the burned area and estimation of pre-fire biomass.</li> <li>- Estimation of fire severity using <math>\Delta</math>NBR index from Landsat 8.</li> <li>- Mapping pre-fire Aleppo pine forest using the Spanish National Forest Map and ALS data.</li> <li>- Application of three burning efficiency factors according to <math>\Delta</math>NBR values.</li> <li>- Conversion of biomass losses to carbon content and, subsequently to CO<sub>2</sub> emissions.</li> </ul> |
| <b>Publication</b>  |  |
| <i>Domingo, D., Lamelas, M.T., Montealegre, A.L., de la Riva, J. 2017. Comparison of regression models to estimate biomass losses and CO<sub>2</sub> emissions using low-density airborne laser scanning data in a burnt Aleppo pine forest. European Journal of Remote Sensing, 50 (1), 384-396. doi: 10.1080/22797254.2017.1336067.</i> |  |
| Specific objectives   | Methods and techniques   |
| 1. Estimating total biomass in Aleppo pine forest stands using low-density ALS data.<br><br>4. Assessing the suitability of variable selection methods.   | <ul style="list-style-type: none"> <li>- Computation of total biomass per plot using tree and shrub allometric equations.</li> <li>- Selection of independent ALS variables using five selection processes.</li> <li>- Comparison of five regression methods to estimate total biomass.</li> <li>- Modelling and comparison between regression models using statistical tests.</li> <li>- Mapping of total biomass.</li> </ul>   |
| <b>Publication</b>  |  |
| <i>Domingo, D., Lamelas, M.T., Montealegre, A.L., García-Martín, A., de la Riva, J. 2018. Estimation of total biomass in Aleppo pine forest stands applying parametric and nonparametric methods to low-density airborne laser scanning data. Forests, 9,158-175. doi: 10.3390/f9030158.</i>  |  |
| Specific objectives   | Methods and techniques   |
| 2. Explore the temporal transferability of  | <ul style="list-style-type: none"> <li>- Computation of seven forest stand variables per plot using allometric equations.</li> </ul>   |

|  |   |
|--|---|
| models for estimating forest stands attributes at regional scale using multi-temporal ALS-PNOA data.   | <ul style="list-style-type: none"> <li>- Generation of concomitant field data using the PHRAGON 2017 model included in the Simanfor simulator.</li> <li>- Assessment of temporal transferability by comparing direct and indirect approaches using multi-temporal ALS data.</li> <li>- Prediction of stand variables and model comparison using statistical tests.</li> </ul>   |
| <b>Publication</b>   |   |
| <p><i>Domingo, D., Alonso, R, Lamelas, M.T., Montealegre, A.L., Rodríguez, F, de la Riva, J. 2019. Temporal Transferability of Pine Forest Attributes Modelling Using Low-Density Airborne Laser Scanning Data. Remote Sensing, 11(3), 261. doi: 10.3390/rs11030261.</i></p>   |   |
| <b>Specific objective</b>  | <b>Methods and techniques</b>   |
| <ol style="list-style-type: none"> <li>1. Quantifying and mapping forest residual biomass.</li> <li>5. Assessing the effect of some ALS parameters and environment conditions in model performance.</li> </ol>   | <ul style="list-style-type: none"> <li>- Computation of forest residual biomass per plot using three biomass fractions: thick, medium and thin branches.</li> <li>- Comparison of variable selection processes and regression methods.</li> <li>- Assessment of the effect of three ALS characteristics on model accuracy: point density, scan angle and canopy penetration pulse.</li> <li>- Assessment of the effect of two environmental characteristics on model accuracy: slope and shrub presence.</li> <li>- Modelling and comparison between regression models using statistical tests.</li> <li>- Mapping and quantifying of forest residual biomass in 87.66% of Aleppo pine forested area in Aragón region.</li> </ul> |
| <b>Publication</b>   |   |
| <p><i>Domingo, D., Montealegre, A.L., Lamelas, M.T., García-Martín, A., de la Riva, J., Rodríguez, F, Alonso, R. 2019. Quantifying forest residual biomass in Pinus halepensis Miller stands using Airborne Laser Scanning data. GIScience and Remote Sensing, 56 (8), 1210-1232. doi: 10.1080/15481603.2019.1641653</i></p> |   |



**Figure 2.** Diagram of the PhD Thesis structure that evidences the relation between the studies carried out.



## 2. Study area, materials and methods

*This chapter describes the study area, which includes four different zones within Aragón region, as well as the material and methods utilized in the research. Firstly, we describe the field inventory data, the allometric equations used for computing the different forest variables and the pre-processing performed to the data. Secondly, the remote sensing information used; the ALS-PNOA data and passive optical images, are presented, including the pre-processes applied. Thirdly, we include the selection and regression methods utilized for modelling different variables at stand level. Then, we define the methods to analyse the temporal transferability of multi-temporal ALS data. Furthermore, the methods utilized to assess the effect of ALS parameters and environmental conditions in model accuracy are presented. Finally, the mapping process and its importance in the generation of information at local or regional scale is addressed.*



## 2.1. Study area

Aleppo pine is the most broadly distributed species from genus *Pinus* in the Mediterranean basin. The regions with higher presence are the north of Africa and Spain, which represents ~850,000 ha, according to Cámara (2001). The wide altitudinal and latitudinal gradient of this distribution allows this species to live from sea level up to 1,600 m in the Saharan Atlas (Cabanillas 2010). Although Aleppo pine is a limestone species, it grows in different types of soils such as siliceous, quartzite or granite. Furthermore, the species is heliophilous, thermophile, xerophile and pyrophyte, being adapted to droughts and wildfires.

This PhD Thesis studied the Aleppo pine forest of Aragón region. Aragón is an Autonomous Community located in the northeast of Spain. This region occupies 47,720.3 km<sup>2</sup> and represents 9.4% of the Spanish territory. Three provinces; Zaragoza, Huesca, and Teruel, constitute this Autonomous Community. Aragón limits to the north with France, to the east with Cataluña and Valencia and to the west part with Castilla-La Mancha, Castilla y León, La Rioja and Navarra.

Three main relief units compose Aragón, the Pyrenees, the Ebro Basin and the Iberian Ranges. The Pyrenees are represented by the Axial Pyrenees, the interior mountains, the Intrapyreanean topographic depression and the exterior mountains (pre-Pyrenees). Axial Pyrenees, constituted by granites, quartzite, slates and limestone, presents the highest altitudes such as Aneto (3,404 m) or Maladeta (3,308 m). The interior mountains include calcareous crests adhered to the axial Pyrenees. The Intrapyreanean topographic depression contains several perpendicular river valleys, ending in “*San Juan de la Peña*” and “*Peña Oroel*” conglomeratic escarps. The pre-Pyrenees, constituted by calcareous rocks, present heights between 1,500 and 2,000 m (Peña & Lozano, 2004).

The “Somontanos” connect the pre-Pyrenees with the left bank of the Ebro Basin. The Ebro Valley was a sea during Mesozoic and Eocene stages. Nowadays it is a topographic depression, covered by tertiary materials and alluvial sediments. The erosion processes have generated different tabular reliefs from 500 up to 800 m in both banks of Ebro River.

The Iberian Ranges is a mountain chain of lower altitude than Pyrenees. The “*Sierra del Moncayo*” constitutes the northwest part, “*Puertos de Beceite*” and “*Gúdar-Maestrazgo*” are located in the eastern part, while “*Javalambre*” and “*Albarracín*” in the southeastern part. Quartzite and Palaeozoic slate are the main materials present in the higher mountains, while Jurassic and Cretaceous limestone and dolomites constitute the lower relief structures.

The climate of Aragón is Mediterranean with continental features, characterized by cold winters, dry summers and irregular and scarce rainfall. The variability in orography modifies the temperatures and precipitations, generating a wide diversity of climatic ambient from semi-desertic areas such as Monegros to high mountains in the Pyrenees. According to Cuadrat (2004), Aragón climate is defined by four main characteristics. The Ebro valley presents low annual precipitation values, ~300 mm m<sup>2</sup>, generated by the shadow effect of Pyrenees and Iberian Ranges; its location in a continental area generates a broad range of temperatures; precipitations are

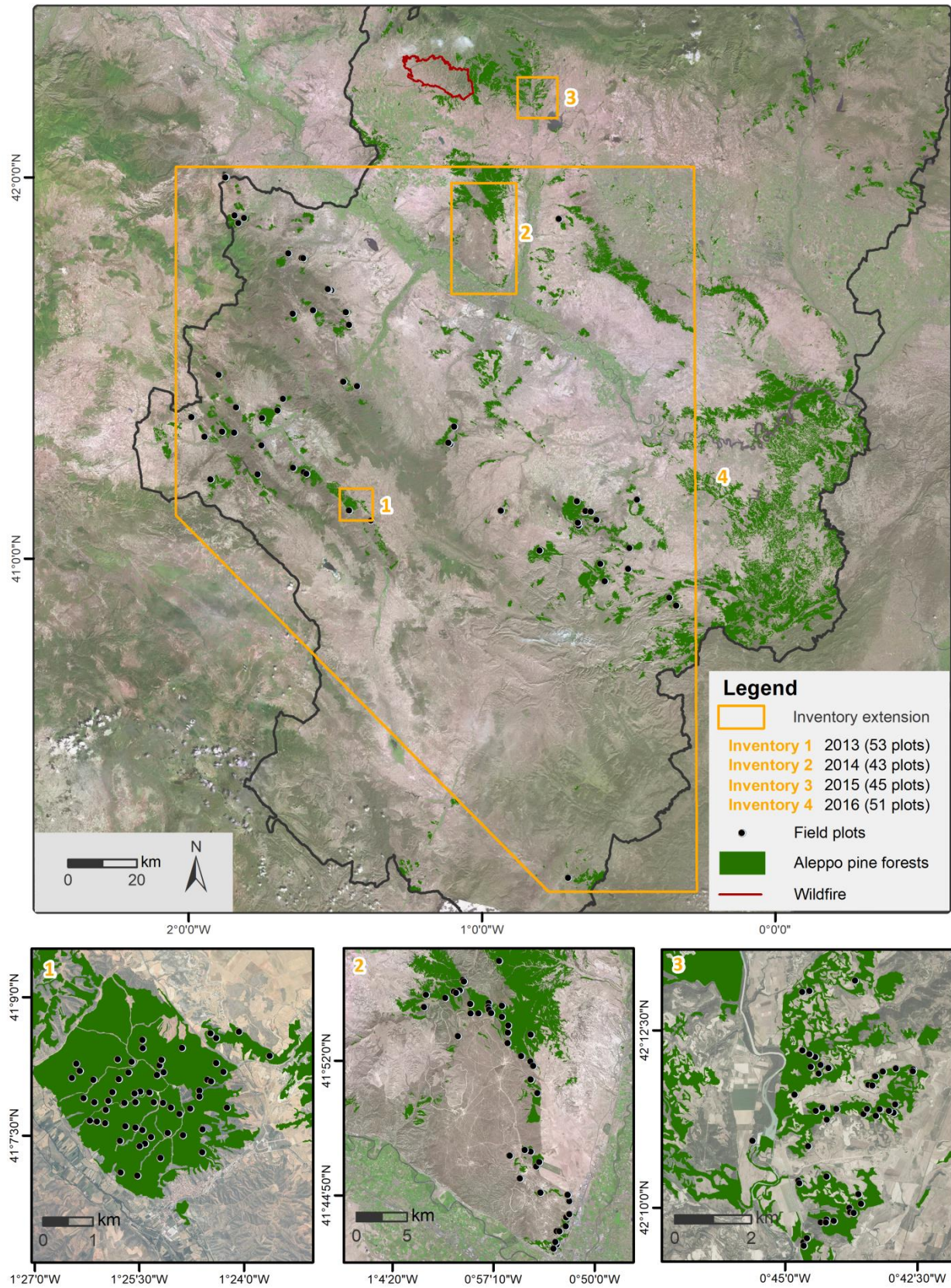
irregular and the winds come from the northwest in winter and southeast in summer, being frequent and heavy.

Aragón includes two Holarctic biogeographical regions: Eurosiberian and Mediterranean. The Eurosiberian region is occupied by forests and grasslands distributed in three altitudinal strata: alpine, subalpine, and montane. The Mediterranean region includes the Ebro Valley, the “Somontanos”, the river valleys located in the right margin of Ebro River and the topographic depression in which Teruel city is situated. *Quercus ilex*, *Pinus halepensis*, *Pinus nigra* and *Juniperus sabina* constitute the species that dominate Mediterranean forests (Longares, 2004). According to the Spanish Forest Map, the forested area in Aragón represents 1.58 million of ha, of which 259,057.45 ha are occupied by Aleppo pine forests. Concretely, 124,473.12 ha are located in Zaragoza, 37,817.19 ha in Huesca and 96,767.14 ha in Teruel regions. In the whole, semi-natural forests represent 211,013.49 ha and afforested forest 48,043.96 ha.

As mentioned before, four study zones were delimited to provide answers to the specific research objectives (Figure 1). The study zones are described below:

- Zone A is located in “Las Cinco Villas” region, northwest of Aragón. The relief is characterized by a topographic depression close to the pre-Pyrenees. Elevations range from 430 to 1150 m above sea level and slopes from 0° to 39°. The climate is Mediterranean with continental features with an annual average precipitation of 525 mm. The Zone A includes two different test areas: “Luna” wildfire and field inventory 3 (Figure 3). Luna wildfire was caused by agricultural machinery on 4<sup>th</sup> July 2015. The fire scorched 14,263 ha, 3,390.4 ha covered by woodland. Field inventory 3 is located in an unburned area close to the wildfire, with similar environmental, climatic and forest characteristics. The unburned Aleppo pine forest is heterogeneous from the structural point of view, being accompanied by an evergreen understorey with species such as *Quercus ilex* subsp. *rotundifolia*, *Quercus coccifera*, *Juniperus oxycedrus*, *Buxus sempervirens* and *Juniperus phoenicea*.
- Zone B is located in the Ebro Basin, Northeast Spain. This zone includes two different test areas: inventory 2 (Figure 3) and inventory 3. Both areas are representative of Aleppo pine Mediterranean forest, occupying 11,400 ha. The inventory 3 area was described above. The majority of Inventory 2 plots are located in the military training center “San Gregorio” (CENAD), located in the north of Zaragoza city. The area presents a hilly topography with elevations ranging from 300 to 750 m and slopes from 0° to 39°. Climate is Mediterranean with continental features, while the average annual precipitation is lower than 350 mm. Most of the pine stands are semi-natural, while the stands located in the south eastern part were planted approximately forty years ago. The evergreen understorey is characterized by xerophiles species such as *Quercus coccifera*, *Juniperus oxycedrus*, *Rosmarinus officinalis* and *Thymus vulgaris*.





- Zone C is located in the Ebro Basin and Iberian Ranges, including a broad part of Aleppo pine forest in Aragón region, except from the forested areas close to the pre-Pyrenees and the stands located in the south of Teruel. This zone includes three different test areas: inventory 1 (Figure 3), inventory 2 and most of the plots from inventory 4. Inventory 2 was described previously. Inventory 1 is located in Daroca Municipality in the Iberian Range. The area includes two forests denominated “*Dehesa de los enebrales*” and “*Valdá y Carrilanga*”, covering 1,102 ha. Both forests were afforested from 1908 to 1979, being occupied by a monospecific Aleppo pine forest with scarce understory. The area presents a hilly topography, with elevations ranging from 860 up to 980 m. Inventory 4 includes several stands from the Middle Ebro Basin up to the Iberian Ranges. The stands were afforested approximately forty to sixty years ago, keeping a low presence of hardwood species. The different environmental sample conditions include a great variability of Aleppo pine forests.
- Zone D represents 197,951.24 ha of the Aleppo pine forested area of Aragón, including all the species distribution range. The Zone D includes four different test areas: inventories 1 to 4 (Figure 3), which have been previously described. The broad area shows a variability of geomorphological forms from topographic depressions up to mountains that reach more than 2,000 m above sea level. Consequently, temperature changes across the altitudinal gradient and the annual precipitation range from less than 350 mm up to 1,000 mm. This variability and the presence of semi-natural and afforested stands are characteristic of Aleppo pine forest at Aragón region.

## 2.2. Materials and methods

### 2.2.1. Field inventory data

Field inventories provide an accurate quantification of forest variables within a small fraction of the study area. Field plot information is the ground-truth evidence, constituting the dependent variables that are estimated for a broad area. In this sense, field plot data must be representative for the study area, capturing the maximum variability to minimize extrapolation errors.

Sampling design is a relevant factor to capture forest stand variability. Several sampling types exist, such as systematic sampling, random sampling or stratified random sampling. Traditional inventories, performed using field campaigns, require the division of forest stands into homogeneous strata, considering factors such as tree species, site productivity, forest development stage, latitude, elevation and stand structure. Thus, performing inventories using ALS data is easier as this technology provides information about some of those factors at a stand level. This PhD Thesis applied a stratified random sampling, considering terrain slope, canopy height and canopy cover variability, according to Næsset & Økland (2002), in monospecific Aleppo pine forest. Field data was acquired in 192 plots in four campaigns performed during 2013, 2014, 2015 and 2016 (hereinafter mentioned as first, second, third and fourth campaign, respectively). Field data were related to ALS data using an area-based approach (Næsset & Økland, 2002). The

number of plots allowed the estimation of forest stand variables at local and regional scales. The specifications of each inventory are described below:

- The acquisition of field data from the first campaign was performed from June to July 2013, in 53 circular plots within the Master Thesis of Jesús Cabrera (Cabrera, 2013). The center point of each circular plot with 15 m radius was positioned using a Leica VIVA® GS15 CS10 real-time kinematic GNSS with a planimetric accuracy of 0.30 m. We used a diameter tape, with millimeter precision, for measuring tree diameter at breast height (dbh) in those trees with a dbh larger than 7.5 cm, which is the standard dbh for inventoried trees in Spain. A Suunto® hypsometer was used for measuring green crown height and tree height of up to 4 randomly selected trees within each plot. The selection of the sample trees, from 7.5 cm up to 42.5 cm, considers the diametric classes defined as representative for the study area in the third national forest inventory. The height for those trees not measured in the field plots was predicted by using a height-diameter model developed from the sampled trees (equation 1). The model performance of the height model gave a RMSE of 1.36 m and  $R^2$  of 0.63. Normality, homoscedasticity and independence or no auto-correlation in the residuals were verified for the fitted model.

$$ht = 0.776 \cdot G^{0.179} \cdot dbh_i^{0.660} \cdot 1.009 \quad (1)$$

where  $ht$  is tree height (m),  $dbh_i$  is the diameter at breast height (cm) and  $G$  is field plot basal area ( $m^2 ha^{-1}$ ).

- The second and third field campaigns use the same inventory methodology. The second campaign includes 43 circular plots acquired from July to September 2014 (Montealegre et al., 2016). The third field plot campaign sampled 45 plots from June to July 2015, being used for objectives 1, 2 and 4. The same GNSS instrument used for the first campaign positioned the 30 m diameter plots, obtaining a planimetric accuracy of 0.15 and 0.18 m in 2014 and 2015, respectively. A Haglöf Sweden® Mantax Precision Blue diameter calliper allowed measuring the dbh of those trees with a dbh larger than 7.5 cm. We used a Haglöf Sweden® Vertex instrument to measure the green crown height and the height for all trees in the plot. Furthermore, the percentage of shrub canopy cover and the average height of the different shrub species that represent the understory were measured (used for objective 2).
- The fourth campaign inventoried 51 field plots in April 2016 being carried out by föra forest technologies within the project RF-64079 (Rural Development Program of Aragón 2014–2020). A Trimble submetric GNSS was used to position the center of each plot with a submetric accuracy in planimetry. A variable plot radius was selected (5.6 m, 8.5 m, 11.3 m, and 14.10 m), in order to obtain data from a similar number of trees in each plot, due to the difference in stand density. A Haglöf Sweden® Mantax Precision Blue diameter calliper allowed measuring those trees with a dbh larger than 7.5 cm. A Haglöf Sweden® Vertex was used for measuring the green crown height and the height of up to 6 trees, the nearest to the plot center. The sample was completed to achieve 100 dominant stems  $ha^{-1}$ ,



considering those with larger dbh. The height for those trees not measured in the field was estimated by using a height-diameter model developed from the sampled trees (equation 2). The model performance of the height model gave a RMSE of 0.80 m and  $R^2$  of 0.93.

$$ht = \left( 1.3^{2.5511} + (H_0^{2.5511} - 1.3^{2.5511}) \cdot \frac{1 - \exp(-0.025687 \cdot dbh)}{1 - \exp(-0.025687 \cdot D_0)} \right)^{1/2.5511} \quad (2)$$

where  $ht$  is tree height (m),  $dbh$  is the diameter at breast height (cm),  $H_0$  is the Assmann dominant height (m) and  $D_0$  is the Assman dominant diameter (cm).

### 2.2.2. Estimation of forest stand variables

The estimation of forest variables at tree level requires the use of allometric equations. These equations generally need a destructive sampling, drying and subsequent weighting of a representative sample. The majority of tree forest species have an allometric equation. However, the accessibility of shrub allometric equations in Spain is more limited, being available for the estimation of some specific variables such as biomass (Montero et al., 2013). This PhD Thesis estimated nine stand variables using allometric equations as a ground-truth. The estimated variables are: stand density (N), basal area (G), squared mean diameter (Dg), dominant diameter (Do), dominant height (Ho), timber volume over bark of stem (V), above ground biomass ( $W_{ag}$ ), total tree biomass (W) and forest residual biomass (FRB). In addition, the equations determined by Montero et al. (2013) for different shrub formations defined in the Spanish Forest Map (MFE) were used to calculate shrub biomass. Finally, total biomass including aboveground tree biomass and shrub biomass (TW) was computed by summing up total tree biomass and shrub biomass.

#### Tree equations

The equations 3 to 16 were used to estimate the above mentioned tree variables.

$$N \text{ (stems } ha^{-1}) = \frac{stems}{a} \quad (3)$$

$$G \text{ (m}^2 \text{ } ha^{-1}) = \frac{\sum_i^n \frac{\pi}{4} d_i^2}{a} \quad (4)$$

$$Dg \text{ (cm)} = 100 \cdot \sqrt{\frac{4 \cdot G}{\pi \cdot N}} \quad (5)$$

$$Do = \frac{\sum_i^k d_i}{a/100} \quad (6)$$

$$Ho = \frac{\sum_i^k h_i}{a/100} \quad (7)$$

$$v_i = \frac{\pi}{40000} \int_0^{h_t} \left[ \left( 1 + 1.121163 \cdot e^{\left( -10.23293 \cdot \frac{h_i}{h_t} \right)} \right) \cdot 0.696362 \cdot d_i \cdot \left( \left( 1 - \frac{h_i}{h_t} \right)^{1.266261 - (0.003553 \cdot E) - 1.865418 \cdot \left( 1 - \frac{h}{h_t} \right)} \right) \right]^2 d_i h_i$$

$$V \text{ (m}^3 \text{ } ha^{-1}) = 10000 \cdot \frac{\sum_i^n v_i}{a} \quad (8)$$

$$W_{ag} \text{ (kg ha}^{-1}\text{)} = \frac{CF \cdot e^a \cdot d_i^b}{a} \cdot 10,000 \quad (9)$$

$$W_s \text{ (kg)} = 0.0139 \cdot d_i^2 \cdot h_i \quad (10)$$

$$W_{b7} \text{ (kg)} = [3.926 \cdot (d_i - 27.5)] \cdot Z; \quad (11)$$

If  $d_i \leq 27.5$  cm then  $Z = 0$ ; If  $d_i > 27.5$  cm then  $Z = 1$

$$W_{b2-7} \text{ (kg)} = 4.257 + 0.00506 \cdot d_i^2 \cdot h_i - 0.0722 \cdot d_i \cdot h_i \quad (12)$$

$$W_{b2+n} \text{ (kg)} = 6.197 + 0.00932 \cdot d_i^2 \cdot h_i - 0.0686 \cdot d_i \cdot h_i \quad (13)$$

$$W_r \text{ (kg)} = 0.0785 \cdot d_i^2 \quad (14)$$

$$W \text{ (kg)} = W_s + W_{b7} + W_{b2-7} + W_{b2+n} \quad (15)$$

$$FRB \text{ (kg)} = W_{b7} + W_{b2-7} + W_{b2+n} \quad (16)$$

where  $d$  is the normal diameter in cm;  $a$  is the area of the plot expressed in  $m^2$ ;  $\sum^n$  refers to the number of trees inside of a plot;  $\sum^k$  refers to the number of  $k$  trees being  $k$  the thickest trees;  $h_i$  is the height of the trees;  $h_t$  is the total tree height;  $E$  is the tree slenderness ( $h_t/d_i$ );  $v_i$  is the volume of each tree in  $m^3$ ;  $CF$  is a correction factor ( $CF = e^{SEE^2/2}$ ) being  $e$  the Euler number and  $SEE$  the standard error (0.151637);  $a$  is (-2.0939) and  $b$  (2.20988) are the specific parameters for Aleppo pine;  $W_s$  is the biomass weight of the stem fraction,  $W_{b7}$  is the biomass weight of the thick branch fraction (diameter larger than 7 cm),  $W_{b2-7}$  is the biomass weight of medium branch fraction (diameter between 2 and 7 cm),  $W_{b2+n}$  is the biomass weight of the thin branch fraction (diameter smaller than 2 cm) with needles, and  $W_r$  is the biomass weight of the roots.

### Shrub equations

The shrub equations used for each formation are presented below (equations 17 to 21):

- Shrub hedges, borders, galleries, etc.:

$$\ln(W_s) = 0.494 \times \ln(CC) \quad (17)$$

- *Quercus coccifera* and *Pistacia lentiscus*:

$$\ln(W_s) = -2.892 + 1.505 \times \ln(h_m) + 0.462 \times \ln(CC) \quad (18)$$

- *Leguminosae aulagoideas* and related shrubs:

$$\ln(W_s) = -2.464 + 0.808 \times \ln(h_m) + 0.761 \times \ln(CC) \quad (19)$$

- *Labiatae* and *Thymus* formations:

$$\ln(W_s) = -1.877 + 0.643 \times \ln(h_m) + 0.661 \times \ln(CC) \quad (20)$$

- General shrub biomass:

$$\ln(W_s) = -2.560 + 1.006 \times \ln(h_m) + 0.672 \times \ln(C) \quad (21)$$

where  $W_s$  is the biomass weight for each species in tons/ha,  $h_m$  is the average shrub height at plot level and  $CC$  is the percentage of shrub canopy cover at plot level.

Equation 17 was applied for *Crataegus monogyna*, *Rhamnus lycioides* and *Rosa canina*; equation 18 was used for *Quercus coccifera*; equation 19 was applied for *Genista scorpius*; equation 20 was used

for *Thymus* sp. Subsequently, general equation 21 was used for the rest of inventoried species: *Rosmarinus officinalis*, *Juniperus oxycedrus*, *Juniperus sabina*, *Buxus sempervirens*, *Genista florida*, and *Salsola vermiculata*. Finally, equation 21 was applied for *Quercus ilex* and Aleppo pine with less than 7.5 cm of dbh, due to the impossibility of using available tree allometric equations.

### 2.2.3. Inventories updating

Modelling forest dynamics is essential for forest management, being a relevant research topic in forestry from 1930s to the present (Tesch, 1980). According to Vanclay (1994) a stand growth model is an abstraction of the natural dynamics of a forest stand, which may consider growth, mortality and other changes in stand composition and structure. Growing models generally include different equations to predict the growth and yield of a forest stand under a wide variety of conditions (Vanclay 1994).

The initial purpose of growth models was to estimate timber yield, improving efficiency of timber production (Shifley et al., 2017). Nowadays, the wide variety of existent models try to explain a broad variety of forest dynamic processes, as for example the effect of climate change on tree growth, by using hybrid models.

A simple classification of growth models considers the level of detail. Accordingly, Vanclay (1994) differentiate between models based in the whole stand, size class models or single-tree models. Model complexity and purpose is also used for classification. The application of empirical growth models based on yield tables or regression equations constitutes a traditionally applied approach. Nowadays, state-space stand-level models (García, 2003), distribution-based models and both, individual-tree models and complex process-based eco-physiological models (Thornley, 2006), have dramatically increased flexibility and realism to forest simulations.

This PhD Thesis applied empiric growth models (objectives 2 and 4) and individual-tree growth models (objective 3) to generate temporally concomitant field data, which was subsequently related to ALS data in order to estimate forest stand variables. The implemented methodology is described below.

#### *Empiric growth models*

The estimation of tree growth for predicting forest residual biomass and total biomass was performed by using yield tables from the Spanish National Forest Inventory (NFI). Specifically, we applied the dbh and height growth values for Aleppo pine between the second NFI (NFI2) and the third one (NFI3). The prediction of the growth values uses regression curves by least squares and considers the difference between both inventories (11 years). Furthermore, a linear interpolation based on these tables allowed obtaining a subtractive value of dbh and height per each diametric class proposed by the NFI (Table 2).

**Table 2.** Values of tree diameter at breast height (dbh) and height growth between the NFI2 and the NFI3, and subtractive values of dbh and height when applying linear interpolated degrowth.

| Diametric<br>class (cm) | dbh<br>growth<br>(mm) | Height<br>growth<br>(m) | Field campaign |               |             |               |             |               |             |               |
|-------------------------|-----------------------|-------------------------|----------------|---------------|-------------|---------------|-------------|---------------|-------------|---------------|
|                         |                       |                         | 2013           |               | 2014        |               | 2015        |               | 2016        |               |
|                         |                       |                         | dbh<br>(mm)    | Height<br>(m) | dbh<br>(mm) | Height<br>(m) | dbh<br>(mm) | Height<br>(m) | dbh<br>(mm) | Height<br>(m) |
| <10                     | 24                    | 1.20                    | -4.36          | -0.22         | -6.55       | -0.33         | -8.73       | -0.44         | -10.91      | -0.55         |
| 10–15                   | 29                    | 1.40                    | -5.27          | -0.25         | -7.91       | -0.38         | -10.55      | -0.51         | -13.18      | -0.64         |
| 15–20                   | 33                    | 1.50                    | -6.00          | -0.27         | -9.00       | -0.41         | -12.00      | -0.55         | -15.00      | -0.68         |
| 20–25                   | 30                    | 1.40                    | -5.45          | -0.25         | -8.18       | -0.39         | -10.91      | -0.51         | -13.64      | -0.64         |
| 25–30                   | 30                    | 1.40                    | -5.45          | -0.25         | -8.18       | -0.39         | -10.91      | -0.51         | -13.64      | -0.64         |
| 30–35                   | 33                    | 1.10                    | -6.00          | -0.20         | -9.00       | -0.30         | -12.00      | -0.40         | -15.00      | -0.50         |
| 35–40                   | 32                    | 1.50                    | -5.82          | -0.27         | -8.73       | -0.41         | -11.64      | -0.55         | -14.55      | -0.68         |
| 40–45                   | 27                    | 1.60                    | -4.91          | -0.29         | -7.36       | -0.44         | -9.82       | -0.58         | -12.27      | -0.73         |
| 45–50                   | 24                    | 1.90                    | -4.36          | -0.35         | -6.55       | -0.52         | -8.73       | -0.69         | -10.91      | -0.86         |
| 50–55                   | 58                    | 1.00                    | -10.55         | -0.18         | -15.82      | -0.28         | -21.09      | -0.36         | -26.36      | -0.45         |
| 55–60                   | 10                    | 0.50                    | -1.82          | -0.09         | -2.73       | -0.14         | -3.64       | -0.18         | -4.55       | -0.23         |

The estimation of shrub growth was performed according to Montero et al. (2013) equations (equations 22 to 25). The equations were developed for different shrub formations determined in the Spanish Forest Map (MFE). It should be noted that no growing equations were developed by Montero et al. (2013) for shrub edges, borders and galleries formation. Consequently, a general equation for shrub (equation 25) was applied for *Crataegus monogyna*, *Rhamnus lycioides*, and *Rosa canina*.

- *Quercus coccifera* and *Pistacia lentiscus*:

$$\ln(W_s) = -4.955 + 1.150 \times \ln(h_m) + 0.463 \times \ln(CC) \quad (22)$$

- *Leguminosae aulagoideas* and related shrubs:

$$\ln(W_s) = -4.479 + 0.715 \times \ln(h_m) + 0.701 \times \ln(CC) \quad (23)$$

- *Labiatae* and *thymus* formations:

$$\ln(W_s) = -4.446 + 0.753 \times \ln(h_m) + 0.573 \times \ln(CC) \quad (24)$$

- General shrub biomass:

$$\ln(W_s) = -4.771 + 0.814 \times \ln(h_m) + 0.676 \times \ln(CC) \quad (25)$$

where  $W_s$  is the biomass weight for each species in tons/ha,  $h_m$  is the average shrub height at plot level, and  $CC$  is the percentage of shrub canopy cover fraction at plot level.

Then, the shrub biomass growing values per year in every plot were calculated summing up the biomass values for each species obtained from equations 22 to 25. These annual growing values

were subtracted from the measured ones considering the difference in years between ALS flight and field data acquisition.

#### ***Individual-tree growth models***

Individual-tree growth models are powerful tools to update stand variables to the ALS mission date being more accurate and consistent than empirical growth models. The applied model, PHRAGON-2017 (Alonso, 2018), is included in the Simanfor simulator platform (Bravo et al., 2012).

The model is specially designed for Aleppo pine afforestation in Aragón, enabling tree-level distance-independent simulation. The model includes a set of equations for diameter over bark growth, diameter under bark growth, diameter under bark–diameter over bark ratio, generalized height–diameter relationship, volume over bark (taper equation) and crown ratio. Furthermore, it includes a 10-year survival model and a classification tree for the regeneration of *Pinus*, *Quercus* and *Juniperus*.

The explanatory variables that need to be included to compute the model are: dbh, total height, stand density (basal area, Hart–Becking index), dominant trees (dominant height, dominant diameter), competition (BALMOD) (Schröder & Gadow, 1999) and site quality (site index). The model uses the site index curves developed for natural Aleppo pine forests in the Central Ebro Valley (Rojo-Alboreca et al., 2017) to calculate site index and dominant height evolution.

The model uses two different approaches depending on future or past projections. When projecting to the future, we need to apply the diameter growth and survival equations to every single tree in each plot. The site index curve is used to forecast the future stand dominant height, and hence estimate the total height of each surviving tree. On the other hand, when projecting stand structure to the past, we need to deploy the diameter under bark growth equation. This procedure, denominate backdating, allows the use of the current stand features to predict the past growth of a tree. Therefore, the estimation of past tree diameters over bark requires the diameter under bark –diameter over bark ratio. The site index dynamic curves allow calculating past dominant height, considering that curves are age-independent functions. Once the diameter and dominant height values are included in the model, the rest of stand variables are directly computed.

#### **2.2.4. ALS-PNOA data**

The capture of ALS data in Spain is associated to the “*Plan Nacional de Observación del Territorio*” known by the acronym PNOT. This plan captures different types of geographical information to characterize the Spanish territory and its evolution over time. PNOT provides valuable information for sustainable environmental management, forestry, agriculture, infrastructures planning, emergencies and security. The PNOT plan includes three national programs:

- National plan for aerial orthophotography (PNOA). The main objective of the PNOA is the acquisition of orthophotography, DTMs generation and ALS data capturing at different



spatial resolutions and considering different update rhythms. PNOA plan is further described below.

- Remote sensing national plan (PNT). The PNT started in 2004 and its main aim is the acquisition of satellite images for the Spanish territory at different spatial resolutions: high, medium and low. The PNT acquires, processes, and distributes these images to all the public administration and public universities.
- Spanish information system for land use (SIOSE). The main purpose of SIOSE is the generation of a geographic database of land use in Spain at 1:25,000 scales. The SIOSE uses information from PNOA and PNT to derive land use cartography that is updated periodically.

PNOT follows a decentralized financial model based on the coordination and co-financing of different public administrations. The main aim of this bottom up model is to avoid duplicities, minimizing the effort and resources. PNOT follows the Infrastructure for Spatial Information in Europe (INSPIRE), providing free access to high quality geographical information.

The aim of PNOA plan is the acquisition of digital aerial orthophotography at 25 and 50 cm resolution, as well as DTMs actualization every 2 or 3 years, depending on the area of interest. The project is co-financed by the General State Administration and the Autonomous Communities. The products provided by the PNOA plan are considered basic data within the INSPIRE directive. These products allow deriving other complex information such as land use, cadastre, forest management or hydrological cartography.

The ALS-PNOA plan, integrated within the PNOA plan, have as main objective the acquisition of ALS data for all Spain. The data classified and colorized using the PNOA orthophotography is publicly available. ALS-PNOA started at 2009 with the aim of generating high resolution DTMs for flood mapping and providing information for infrastructures planning. Although at the beginning only the DTMs were available, nowadays the ALS point clouds of the whole Spanish territory are accessible for all the citizens.

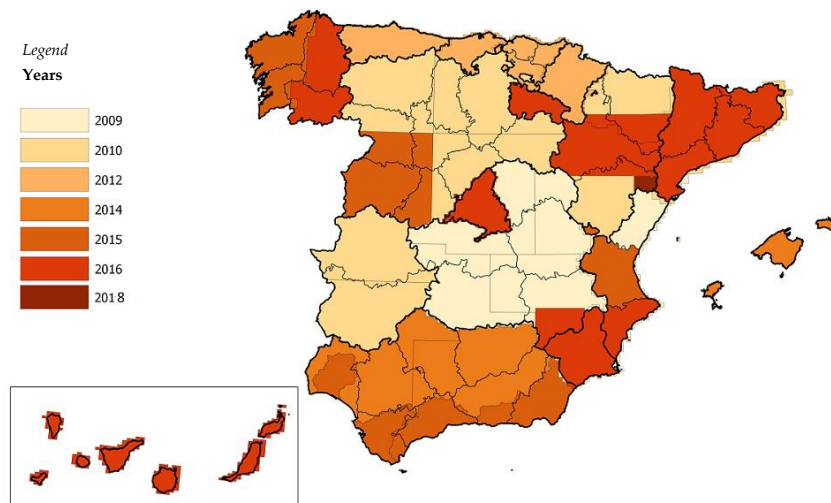
Two ALS-PNOA coverages have been acquired with low point densities. The first one covers all Spain with a nominal point density of 0.5 points  $\text{m}^{-2}$ . The second coverage has a higher point density of 1 point  $\text{m}^{-2}$  and has been only captured in some areas. Thus, the second coverage will continue during the next years as a 6 years ALS flight updating is considered. Both ALS surveys have been co-financed by the “Ministerio de Fomento”, the “Ministerio de Agricultura, Alimentación y Medio Ambiente”, the “Ministerio de Hacienda y Administraciones Públicas” and the Spanish “Comunidades Autónomas” (Figure 4).

A wide variety of applications can be developed using this ALS-PNOA data:

- Generation of high resolution DTMs.
- Generation of Surface models for characterizing urban areas, vegetation, electric lines, and road lines.
- Analysis of flooding areas.

- Hydrographic models.
- Automatic detection of surface and urban changes.
- Pasture admissibility within the common agricultural policy.
- Visibility analysis.
- Aspect mapping.
- Fuel models cartography.

The data from both coverages are available at the National Center for Geographic Information (CNIG <http://centrodedescargas.cnig.es>). The data, distributed in 2x2 km tiles in \*.laz format, is classified and coloured using Red Green Blue (RGB) orthophotography. The first coverage acquisition of ALS data and orthophotography was not concomitant, but the acquisition of both products was simultaneous for the second coverage. Furthermore, the ALS data have orthometric Z values being provided in ETRS89.



**Figure 4.** Most recent ALS-PNOA flight. Modified from: <https://pnoa.ign.es/estado-del-proyecto-lidar>.

The \*.laz file is a \*.las zipped file. The \*.las file is the public file format for the interchange of 3D point cloud data (ASPRS, 2019). Although the format was primarily developed for exchanging LiDAR data, supports the exchange of any 3D data type. The main information of ALS- PNOA data is the x, y and z coordinates of each point. Furthermore, the files include several attributes associated to each point. A description of the attributes according to \*.las format (ASPRS, 2019) is presented below:

- Intensity defines the quantity of energy scattered by each laser return. The intensity presents values from 0 to 255.
- Return number refers to the pulse return number for a given output pulse. The emitted laser pulses might have up to five returns per pulse, depending on the surface type and sensor capabilities between others. The first return is assigned as 1, the second return as 2 and so on.

- Number of returns emitted for a given pulse. For example, a point may be return number three within up to four total returns.
- Scan direction refers to the direction in which the scanner mirror was traveling at the time of the output pulse. The values range from 1 to 0. Positive values indicate that the scanner is moving to the left side of the in-track direction to the right side while negative values are the opposite.
- Edge of Flight line is a numeric identification to determine whether the point is located in the edge of the scanning flight line. When the point is the last one a given scan line before it changes direction or the mirror facet have a value of 1. On the other hand, when the value is 0 there will be no change of direction.
- Classification. When a point cloud is classified, the points are labelled according to the object that reach. The codification of ALS point clouds is defined by the ASPRS for \*.las formats 1.1, 1.2, 1.3 and 1.4. These are included in Table 3.

*Table 3. ASPRS Standard point classification.*

| Classification Value | Meaning                       |
|----------------------|-------------------------------|
| 0                    | Created, Never Classified     |
| 1                    | Unclassified                  |
| 2                    | Ground                        |
| 3                    | Low Vegetation                |
| 4                    | Medium Vegetation             |
| 5                    | High Vegetation               |
| 6                    | Building                      |
| 7                    | Low point (Noise)             |
| 8                    | Model Key-Point (Mass Point)  |
| 9                    | Water                         |
| 10                   | Reserved for ASPRS Definition |
| 11                   | Reserved for ASPRS Definition |
| 12                   | Overlap points                |
| 13-31                | Reserved for ASPRS Definition |

- Scan Angle Rank during the flight. This attribute determine the scan angle value expressed in degrees. The values range from -90 up to 90. Nadir pulses are codified with the value 0. Positive values refer to pulses located at the right side of the nadir and negative pulses are located at the left side of the nadir.
- User data. This field may be used at the user's discretion.
- Point Source ID. The value indicates the source from which a point originated. Values range from 1 up to 65,535 inclusive.

- GPS time indicates the date and time of GNSS laser point registration when the plane is flying. The time is expressed in seconds.

**Table 4.** Technical specifications of LiDAR-PNOA plan for Aragón region from the first and second coverage.

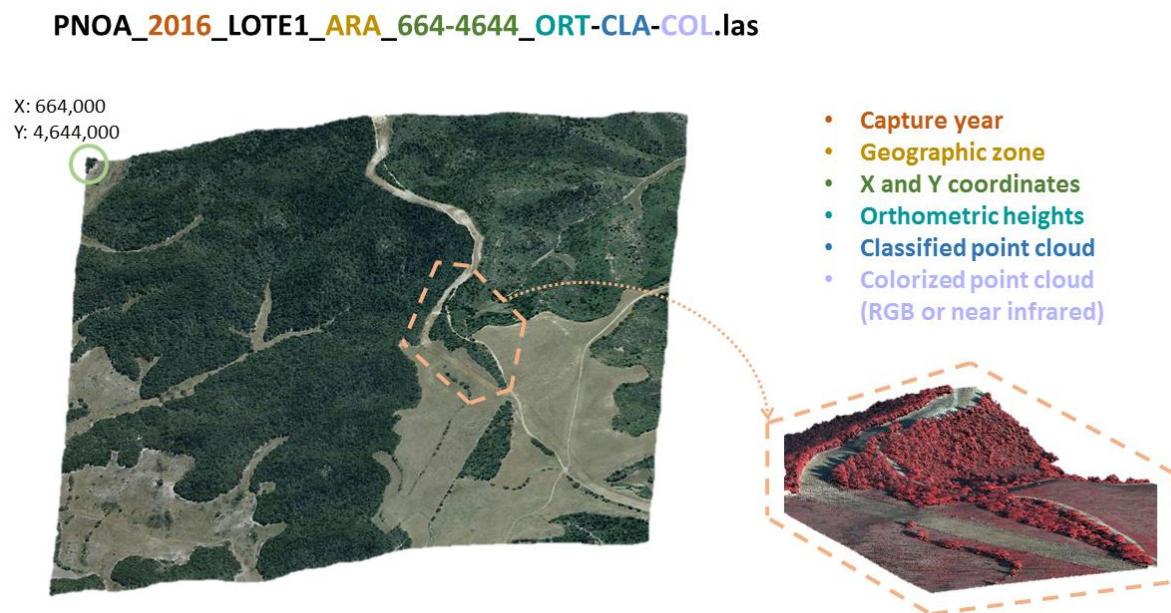
| Characteristic   | Description  |                 |
|--|--|-----------------|
|  | First coverage   | Second coverage |
| Sensor   | Leica ALS-50   | Leica ALS-80    |
| Geodetic reference system                              | ETRS89   |                 |
| Cartographic projection                                | Universal Transversal Mercator (UTM) 30 and 31   |                 |
| Geoid model  | EGM2008-REDNAP   |                 |
| Field of view (FOV)                                    | The maximum allowed FOV is 50°   |                 |
| Scanning frequency                                     | Minimum of 70 Hz and up to 40 Hz with a FOV of 50°   |                 |
| Pulse frequency  | Minimum of 45 kHz with a FOV of 50° and up to 3,000 m  |                 |
| Point density  | 0.5 points m <sup>-2</sup> implying a spacing between points ≤1.41 m   |                 |
| Radiometric resolution for multiple intensities        | Dynamic range with at least 8 bits   |                 |
| Ability to capture multiple returns for the same pulse | Up to 4 returns per pulse when the vertical distance is higher than 4 m  |                 |
| GNSS navigation system                                 | Double frequency GNSS with at least 2 Hz   |                 |
| Inertial system (IMU/INS)                              | Data register frequency ≥200Hz and drift <0.1° h <sup>-1</sup>   |                 |
| Plane velocity when LiDAR data capturing               | Variable   |                 |
| Flight height  | Variable   |                 |
| Maximum flight line length                             | 90 km  |                 |
| Horizontal and vertical precision after processing     | The horizontal global precision at nadir have a RMSE <sub>x,y</sub> <30 cm (1 sigma), while the vertical global position at nadir have a RMSE <sub>z</sub> <20cm (1 sigma). Errors up to 3xRMSE in dense vegetation or steep slopes might occur. The edge of flight line might present an error up to 2xRMSE |                 |
| Altimetry precision and maximum error                  | ≤0.40 m for 95% of the cases. Points cannot have an error higher than 0.60 m   |                 |
| Altimetry discrepancies between flight lines           | ≤0.40 m  |                 |
| Point cloud format                                     | *.las 2x2 km tiles   |                 |
| LAS file classification                                | Automatically classified (ground, vegetation, buildings, overlap)  | Non classified  |

Data from both ALS-PNOA coverages are utilized in this PhD Thesis. In this sense, more than 4,000 \*.las files from the first coverage were downloaded from the CNIG webpage. Furthermore, the Geographic Institute of Aragón (IGEAR) and CNIG provided 147 tiles from the second coverage before being included in CNIG platform.

The flight technical specifications, the point cloud characteristics and the file nomenclature is included in Table 4, Table 5 and Figure 5, subsequently.

*Table 5. Characteristics of LiDAR-PNOA data according to the study area.*

| Characteristics               | Zone A                        | Zone B  | Zone C                                      | Zone D  |
|-------------------------------|-------------------------------|---|---|---|
| Acquisition date              | October 2010 (first coverage) | October 2010, January 2011 and February 2011 (first coverage) | July 2010 to February 2011 (first coverage) | July 2010 to February 2011 (first coverage)<br>September to November 2016 (second coverage) |
| Sensor                        | Leica ALS50                   | Leica ALS50   | Leica ALS50                                 | Leica ALS50 (first coverage)<br>Leica ALS80 (second coverage)                               |
| Average flight height (m)     | 3146                          | 3022  | 3240  | 3138 (first coverage), 2943 (second coverage)   |
| Average plane velocity (km/h) | 241                           | 241   | 184   | 184 (first coverage), ~240 (second coverage)  |
| Number of tiles               | 42                            | 690   | 3800  | 147 (first coverage) and 147 (second coverage)  |



*Figure 5. Nomenclature of LiDAR-PNOA files.*

### **2.2.5. ALS pre-processing and metric computation**

Quality control of ALS point clouds is required for accurate pre-process the data. The ALS data provided by the PNOA is not raw data, but users should check the quality of downloaded point clouds before performing any further analysis. The first ALS processing step considered was the removal of noise return hits. ALS-PNOA data is classified and noise returns are denoted as class 7. These points were removed using LAStools software implemented in ArcGIS 10.5.

Overlapping strips might generate horizontal and vertical discrepancies in laser scanning surveys playing an important role in quality control (Vosselman, 2012). Two approaches could be considered when dealing with overlapping strips: removing those discrepant strips or using methods to compensate for discrepancies between two datasets in the overlapping areas (Latypov, 2002). In this PhD the validity of overlapping returns was verified by visualizing the 3D point clouds, generating reports about ALS characteristics and analysing the generated DEMs. The analysis performed determined that some tiles from the south of Aragón, named as “ARA\_SUR” present vertical and/or horizontal displacements. These displacements are classified with the number 12 and, consequently, were removed using the same procedure as noise return removal.

The ALS point clouds in forested environments contain returns from several surface objects, such as shrubs, trees, electrical wires and buildings that should be separated from ground returns (Montealegre et al., 2015a). The process of separating the ground and non-ground points, performed prior to DEM generation, is called filtering or classification. This is a key process in forestry applications, allowing normalizing point clouds to determine aboveground return heights. There are several filtering algorithms which can be classified in four types (Meng et al., 2010; Sithole & Vosselman, 2004): interpolation-based, slope-based, segmentation-based and morphological methods. Most filtering methods provide accurate results in flat and non-complex areas while steep slopes and complex forest stands increases classification errors (Sithole & Vosselman, 2004).

Previous research developed by Montealegre et al. (2015a) compared seven filtering methods implemented in open software in Aleppo pine forested areas with moderate to steep slopes. The study determined that Multiscale Curvature Classification (MCC) method, developed by Evans & Hudak (2007), was the most accurate to filter ALS-PNOA point clouds in these Mediterranean environments. Accordingly, in this PhD MCC algorithm, implemented in MCC-LiDAR software, was selected for classifying the point clouds.

MCC is an iterative-interpolation-based filter. The algorithm calculates an interpolated surface by using a thin-plated spline and discards the ALS returns that exceed a threshold curvature. MCC creates three scale domains defining three processing window sizes. The algorithm iterates in the three scale domains until the number of remaining returns changes by less than 1%, less than 0.1%, and less than 0.01%, respectively (Evans & Hudak, 2007). Two parameters must be defined for running MCC: the scale parameter (s) and the curvature threshold (t). The scale depends on object sizes and ALS point spacing, while the curvature threshold is related to curvature tolerance for a

scale domain. In this research, both parameters were determined according to Montealegre et al. (2015b). The scale was set to 1 m and the curvature threshold to 0.3.

ALS data create a random sampling of the terrain surface, being necessary to apply interpolation processes to generate a continuous surface (Vosselman & Maas, 2010). The selection of the appropriate interpolation algorithm and DEM spatial resolution is relevant as may constitute a source of inaccuracy in vegetation metrics prediction (Aguilar et al., 2006). Several interpolation methods, such as natural neighbour, Triangulated Irregular Network (TIN) to raster, kriging, point to raster, are commonly implemented in Geographical Information System (GIS) software.

Montealegre et al. (2015b) compared the suitability of six interpolation routines over a range of terrain roughness in Aleppo pine forested stands. The analysis determined that the TIN to raster interpolation method produces the best result when generating a DEM with 1 m resolution. Consequently, the TIN to raster routine, implemented in ArcGIS 10.5 software, was selected for interpolating and generating the DEMs in this research.

TIN to raster method includes two procedures. Firstly, the interpolation method build a terrain surface using Delaunay irregular triangles by connecting ALS returns. The elevation is recorded for each triangle node, while elevations between nodes can be interpolated to generate a continuous surface. Secondly, the TIN structure is transformed to a raster structure using natural neighbour interpolation from previous triangles nodes to generate a value at the center of each raster cell.

The statistical metrics derived from ALS data were related to field variables to fit regression models. The computation of ALS metrics requires the clipping of the point cloud to the spatial extent of each field plot, being performed using the “*ClipData*” command of FUSION LDV 3.60 open source software (McGaughey, 2009). Furthermore, the height of the point cloud returns were normalized using the DEM to compute a wide range of statistics using the “*Cloudmetrics*” command (Table 6). These statistical metrics are commonly used as independent variables in forestry (Evans et al., 2009).

The computation of ALS metrics normally requires the application of a threshold value to remove ground and/or understory laser hits. In this PhD Thesis, several tests were performed to determine the more suitable thresholds according to the type of estimated stand variable and application. A threshold value of 2 m height was selected according to shrub height agreeing with Nilsson (1996) and Næsset & Økland (2002) for predicting the analysed forestry metrics except from total biomass, that required the selection of a different threshold to include understory laser hits. The explored thresholds ranging from 0 up to 1 m determined that 0.2 is the most suitable one being related to the ALS sensor precision in coordinate Z, which has a root mean square error below 0.2 m.

Table 6 describes the computed ALS metrics structured in three main groups: metrics related with canopy height, metrics associated to canopy height variability and canopy density metrics. The

computed metrics might be selected to be used in modelling and all of them present a coherent relationship with vegetation structure (Evans et al., 2009; McGaughey, 2009).

**Table 6.** Derived metrics from ALS point clouds, where  $x_i$  is the height value of the return,  $N$  is the total number of observations,  $r_i$  is the return, and  $p$  is the pulse.

|                             | Metric   | Description  |
|-----------------------------|--|--|
| Canopy height metrics (CHM) | Percentiles of the return heights 1, 5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 95 and 99 ( $P_{01}$ , $P_{05}$ , $P_{10}$ , etc.) | <p>The percentiles were computed according to the following methodology:</p> $(N - 1)P = I + f \begin{cases} I \text{ is the integer part of } (N - 1)P \\ f \text{ is the fractional part of } (N - 1)P \end{cases}$ <p>where <math>N</math> is the number of observation and <math>P</math> is the percentile value divided by 100.</p> <p>if <math>f = 0</math> then Percentile value = <math>x_{i+1}</math></p> <p>if <math>f &gt; 0</math> then Percentile value = <math>x_{i+1} + f(x_{i+2} - x_{i+1})</math></p> <p>where <math>x_i</math> is the observation value considering that observations are ranked in ascending order.</p>  |
|                             | Minimum elevation  | $x_i \text{ minimum}$  |
|                             | Mean elevation   | $\frac{\sum_{i=1}^N x_i}{N}$   |
|                             | Mode elevation   | $x_i$ value more frequent in the plot  |
|                             | Elevation quadratic mean   | $\left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right)^{\frac{1}{2}}$  |
|                             | Elevation cubic mean   | $\left( \frac{1}{N} \sum_{i=1}^N x_i^3 \right)^{\frac{1}{3}}$  |
|                             | L moments ( $\lambda_1$ to $\lambda_4$ )   | $\hat{\lambda}_1 = \frac{1}{n} \sum_{i=1}^N x_{(i)}$ $\hat{\lambda}_2 = \frac{1}{2} \frac{1}{n} \sum_{i=1}^N \left( {}^{i-1}C_1 - {}^{n-1}C_1 \right) x_{(i)}$ $\hat{\lambda}_3 = \frac{1}{3} \frac{1}{n} \sum_{i=1}^N \left( {}^{i-1}C_2 - 2 {}^{i-1}C_1 {}^{n-1}C_1 + {}^{n-1}C_2 \right) x_{(i)}$ $\hat{\lambda}_4 = \frac{1}{4} \frac{1}{n} \sum_{i=1}^N \left( {}^{i-1}C_3 - 3 {}^{i-1}C_2 {}^{n-1}C_1 + 3 {}^{i-1}C_1 {}^{n-1}C_2 - {}^{n-1}C_3 \right) x_{(i)}$ <p>where <math>x_{(i)}</math>, <math>i=1, 2, \dots, n</math>, are sample values ranked in ascending order and</p> ${}^m C_k = \binom{m}{k} = \frac{m!}{k!(m-k)!}$ <p>is the number of combinations of any <math>k</math> items form <math>m</math> items and is equal to zero when <math>k &gt; m</math>.</p> |
|                             | Maximum elevation  | $x_i \text{ maximum}$  |



|  |  |  |
|--|--|--|
| Canopy height variability metrics (CHVM) | Standard deviation of point height distribution                              | $\sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$  |
|  | Variance of point height distribution ( $\sigma^2$ )                         | $\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$   |
|  | Coefficient of variation of point height distribution                        | $\frac{\sigma}{\mu} 100$   |
|  | Skewness of point height distribution  | $\frac{\sum_{i=1}^N (x_i - \mu)^3}{(N - 1)\sigma^3}$   |
|  | kurtosis of point height distribution  | $\frac{\sum_{i=1}^N (x_i - \mu)^4}{(N - 1)\sigma^4}$   |
|  | Interquartile distance of point height distribution                          | $[P_{75}(x) - P_{25}(x)]$  |
|  | Average Absolute Deviation of point height distribution                      | $\frac{\sum_{i=1}^N (x_i - \mu)}{N}$   |
|  | L moment coefficient of variation of point height distribution ( $\tau_2$ )  | $\frac{\lambda_2}{\lambda_1}$<br>$0 < \tau_2 < 1$  |
|  | L moment skewness of point height distribution ( $\tau_3$ )                  | $\frac{\lambda_3}{\lambda_2}$<br>$-1 < \tau_3 < 1$   |
|  | L moment kurtosis of point height distribution ( $\tau_4$ )                  | $\frac{\lambda_4}{\lambda_2}$<br>$\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1$  |
| Canopy density metrics (CDM)             | Percentage of first returns above a height-break, above the mean or the mode | $\frac{\sum_{i=1}^N r_{i \text{ first returns}} > \text{heightbreak}}{\sum_{i=1}^N r_{i \text{ first returns}}} 100$ |
|  | Percentage of all returns above a height-break, above the mean or the mode   | $\frac{\sum_{i=1}^N r_i > \text{heightbreak}}{N} 100$  |
|  | Canopy relief ratio  | $\frac{\mu - x_{i \text{ minimum}}}{x_{i \text{ maximum}} - x_{i \text{ minimum}}}$                                  |
|  | All returns above a height-break, above the mean or the mode x 100           | $\frac{\sum_{i=1}^N r_{i \text{ all returns}} > \text{heightbreak}}{\sum_{i=1}^N r_{i \text{ all returns}}} 100$     |

### 2.2.6. Optical data and spectral indices

Although the PhD Thesis mainly focus on the use of ALS-PNOA data for forestry applications, the combination with optical data allowed to estimate fire severity and, subsequently, determine the CO<sub>2</sub> emissions.

Wildfires constitute a socio-environmental hazard in pyrophyte Aleppo pine Mediterranean ecosystems. Fire severity usually reaches high levels, generating important changes in vegetation vigour, colour, water content as well as forest composition, structure and density. Optical remote sensing data have been widely used to characterize fire severity (García-Llamas et al., 2019) considering infrared spectral regions.

Landsat program is the longest middle resolution remote sensing program and provides coherent and continuous global data since 1972. Specifically, Landsat 8 was launched in 2013 carrying on board the Operational Land Imager (OLI) sensor. The mission presents a polar low earth orbit and 16 days of temporal resolution. The spatial resolution of the 9 optical spectrum bands is 30 m, except from the panchromatic band with 15 m resolution. The most valuable bands to characterize fire severity are number 5, which corresponds to the near infrared (NIR), and 7, which refers to the short wave infrared (SWIR). NIR is linked to foliar structure and morphology while SWIR is related to vegetation and soil water content (Soverel et al., 2010).

In this research, two images were selected to characterize pre and post-fire conditions and estimate fire severity. The pre-fire image was acquired on June 30 (path 200, row 31) 2015 and the post-fire image on July 9 (path 199, row 31) 2015. The selection of these images was conditioned by the need for temporary proximity between images in order to capture the variability of vegetation spectral response. The images were provided by the United States Geological Survey (USGS) and downloaded using Earth Explorer platform. Both images present a high processing level. Images were geometrically and radiometric corrected. Specifically, the Landsat Surface Reflectance Code (LaSRC) was applied according to Vermote et al. (2016). The algorithm uses a radiative transfer model, climate data from MODIS and the coastal aerosol band to perform those corrections.

The use of spectral indices, derived from the combination of different spectral bands, provides relevant information about surface properties (Chuvieco, 2010). The Normalized Burn Ratio (NBR) index, developed by Key & Benson (2006), relates Landsat band 5, from NIR, with band 7 from SWIR. NBR values range from -1 to 1. Surfaces with no presence of vegetation or low vegetation vigour present negative values while photosynthetically active areas show positive values. In this PhD Thesis, NBR was calculated for pre-fire and post-fire images according to equation 27. Then, the Differenced or Delta NBR index ( $\Delta\text{NBR}$ ) was computed to provide a quantitative measure of the burned area (equation 28).

$$\text{NBR} = (\rho_{\text{NIR}} - \rho_{\text{SWIR}}) / (\rho_{\text{NIR}} + \rho_{\text{SWIR}}) \quad (27)$$

$$\Delta\text{NBR} = \text{NBR}_{\text{prefire}} - \text{NBR}_{\text{postfire}} \quad (28)$$

where  $\rho_{\text{NIR}}$  (near infrared) and  $\rho_{\text{SWIR}}$  (short wave infrared) refer to bands 5 and 7 Landsat 8 OLI reflectance, respectively.

The  $\Delta\text{NBR}$  values were multiplied by 1,000 to provide a valid continuous range of values. Negative values are associated with fast regrowth from herbaceous, while positive values are related to different degree of fire severity.

### 2.2.7. Wildfire biomass loss and CO<sub>2</sub> emission estimation

Wildfires constitutes a relevant source of carbon monoxide emissions (Pétron et al., 2004) in the Mediterranean basin, which yearly records an average of 45,000 fires (Oliveira et al., 2012). The account of carbon dioxide emissions is relevant for understanding carbon cycling and provides information to develop climate regulation policies (Mieville et al., 2010). Thus, different approaches have been tested to estimate the emissions for wildfire as the one proposed by the Intergovernmental Panel on Climate Change (IPCC) (equation 29):

$$L_{fire} = A \times B \times C \times D \times 10^{-6} \quad (29)$$

where  $L_{fire}$  refers to the quantity of GHG released due to fire (tonnes of GHG),  $A$  is the burned area (ha),  $B$  is the mass of “available” fuel including biomass, ground litter and dead wood (tons ha<sup>-1</sup>),  $C$  is the combustion efficiency or fraction of combusted biomass (dimensionless),  $D$  is the emission factor (g kg<sup>-1</sup> of dry matter burnt).

A similar approach, proposed by De Santis et al. (2010), adapted the IPCC one to estimate fire GHG emissions, using remote sensing techniques. In this sense, four steps were required: the delimitation of the burned area; the estimation of pre-fire biomass; the assessment of the fraction of biomass consumed by the fire or burning efficiency, associated with fire severity and; the use of conversion factors to estimate GHG emissions.

Traditionally the estimation of pre-fire biomass was performed by using field data and allometric equations, while post-fire biomass was assessed either by visual examination (Roy et al., 2005) or field-based weighting (Sá et al., 2005). However, the implementation of remote sensing techniques provided new methodological approaches. In this research, ALS data (used to determine pre-fire biomass) and passive optical data (applied to estimate fire severity) were combined with the final objective of estimating biomass losses and, subsequently CO<sub>2</sub> emissions. The following five steps were conducted in this approach:

- i. Estimation of Aleppo pine pre-fire biomass using ALS data captured previously to the occurrence of the wildfire.
- ii. Estimation of fire severity.  $\Delta NBR$  index is computed using two Landsat 8 OLI images from June 30<sup>th</sup> and July 9<sup>th</sup> 2015.
- iii. Mapping of pre-fire Aleppo pine forest. The Spanish National Forest Map as well as a canopy height model derived from the ALS data were used to delimit the burned stands.
- iv. Selection of burning efficiency factors and biomass losses quantification. Most approaches consider that forest biomass is completely consumed by fire (French et al., 2004). However, more accurate approaches from our point of view consider different burn severity levels, related to different biomass losses. Thus, in this research three burning efficiency factors were applied to three different severity levels (De Santis et al., 2010). Key & Benson (2006) generic severity ranges were reclassified to match the three burning efficiency factors (Table 7). Low burning efficiency refers to low consumption of leaves and very low consumption of branches. Moderate burning efficiency denotes intermediate consumption

of leaves and moderate consumption of small branches. High burning efficiency corresponds to a complete consumption of leaves and high loss of small branches and twigs.

**Table 7.** Severity levels,  $\Delta NBR$  generic ranges defined by Key & Benson (2006) and burning efficiency factors used to estimate biomass losses following De Santis et al. (2010).

| Severity level         | $\Delta NBR$ range | Burning efficiency factors |
|------------------------|--------------------|----------------------------|
| Unburned               | -100 to +99        | 0.00                       |
| Low severity           | +100 to +269       | 0.25                       |
| Moderate-low severity  | +270 to +439       | 0.42                       |
| Moderate-high severity | +440 to +659       | 0.57                       |
| High severity          | +660 to +1300      |                            |

- v. Conversion of biomass losses to carbon content and, subsequently, to  $CO_2$  emissions. Biomass can be converted to carbon content by applying conversion factors. The most common factor is 0.5, but, in this PhD Thesis, a value of 0.499 was set for Aleppo pine, in accordance with Montero et al. (2005). These authors determined several conversion factors for Mediterranean species based in field work. Secondly, the conversion from carbon to  $CO_2$  was performed according to Trozzi et al. (2002) equation. This equation includes the variables proposed by Seiler & Crutzen (1980), while modifying the specific factors to better represent Mediterranean environments (equation 30)

$$CO_2 = \varepsilon * \delta * C \quad (30)$$

where  $\varepsilon$  is the fraction of total carbon emitted as  $CO_2$  (0.888);  $\delta$  is the factor of conversion from the emissions in ton of carbon to the emissions in ton of  $CO_2$  (44/12); and  $C$  is the carbon content.

## 2.2.8. Modelling: variable selection, regression methods and validation

### Variable selection

Variable selection includes a variety of methods to reduce the dimensionality problem as well as deal with large feature sets, improving model performance, reducing time and costs and generating understandable models (Guyon & Elisseeff, 2003). This PhD Thesis explored the performance of five selection processes to estimate forest stand variables: Spearman's rank correlation, Stepwise selection, Principal component analysis (PCA) and Varimax rotation, Last absolute shrinkage and selection operator (LASSO) and All subset selection. These methods are described below.

Spearman's rank correlation coefficient is generally denoted by the Greek letter  $\rho$  (rho) (Spearman, 1904). This coefficient determines the strength and direction of the relationship between two

variables being carried out using *“cor.spearman”* function in R environment. Although other correlation coefficients as Pearson or Kendall exist, Spearman’s rank coefficient has been widely used for forestry applications, showing a uniform power for linear and non-linear relationships. The rho coefficient varies between +1 and -1. The value +1 indicates a perfect positive degree of association between two variables while the value -1 denotes a perfect negative degree of association. The weaker the correlation, the closer to 0 is the value. Thus, the correlation value of 0 indicates a null relationship between the analysed variables. In this PhD, the selection of the ALS variables was made considering a minimum positive and negative rho value, ranging from 0.2 up to 0.5. Depending on the higher or lower strength of the relationship between the variables the range was changed to better reduce variable redundancy and multicollinearity.

Stepwise selection, or also called stepwise regression, allows selecting predictive variables in an automatic procedure (Efroymson, 1960). The method iteratively adds or drops variables at several steps to determine the best subset of variables, which denotes the best model performance and lower prediction error. The predictor, that is added or dropped in each step, is based in the Akaike information criterion (AIC). There exist three ways of performing stepwise regression: forward selection, backward selection and bidirectional stepwise selection. Forward selection begins with no variables in the model, iteratively selects the variables that contribute the most and stops adding predictors when the improvement is no longer statistically significant. Backward selection begin with all variables in the model, iteratively drops the one that contribute the least and stops dropping predictors when all the selected variables are statistically significant. Bidirectional stepwise selection combines forward and backward selections. Firstly, this method begins with forward selection and then drops those variables no statistically significant using backward selection. The stepwise selection was carried out using *“step”* function in R environment.

PCA is a dimension-reduction method that reduces large sets of predictors to a smaller size, keeping the majority of the information. Although is a classic statistical method that has been broadly used in many research fields, it is not a common approach in ALS metric selection (Silva et al., 2016). PCA was computed using R package *“lattice”* and specifically with the function *“prcomp”*, which uses the singular value decomposition to examine the covariance and correlations between individuals. According to the Kaiser Criterion, in his research the generated components with greater values than 0.1 were retained. Then, a Varimax rotation, which maximizes the sum of the variance, was applied to better interpret the PCA results (Darlington & Horst, 1966; Kaiser, 1958).

LASSO is a method that performs two processes: regularization and feature selection. The method was popularized and improved by Tibshirani (1996). This technique applies a shrinking or regularization process to penalize some of the coefficients to zero by using a penalty term, which refers to the sum of the absolute coefficients. Those variables that have a non-zero coefficient after regularization will be selected to be part of the model, minimizing prediction errors and generating interpretable models. LASSO was computed in R using the *“glmnet”* package.

All subset selection methods allow a suitable group of metrics being selected, while disregarding

the rest of the variables (Miller, 2002). Concretely, best subsets regression, being part of all subset selection methods, test all possible combination of variables and selects the best ones according to statistical criteria. Several search approaches have been developed for best subset selection. In this sense, four search approaches were tested: exhaustive, backward, forward and sequential replacement. Exhaustive approach uses efficient branch-and-bound algorithms to search for the best subsets of variables. Backward and forward algorithms have been previously described when using stepwise selection. Sequential replacement (Seqrep) is a synonym of stepwise selection combining forward and backward selection. The four explored approaches was implemented using R package “*leaps*” and specifically “*regsubsets*” function, which includes a wrapper that improves leaps functionalities. Furthermore, a maximum number of model variables can be determined.

### ***Regression methods***

The regression analysis constitutes the process of estimating or modelling the relationship between variables, which can be performed using either parametric or non-parametric regression methods. Although the application of non-parametric methods have increased popularity (Bollandsås et al., 2013b; Chirici et al., 2008; Liaw & Wiener, 2002), few studies focused on the comparison of different non-parametric algorithms respect to parametric methods. Furthermore, the majority of studies were performed using high-density point clouds (Latifi et al., 2010; Gleason & Im, 2012; Gagliasso et al., 2014; García-Gutiérrez et al., 2015; Görgens et al., 2015). The effect of using parametric or non-parametric regression methods on model accuracy for estimating stand variables with ALS data, as well as, the effect of dataset size on model performance was also addressed. Specifically, this PhD Thesis explored the performance of eight regression methods to estimate forest stands variables: a multivariate linear regression model, two machine-learning algorithms and five regression trees structures. These methods are described below.

Multivariate linear regression (MLR) is a parametric method, widely employed to estimate forest stand variables using ALS data (Means et al., 1999; Næsset & Økland, 2002; Watt et al., 2013). This method has the same structure as simple linear regression, but presents a higher number of predictors and builds the predictions based on a hyperplane (Mardia et al., 1979). Model coefficients for the predictors are computed considering the minimum quadratic differences between the observed and predicted values (equation 31):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (31)$$

where  $Y$  refers to the independent variable,  $\beta_0$  is the constant value,  $\beta_n$  is the regression coefficient for the predictor or independent variable  $X_n$ .

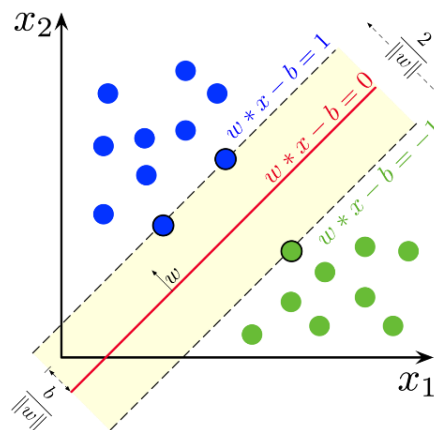
The computation of parametric methods requires the verification of the assumptions of linear regression models. In accordance with García et al. (2012) six assumptions were tested:

- Normality of the residuals. Residuals are normally distributed with an average value of zero. Normality was tested using Shapiro-Wilk and Kolmogorov-Smirnov statistical tests.

- Homoscedasticity. Residual variance should be constant for each independent variable. The homoscedasticity was tested using the Breusch-Pagan statistical test.
- Linearity between the dependent and independent variables is crucial for parametric regression methods. Linearity was verified using the RESET test.
- Independence. Residuals should be independent between them as random variables. Durbin-Watson test was used for confirming independence assumption.
- No-collinearity. There is not lineal relationship between the independent variables. The variance inflation factor (VIF) was computed to test no-collinearity assumption.
- Absence of atypical values. The presence of these values biases model equation. The Bonferroni and Cook distance test were used to verify the assumption.

Sample size was also considered when modelling. According to Hair et al. (1999) one predictor variable was included for each 15 to 20 samples to avoid overfitting. Furthermore, logarithmic transformation of dependent and independent variables were explored in the cases where statistical hypothesis of linear regression models could not be fulfilled (García et al., 2012; Means et al., 1999). Finally, the improvement of the goodness of fit of the models was also verified.

Support vector machine (SVM) is a supervised learning model that has associated learning algorithms to analyse and recognize patterns. SVM can be applied either for classification or regression analysis. This method considers that input data are separable in space (Mountrakis et al. 2011). SVM builds a hyperplane or a set of hyperplanes in an infinite-dimensional space. The method tries to find the optimal separation between classes determining those hyperplanes with maximum separability (Figure 6). The data located in the hyperplane are the most difficult to classify since they have lower separability between classes; they are called support vectors. A SVM model requires the definition and optimization of several parameters: type of kernel, cost and gamma of the kernel function. Two types of kernels were computed in this PhD Thesis: linear and radial. SVM was implemented by using R package “e1071”. In both SVM models, the parameter cost was defined in the interval 1-1,000, and the parameter gamma in the interval 0.01-1, applying the best parameters after tuning the model.



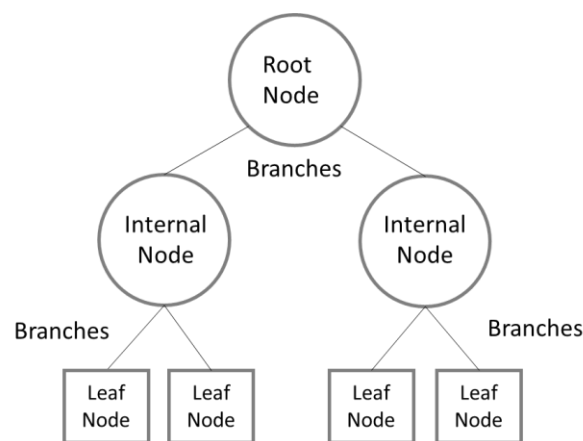
**Figure 6.** SVM example (source: <https://en.wikipedia.org>).  $x$  is a  $p$ -dimensional real vector and  $w$  is the normal vector to the hyperplane.

Random forest (RF) is an ensemble learning method that uses decision trees as base classifier. RF is a non-parametric method that can be either applied for classification or regression.

RF combines a decision tree that depends on the values of a random vector sampled, independently and with the same distribution for all trees in the forest (Breiman, 2001). The algorithm adds randomness to bagging and increase the diversity of decision trees by growing them from different subsets. In each decision tree, RF divides the nodes by using the best variables from a random sample. RF was implemented through R package “randomForest” (Liaw & Wiener, 2002) and “caret” (Van Essen et al., 2001).

The RF model was adjusted using two parameters: the number of trees to growth (ntrees) and the number of variables selected randomly at each split (mtry). They were in the intervals 1-3,000 and 1-3, respectively.

Decision trees method partitions the data into branches and then continues subdividing the data into smaller groups in a process called recursive partitioning. In simpler words, the space is divided in simpler regions that are more manageable (Quinlan, 1987). There are two types of decision trees: classification trees and regression trees. Classification trees are associated to categorical variables while regression trees are linked to continuous target variables. The trees have different components: the root node of the tree, the branches (segments of the tree that connect the nodes), the internal nodes (where the predictor space is split) and the terminal nodes (Figure 7). Five regression trees structures based on “If-Then” rules were computed using the R package “CORElearn”. The differences between them, described below, refer to the regression model considered in the leaf nodes.

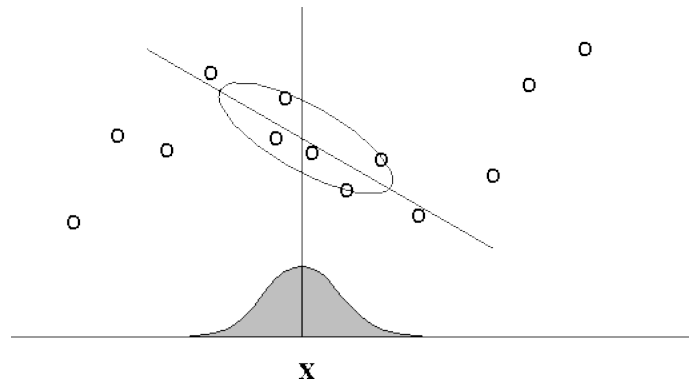


**Figure 7.** Components of a decision tree.

Locally weighted linear regression (LWLR), locally weighted regression (LWR), or loess, is a non-parametric method that uses local functions to make predictions. LWR divides the space and creates a local model based on neighbouring data for each point of interest. The local model is a linear regression weighted by a tricubic kernel (Cleveland & Devlin, 1988) centered at  $x$  (Figure 8). Thus, the dependent variable is smooth by a function of the independent variables. The coefficient



of smoothness is fitted by computing weighted mean square error and considering a distance function (tricubic kernel).



**Figure 8.** LWR example (source: <http://www.cs.cmu.edu/>).

Linear model with a minimum length principle (MDL) is based on the rule developed by (Rissanen, 1978). The decision tree searches for regularities in a set of data that can be used to compress the data by using fewer symbols, from a finite alphabet, than needed to explain the data faithfully.

Reduced linear model (RLM) is an implementation of the M5' in "CORElearn" R package. M5' was developed by Wang & Witten (1996) as a slight modification of the M5 decision tree invented by Quinlan (1987). The tree model splits the data based on the standard deviation of the class values that reach a node. Then, a linear model is calculated for each node of the unpruned tree. Finally, the model is simplified using an exhaustive search to remove those variables that contribute little to the model in order to minimize the estimated error.

K nearest neighbor (KNN) is a non-parametric and one of the simplest machine learning algorithms. KNN belongs to the supervised learning category and is a lazy learning method. Lazy learning methods generalize the data during the testing phase while reducing the training phase being easily adaptable to changes. KNN is based on the k-NN algorithm developed by Fix & Hodges (1951), which includes two phases. Firstly, k nearest neighbours are searched using the complete dataset and considering an established Euclidean distance. Then, the mean of the k-most similar instances is used for the prediction.

Weighted k nearest neighbors (WKNN) is a refinement or extension of the KNN algorithm. The algorithm considers that the observations that are particularly close to the new observation should have a higher weight, which is calculated according to the distance between observations. In this sense, it gives greater weights to the closer neighbours. The weighted mean of the k nearest neighbours is used for the prediction (Hechenbichler & Schliep, 2014).

### ***Model accuracy and comparison***

Model validation is essential to verify the accuracy and credibleness of a model when working

with different datasets. The diverse validation methods found in literature are suited to the dataset size. Leave-one-out cross-validation (LOOCV) and k-fold validation are better suited to smaller to medium size samples while validation by splitting the sample into training and test sets is associated to medium to large samples. According to the number of field plots two types of validations were computed in this PhD Thesis: LOOCV was used for small samples while validation by splitting the sample was used when the sample was bigger.

LOOCV is one of the most common techniques used with small dataset size in order to not further reduce the sample (Andersen et al., 2005). This cross-validation method drops one sample from the dataset –one field plot- in each iteration, predicting the values using all the other samples from the dataset. The number of iterations is equal to the number of samples in the dataset. An average value of model performance, generated for the coefficients of each iteration, is computed being expressed as RMSE, %RMSE, Bias or  $R^2$  statistics. The comparison between the initial and the validated model denotes credibility of the model when the statistical values are similar.

Validation by splitting the sample is generally computed for medium to large datasets sizes. In this research, the original sample was split into a training set of 75% of the cases and a test set of 25% of the cases. The model uses the training set to test model variables and parameters and the performance of the model is evaluated with a different data set, the test dataset. The existence of similar values in the models generated with the training and test datasets generates a credible and robust model. The presence of differences between training and test might be caused by overfitting in the training phase of modelling.

In this PhD Thesis, following the recommendation from García-Gutiérrez et al. (2015), LOOCV or validation by splitting the sample was executed 100 times to increase results robustness in those methods with higher randomness.

Several statistics of goodness of fit can be computed to describe model performance. In this research root mean square error (RMSE) (equation 32), relative RMSE respect to the mean (%RMSE) (equation 33), bias (equation 34) and  $R^2$  after validation (equation 35) were calculated. The comparison between models was performed using RMSE, %RMSE and bias in all the regression methods, while  $R^2$  was not applied with non-parametric models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (32)$$

$$\%RMSE = \frac{RMSE}{\bar{y}} \times 100 \quad (33)$$

$$Bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (34)$$

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} 100 \quad (35)$$

where  $y_i$  is the observed value for plot  $i$ ;  $\hat{y}_i$  is the predicted value for sample plot  $i$ ;  $n$  is the number of plots and  $\bar{y}$  is mean observed value for all plots.

The use of the abovementioned statistics allowed us to compare the performance between the generated models, but it did not established whether the models were equivalent or there were statistically significant differences between them. Several statistical tests are design to determine these assumptions, being associated to the type of model (parametric or non-parametric). In this sense, Friedman non-parametric test was selected to determine whether models were equivalents and, subsequently, Nemenyi post-hoc test was used to determine whether the differences between the models were statistically significant. These methods were chosen, as proposed by Stojanova et al. (2010), being suitable to compare non-parametric models with a multiple relationship (one to many). The application of both tests was carried out using the RMSE model values for each dataset sample.

Friedman test allows testing differences between groups when the dependent variable is continuous, as the case of stand variables. It is considered an alternative to ANOVA test for parametric models. This method ranks each row –field plot- together and considers the values of ranks by columns –models-. The null hypothesis determines that models are equivalent, while the alternative hypothesis denotes difference between models.

In those cases of rejection of null-hypothesis in Friedman test, implying that the models are not equivalent, the Nemenyi (1963) post-hoc test was used. This pair-wise test is applied after a multiple comparison test such as Friedman test. This method determines whether the differences between the models were statistically significant, with a significance level of 0.05.

### **2.2.9. Temporal transferability**

Two approaches exist in ALS literature for modelling forest stand variables using multi-temporal ALS data: direct and indirect approaches. The direct approach fits one model for one point in time and estimates the forest variable in another point in time (Cao et al., 2016; Zhao et al., 2018). The main hypothesis of direct approach is that models can be imputed across time, in the same way they can be imputed across space. On the other hand, the indirect approach estimates the forestry variables in different dates by adjusting different models in each point in time (Næsset & Gobakken, 2005).

The main benefit of direct approach is that reduces modelling time and fieldwork cost, allowing model temporal transferability. The generation of models requires the selection of variables and the subsequent model parametrization, constituting relevant tasks for the prediction of forest variables. The model that is going to be temporally transferred is generated by relating ALS data with ground-truth data for one point in time, while second fieldwork campaign might be reduced or even not needed when the time between the ALS surveys is not large (Noordermeer et al., 2018). As stated by Fekety et al. (2015), integrating temporally disparate inventory data using multi-temporal ALS acquisitions could reduce inventory costs. This approach benefits not only forest managers but also enterprises devoted to forest inventories. Furthermore, this methodology could serve as an alternative to design the temporal gap between ALS flights to predict accurately forest variables over time.

The indirect approach requires higher cost of modelling and data capturing. Thus, we found different results of the evaluation of the two approaches in the literature (Næsset & Gobakken, 2005; Noordermeer et al., 2018; Zhao et al., 2018). Some authors get slightly better performance of the direct approach in the prediction of biomass and carbon fluxes (Bollandsås et al., 2013; Cao et al., 2016; Skowronski et al., 2014) while Meyer et al. (2013) and Zhao et al. (2018) achieved better results with the indirect approach.

This PhD Thesis explores the benefits of the two mentioned approaches in the prediction of seven forestry attributes at regional scale: stand density, basal area, squared mean diameter, dominant diameter, tree dominant height, timber volume and total tree biomass. We performed a comparison of the direct and indirect approaches to assess temporal transferability. Firstly, following the indirect approach, two different models were fitted for the available ALS-PNOA data (2011 and 2016), estimating the stand attributes for each point in time, using different ALS-metrics and model parameters. In this approach inventory data was updated using single-tree-growth models to generate concomitant information to the ALS-PNOA years (see Section 2.2.3). Secondly, the direct approach was tested. The models, fitted for one point in time, were extrapolated to the other point in time using the same variables and model parameters. The process was performed two times: the 2011 models were extrapolated to 2016, and inversely the 2016 models were extrapolated to 2011. The modification of the original methodology tested the validity of multi-temporal ALS data to perform future or retrospective analyses.

#### **2.2.10. Variable influence assessment**

ALS sensors have different configurations that determine the final data characteristics and quality of the point clouds. These configurations are normally different from one flight to another, varying according to the specific sensor capabilities, and might have an effect in model accuracy. In this sense, the effect of three ALS characteristics (point density, scan angle, canopy penetration pulse) and two environmental conditions (slope and shrub presence) on the prediction of forest residual biomass have been addressed.

The effect of point density has been explored in the prediction of different forest inventory attributes such as height, basal area or volume (Roussel et al., 2017; Gobakken & Næsset, 2008). However, less studies have considered point density effect on biomass prediction (García et al., 2017; Singh et al., 2015; Ruiz et al., 2014). The effect of scan angle on tree height prediction has been analysed by Disney et al. (2010), Holmgren (2004), Liu et al. (2018) and Montaghi (2013). Disney et al. (2010) proposed to minimize the use of data collected at scan angles greater than  $\sim 15^\circ$ . According to Holmgren (2004) and Liu et al. (2018), the prediction of canopy closure is affected by the scan angle, being necessary to avoid off-nadir angles from  $23$  to  $38^\circ$ . Similar results were concluded by Montaghi (2013) in the prediction of forestry metrics using an area based approach, determining that scan angles higher than  $20^\circ$  had a great effect in forest parameters predictions.

The structural characteristics and ALS flight settings modify canopy pulse penetration (CPP), decreasing DTM accuracy (Cowen et al., 2000; Hollaus et al., 2006; Hyypä et al., 2000). The higher

the density of the forest, the lower CPP rates are found (Hollaus et al., 2006). Secondary effects of densely covered areas would be a decrease in CPP in lower strata (Chasmer et al., 2006b; Wasser et al., 2013) and the decrease of DTM accuracy (Cowen et al., 2000; Hollaus et al., 2006; Hyypä et al., 2000). In contrast, leaf off conditions in deciduous forest improves CPP rates (Hill et al., 2009; Wasser et al., 2013).

The presence of steep slopes reduces the accuracy in tree height prediction (Breidenbach et al., 2008; Clark et al., 2004; Ørka et al., 2018), as well as in tree diameter, basal area, number of stems and volume (Ørka et al., 2018). Furthermore, decreases the ability to detect tree tops (Khosravipour et al., 2015). Although ALS predictions are affected by the increase of slope, the detected effect was not severe. Slope effect might be partially explained by the lower accuracy of DTMs in these areas, considering that filters have more difficulties on determining ground points on steep slopes (Montealegre et al., 2015a).

The effect of shrub presence has not been previously analysed. The hypothesis that the presence of shrub might affect the prediction of forest residual biomass is associated to the lower CPP in those areas and, consequently, a lower number of ground returns, which may affect DTM generation.

Two methodological approaches were tested to analyse the effect of the five abovementioned variables in forest residual biomass prediction. Firstly, a graphical assessment using boxplots, including the average mean prediction error per defined class and variable was applied. In addition, several statistical tests, described below, were applied to determine whether the differences between model performances were statistically significant.

The defined classes for each analysed variable are the following:

- Point density. Plots were classified into two classes: up to 1 point m<sup>-2</sup> and higher than 1 point m<sup>-2</sup> according to Montealegre et al. (2015b). Jakubowski et al. (2013) and García et al. (2017) also concluded that this threshold allows to have relatively high accuracies in forest parameters modelling.
- Scan angle. Three scan angle classes were tested. The first class was set close to nadir with average scan angle of up to 5°. The second class was defined as plots with an average scan angle between 5 and 15°. The third class included those plots with an average scan angle higher than 15°. The breakpoint between the second and third classes was defined by the maximum average scan angle established by the PNOA mission in order to reach the minimum nominal point density specified by the project.
- Canopy pulse penetration. The influence of four categories were analysed: 0%-25%, 25%-50%, 50%-75% and 75%-100% following Montealegre et al. (2015b). The proportion of the pulses that penetrate canopy and reach the ground was calculated using a ground tolerance of 2 m.
- Terrain slope. The effect of the terrain slope was assessed using two categories: smooth slopes of up to 15% and steep slopes higher than 15%.
- Shrub presence was binary categorized in plots with and without understory.

The mean predicted error (MPE) was selected to analyse the differences between the established classes. MPE was obtained for each field plot and, afterwards, the average mean per class was computed (equation 36). Furthermore the percentage of mean predicted error (%MPE) respect to the observed mean value was computed according to equation 37.

$$MPE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (36)$$

$$\%MPE = \frac{MPE}{\bar{y}} \times 100 \quad (37)$$

where  $y_i$  is the observed value for plot  $i$ ;  $\hat{y}_i$  is the predicted value for sample plot  $i$ ;  $n$  is the number of plots and  $\bar{y}$  is mean observed value for all plots.

Previous to significance analysis, normality and homogeneity test were computed. After considering logarithmic and square root transformation we concluded that variables were not normally distributed. In this sense, non-parametric Mann-Whitney and median tests were applied for analysing differences between two categories. Kruskal Wallis test was applied for those analyses with more than two classes.

The Mann-Whitney test is considered the main non-parametric alternative to the independent sample t-test. This test is designated to compare two populations. The null hypothesis of the test establishes that both samples come from different populations, while the alternative to the null hypothesis indicates that both samples come from the same population.

The median test is a non-parametric one used to determine whether two or more samples differ in their central tendency or median value, consequently it can be inferred whether the samples are from the same population or not. The null hypothesis establishes that both samples come from the same population.

Kruskal Wallis test is considered the main non-parametric alternative to the One Way ANOVA test. This method is a rank-based test that determines whether the medians of two or more groups are different. The null hypothesis considers that the samples are from the same population, while the alternative hypothesis establishes that at least one of the samples come from a different population.

#### **2.2.11. Mapping of forest variables**

The mapping of forest stand variables have been traditionally performed by linking tree metrics to the estimated stand variable using allometric equations and, subsequently, extrapolating these estimates at stand-level or regional scale (Boudreau et al., 2008). The use of remote sensing tools provides a greater overview of large areas and with higher temporal resolution (Castro et al., 2003). The use of optical data and, specially, the characterization of disturbance history with temporal series, have been proposed for estimating forestry variables (Cohen et al., 1996; Coops & Waring, 2001). However, the information captured by passive optical sensors tends to saturate under closed canopy conditions and in dense forests (Lu, 2006). Although active SAR remote

sensing improved the accuracy on forestry predictions (Tanase et al., 2014), there are still difficulties in heterogeneous and dense forests (Hyde et al., 2006).

ALS data have been proven as the most suitable technique for mapping 3D structure (Zhao et al., 2018). However, the use of ALS data at regional scales is still limited by the acquisition costs, the absence of global coverage and the huge data volumes that need to be processed. The use of ALS data for sampling some areas or by creating strips and the subsequent fusion with passive optical data have been explored for estimating forest stand variables at regional scales (e.g.: Matasci et al., 2018; Pflugmacher et al. 2014). Recently, the increase of ALS data at regional scale or country level, as the case of Spain, has opened new opportunities. In this sense, although computational effort is required, the improvement in hardware and software has allowed processing time reduction. In this sense, this PhD Thesis explored the use of ALS data at regional scales, providing cartographic information to forest managers.

In this research, the generation of cartography implied the prediction of a dependent variable for the whole study area by applying a regression model previously generated with the above mentioned methodology. Pixel size is one of the most relevant factors to be considered in mapping. The determination of pixel size should consider field plot size and ALS point density, but might be modified to account for specific requirement of forest managers. Commonly, pixel size is similar to field plot size. An increase in pixel size might decrease mapping accuracy when the number of ALS returns per pixel is low. Furthermore, independent variables that are included in regression models should be converted to raster format. In this sense, FUSION software, specifically designated to work with ALS data for forestry purposes, was used to generate the raster files using “*Gridmetrics*” and, subsequently “*CSVGrid*” commands. The generation of the raster for the dependent variable was performed in an R environment following three steps: running the best selected model; reading the independent variables in raster format; spatializing the results.





### 3. Research contributions

*The papers that constitute the PhD Thesis body are entirely included in this chapter as required by compendium PhD Thesis type. The papers provide different forest applications using ALS-PNOA data and field surveys: the estimation of CO<sub>2</sub> emissions generated by wildfires, several biomass fractions and inventory variables. Furthermore, the papers compare several selection and regression methods within different conditions, analysing the effect of ALS characteristics and environmental variables in modelling error and assessing temporal model transferability.*



### 3.1. Comparison of regression models to estimate biomass losses and CO<sub>2</sub> emissions using low density airborne laser scanning data in a burnt Aleppo pine forest





Comparación de modelos de regresión para estimar las pérdidas de biomasa y emisiones de CO<sub>2</sub> utilizando datos de escáner láser aeroportado de baja densidad en bosques de Pino carrasco afectados por el fuego

#### RESUMEN

El conocimiento de la pérdida de biomasa forestal producida por un incendio puede ser de utilidad para la estimación de las emisiones de los gases de efecto invernadero a la atmósfera. Este estudio se centra en la estimación de la pérdida de biomasa y las emisiones de CO<sub>2</sub> por la combustión de masas forestales de Pino carrasco en un incendio ocurrido en el municipio de Luna (España). La disponibilidad de datos de escáner láser aeroportado (ALS) de baja densidad permitió estimar la biomasa arbórea pre-fuego. Se realizó una comparación de nueve modelos de regresión con objeto de relacionar la biomasa, estimada en 46 parcelas de campo, con distintas métricas extraídas de los datos ALS. El método de regresión lineal multivariante seleccionado como óptimo, incluyó entre las variables independientes el porcentaje de primeros retornos sobre 2 m y el percentil 40 de la altura de los retornos. El modelo se validó utilizando una técnica de validación cruzada "*leave-one-out-cross-validation*" (RMSE: 6.1 ton ha<sup>-1</sup>). Las pérdidas de biomasa se estimaron utilizando una aproximación en tres fases: (i) la severidad del incendio fue obtenida utilizando el índice de diferencia normalizado ( $\Delta NBR$ ), (ii) los pinares de Pino carrasco se delimitaron utilizando el Mapa Forestal Nacional y datos ALS y, (iii) tres factores de eficiencia de combustión se aplicaron considerando los niveles de severidad. La biomasa post-fuego se transformó en emisiones de CO<sub>2</sub> (426.754,8 ton). Este estudio evidencia la utilidad de los datos ALS de baja densidad para estimar de forma precisa la biomasa pre-fuego y evaluar las emisiones de CO<sub>2</sub> en masas forestales mediterráneas de Pino carrasco.



## Comparison of regression models to estimate biomass losses and CO<sub>2</sub> emissions using low-density airborne laser scanning data in a burnt Aleppo pine forest

Darío Domingo <sup>a</sup>, María Teresa Lamelas-Gracia <sup>a,b</sup>, Antonio Luis Montealegre-Gracia <sup>a</sup>  
and Juan de la Riva-Fernández <sup>a</sup>

<sup>a</sup>GEOFOREST-IUCA, Departamento de Geografía y O.T., Universidad de Zaragoza, Zaragoza, Spain; <sup>b</sup>Centro Universitario de la Defensa de Zaragoza, Academia General Militar, Zaragoza, Spain

### ABSTRACT

The knowledge of the forest biomass reduction produced by a wildfire can assist in the estimation of greenhouse gases to the atmosphere. This study focuses on the estimation of biomass losses and CO<sub>2</sub> emissions by combustion of Aleppo pine forest in a wildfire occurred in the municipality of Luna (Spain). The availability of low point density airborne laser scanning (ALS) data allowed the estimation of pre-fire aboveground forest biomass. A comparison of nine regression models was performed in order to relate the biomass, estimated in 46 field plots, to several independent variables extracted from the ALS data. The multivariate linear regression selected model, including the percentage of first returns above 2 m and 40th percentile of the return heights, was validated using a leave-one-out cross-validation technique (6.1 ton/ha root mean square error). Biomass losses were estimated in a three-phase approach: (i) wildfire severity was obtained using the difference normalized burn ratio ( $\Delta NBR$ ), (ii) Aleppo pine forest was delimited using the National Forest Map and ALS data and (iii) burning efficiency factors were applied considering severity levels. Post-fire biomass was then transformed into CO<sub>2</sub> emissions (426,754.8 ton). This study evidences the usefulness of low-density ALS data to accurately estimate pre-fire biomass, in order to assess CO<sub>2</sub> emissions in a Mediterranean Aleppo pine forest.

### ARTICLE HISTORY

Received 31 January 2017  
Revised 22 May 2017  
Accepted 24 May 2017



### KEYWORDS

ALS; aboveground tree biomass losses; CO<sub>2</sub> emissions; Aleppo pine; burn severity

## Introduction

Wildfires are a socio-environmental hazard in Mediterranean ecosystems, acting as a source of greenhouse gases (GHGs) emissions to the atmosphere (Akagi et al., 2013; Andreae et al., 1988; Seiler & Crutzen, 1980; Van Der Werf et al., 2010; Wiedinmyer et al., 2011). Consequently, fires are able to alter the carbon cycle behaviour at regional or even global scales (Narayan, Fernandes, Van Brusselen, & Schuck, 2007), as well as to decrease the effect of carbon sequestration by forest ecosystems (Van Der Werf et al., 2006; Wiedinmyer & Neff, 2007). In the Mediterranean basin, an average of 45,000 fires is recorded yearly (Oliveira, Oehler, San-Miguel-Ayanz, Camia, & Pereira, 2012), increasing the albedo and determining the current landscape (Pausas, Llovet, Rodrigo, & Vallejo, 2008). Although these values and the resulting emissions are variable in time and space, biomass burning contributed significantly in the total direct carbon monoxide (CO) emissions (Pétron et al., 2004). These natural or anthropogenic disturbances might be enhanced by climate change, increasing fire risk (Moriondo et al.,

2006) particularly in summer months (Sebastián-López, Salvador-Civil, Gonzalo-Jiménez, & SanMiguel-Ayanz, 2008). In Spain, fire statistic registers show a reduction in the number of fire events during the last decade (2001–2010), as well as in the total burned area (Rodrigues, Ibarra, Echeverria, Perez-Cabello, & de la Riva, 2014; San-Miguel-Ayanz et al., 2012). However, the occurrence of large fires (>500 ha) has increased. In 2015, 39% of the total area affected by fires was burned in a large fire (MAGRAMA, 2016a). Moreover, if fire recurrence is high, regeneration process might fail for even species with high resilience such as Aleppo pine (*Pinus halepensis* Mill.) (Pausas et al., 2008), influencing carbon sequestration. Under this context, scientists, fire managers and decision-makers require the most accurate information available related to fire emissions and its impact on the environment and population. The account of carbon dioxide (CO<sub>2</sub>) emissions is essential for climate regulation policies and the evaluation of the effects of these policies (Mieville et al., 2010), as well as for understanding the services that forest provide to societies (Lal, 2008; Pan et al., 2011).

**CONTACT** Darío Domingo  [ddomingo@unizar.es](mailto:ddomingo@unizar.es)  GEOFOREST-IUCA, Departamento de Geografía y O.T., Universidad de Zaragoza, C/Pedro Cerbuna 12, 50009, Zaragoza, Spain

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

GHG emissions from fires estimation require (i) the delimitation of the burned area, (ii) the estimation of pre-fire biomass, (iii) the assessment of the fraction of biomass consumed by fire, also defined as burning efficiency (De Santis, Asner, Vaughan, & Knapp, 2010) and (iv) the use of conversion factors to estimate GHG emissions. However, little research has been conducted on quantifying pre-fire biomass and biomass consumed by fire. According to De Santis et al. (2010), biomass consumption was traditionally estimated using a two-step methodology which includes (i) the estimation of pre-fire biomass by applying allometric regression equations using destructive sampling or biomass values per species and (ii) the post-fire biomass estimated by field-based weighting (Prasad et al., 2001; Sá, Pereira, & Silva, 2005; Ward et al., 1996) or by visual examination (Roy, Jin, Lewis, & Justice, 2005). An alternative approach is based on the use of remote sensing imagery for pre-fire biomass estimation. Despite the wide acceptance of the use of optical and radar remote sensing to estimate forest attributes such as biomass (Chuvieco, 2009; Leboeuf, Fournier, Luther, Beaudoin, & Guindon, 2012; Le Toan, Beaudoin, Riom, & Guyon, 1992; Tanase, de la Riva, Santoro, Pérez-Cabello, & Kasischke, 2011), airborne laser scanning (ALS) is considered one of the best techniques for forest structural parameters estimation (Lefsky, Cohen, Parker, & Harding, 2002; Maltamo, Næsset, & Vauhkonen, 2014; Vosselman & Maas, 2010).

In this sense, some studies have used low-density ALS data to estimate forest parameters such as tree height, crown diameter, basal area, stem density, volume (Guerra-Hernández, Tomé, & González-Ferreiro, 2016a; Hayashi, Weiskittel, & Sader, 2014; Holopainen et al., 2010; Mehtätalo, Virolainen, Tuomela, & Packalen, 2015; Montealegre, Lamelas, de la Riva, García-Martín, & Escribano, 2016; Næsset, 2002; Næsset & Økland, 2002; Popescu, Randolph, & Ross, 2003) as well as biomass (García-Gutiérrez, Martínez-Álvarez, Troncoso, & Riquelme, 2015; Guerra-Hernández et al., 2016b; Hall, Burke, Box, Kaufmann, & Stoker, 2005; Montagnoli et al., 2015; Shendryk, Margareta, Leif, Natascha, 2014). In addition, some of them have compared different point densities (González-Ferreiro et al., 2013; Singh, Gang, James, & Ross, 2015). However, few studies have been focused on comparing different algorithms to estimate forest parameters and they were all applied to high-density point clouds (Gagliasso, Hummel, & Temesgen, 2014; García-Gutiérrez et al., 2015; Gleason & Im, 2012; Görgens, Montagni, & Rodriguez, 2015; Latifi, Nothdurft, & Koch, 2010).

The main objective of this study is to estimate the CO<sub>2</sub> emissions derived from the consumption of the aboveground tree biomass (AGB), which refers to the total biomass of the trees considering stem, branches and needles, in a heterogeneous Aleppo pine forest,

located in Aragón Region (Spain). To achieve this goal, a discrete, multiple-return, low point density ALS data and field plots representative of pine stands were used to fit and validate the AGB models. A secondary objective was the comparison of different regression models, including machine learning.

Besides, the majority of previous approaches to CO<sub>2</sub> estimation assume that biomass is completely consumed. However, during wildfires in conifer stands in some cases only the needles and the small fine twigs of the pine crowns are consumed (Call & Albin, 1997; Mitsopoulos & Dimitrakopoulos, 2007; Scott & Reinhardt, 2001). Consequently, different combustion factors were applied to avoid assuming that biomass was completely consumed by the fire (French, Goovaerts, & Kasischke, 2004). The fire severity levels were extracted from the difference normalized burn ratio ( $\Delta$ NBR) spectral index applied to Landsat 8 OLI images.

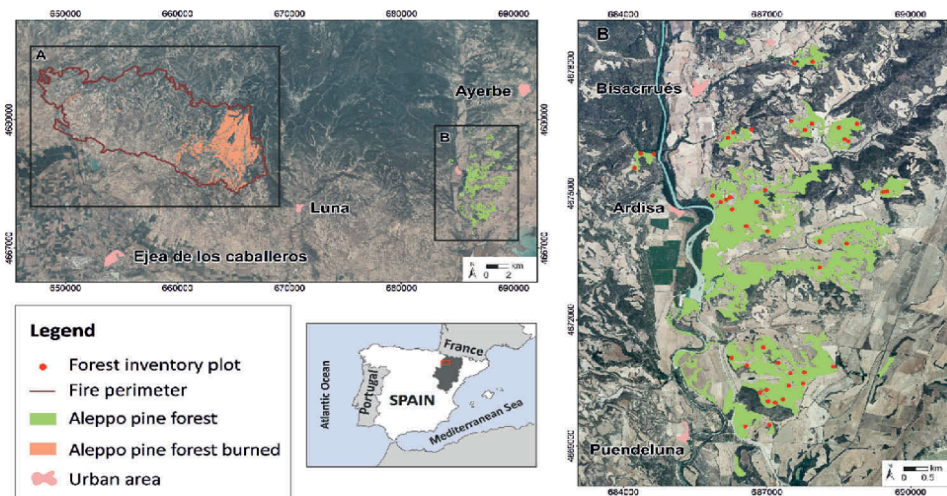
## Materials

### Study area

The study area, burned on 4 July 2015, is located in Luna municipality, northeast of Spain (42°12'N, 0°45'W). Aleppo pine has a high potential of ignition and represents almost 50% of the forested area in Aragón and is well adapted to these Mediterranean environmental conditions. The fire scorched in the area of 14,263 ha, of which 3390.4 ha was woodland. Those forested areas were covered in a 62.3% by monospecific Aleppo pine. As can be observed in Figure 1, for forest inventory purposes, the field campaign to estimate AGB was conducted in a close unburned area (Figure 1(b)). The proximity between both sites (see Figure 1(a) and (b)) and the similarity on environmental characteristics such as slope, climate and vegetation enable to extrapolate the AGB model to the burned area (Figure 1(a)). This similarity was previously evaluated by comparing some variables derived from the ALS metrics such as slope, canopy cover and tree height.

These heterogeneous pine forests from the structural point of view appear fragmented into stands of variable size, accompanied by an evergreen understorey with species, such as: *Quercus ilex* subsp. *rotundifolia*, *Quercus coccifera*, *Juniperus oxycedrus*, *Buxus sempervirens* and *Juniperus phoenicea*.

Climate of the region is Mediterranean with continental features. Annual precipitation is medium-low and irregular, averaging 525 mm and mostly occurring in autumn and spring. Winter has a monthly mean temperature less than 10°C, whereas summers have temperatures of ~20°C. The study area is characterized by a hilly topography, with elevations ranging from 430 to 1150 m above sea level and slopes



**Figure 1.** Study area. Aleppo pine forest stands, inside the perimeter of the Luna wildfire (a); location of the 46 forest inventory plots (b). High spatial resolution orthophotography (PNOA-2012) (IGN, 2017a) is used as backdrop.

from 0° to 39°. The lithology of the study area corresponds to Miocene shales and sandstones, alternating with conglomerates.

### Field plot data

Field data were acquired in 46 circular plots, 15 m radius at the unburned area during June and July 2015 (Figure 1(b)). The location of the field plots was selected, within the limits of the Aleppo pine stands at the unburned area, using a stratified random sampling technique, in order to achieve a representative sample of the variability of the terrain (Næsset & Økland, 2002), forest structure and tree density (Montealegre et al., 2016). Thereby, terrain slopes, tree height and canopy cover of the study area were derived from ALS point cloud to define homogeneous areas.

The centre of the selected plots (Figure 1(b)) was located in the field using a Leica VIVA GS15 CS10 GNSS real-time kinematic Global Positioning System. The average accuracy of the planimetric coordinates was 0.18 m. Tree breast height diameter (dbh) was measured at 1.3 m, using a Mantax Precision Blue diameter caliper (Haglöf Sweden®). It should be noted that only the trees with a dbh > 7.5 cm were measured in each plot. The AGB was calculated for each plot according to Montero, Ruiz-Peinado, and Muñoz (2005) allometric equation and extrapolated to per hectare biomass value (kg of dry biomass per ha) considering the plot area (Equation (1)).

$$\text{Biomass(kg/ha)} = \frac{CF \cdot e^a \cdot \text{dbh}^b}{A_{\text{plot}}} \cdot 10,000, \quad (1)$$

where  $CF$  is a correction factor ( $CF = e^{SEE^2/2}$ ) being  $e$  the Euler number and  $SEE$  the standard error

(0.151637);  $dbh$  is breast height diameter in cm;  $a$  (−2.0939) and  $b$  (2.20988) are the specific parameters for Aleppo pine; and  $A_{\text{plot}}$  is the area of each plot (706.8 m<sup>2</sup>).

These data act as ground truth to adjust and validate the AGB predictive model, which would be extrapolated to the burned area (Figure 1(a)) to estimate pre-fire biomass. The extrapolation of the AGB model was carried out in a Geographical Information System (GIS) environment using the selected Light Detection and Ranging (LiDAR) metrics and the coefficients of the model.

### Remote sensing data

The ALS data were captured for the burned and unburned area in several surveys carried out 4 years before the fire ignition between January and February 2011, using a small-footprint oscillating-mirror airborne Leica ALS60 discrete-return sensor. The Spanish National Plan for Aerial Orthophotography (PNOA) provided these data with a nominal density of 0.5 point/m<sup>2</sup> (IGN, 2017b). Data were delivered by the National Geographic Information Centre (CNIG) in 2 × 2 km tiles of raw data points in LAS binary files format v. 1.2. The  $x$ ,  $y$  and  $z$  coordinates were provided in UTM Zone 30 ETRS 1989 geodetic reference system and orthometric heights. The point cloud was captured with up to four returns measured per pulse. The ALS60 sensor was operating in 1.064 μm wavelength, 0.22 mrad beam divergence and ±29 scan angle degrees from nadir. The ALS point cloud density was 1.5 point/m<sup>2</sup>, considering all returns with a vertical accuracy better than 0.2 m for the area burned on 4 July 2015 and for the unburned one. The temporal lag between ALS acquisition data at the unburned area, captured in 2011, and fieldwork



campaign, performed in June and July 2015, was considered appropriate, as no significant changes took place in the study area in that period.

Pre-fire and post-fire Landsat 8 OLI land surface reflectance images, acquired on June 30 (path 200, row 31) and July 9 (path 199, row 31) 2015, were selected for  $\Delta$ NBR index calculation for the burned area and downloaded from USGS (2016). These products are generated using Landsat Surface Reflectance Code (LaSRC) algorithm (Vermote, Justice, Claverie, & Franch, 2016).

## Methods

The two-phase approach methodology includes the pre-fire biomass estimation through the comparison of different models and the estimation of biomass losses by applying three burning efficiency factors to assess the CO<sub>2</sub> emissions to the atmosphere (Figure 2).

### Pre-fire AGB estimation

This section describes the process followed for ALS data processing, as well as the generation of pre-fire AGB model.

### ALS data processing

The first processing step was noise point removal, which included verification of the overlapping returns. Thereafter, ALS point clouds were filtered using the multiscale curvature classification algorithm (Evans & Hudak, 2007) to extract the ground points. This algorithm, implemented in the MCC 2.1 command-line tool, is suitable for this environment according to

Montealegre, Lamelas, and de la Riva (2015a). Then, a digital elevation model (DEM) with a 1 m size grid was generated using the Point-TIN-Raster interpolation method (Renslow, 2013), following Montealegre, Lamelas, & de la Riva, (2015b). The normalized heights were obtained by the subtraction of the ground elevation value of the DEM from each point height. The normalized ALS tiles were clipped to the spatial extent of each field plot (Figure 1(b)). Furthermore, a wide range of statistical metrics commonly used as independent variables in forestry were calculated (Evans, Hudak, Faux, & Smith, 2009) using FUSION LDV 3.30 open source software (McGaughey, 2008). It should be noted that ALS-derived variables were generated after applying a threshold value of 2 m height so as to remove ground and understory laser hits according to Nilsson (1996) and Næsset and Økland (2002).

### Model for estimating pre-fire AGB

With the aim of comparing the predictive performance of different regression methods for the estimation of AGB, eight regression methods were analysed: a multivariate linear regression (MLR) model, two machine learning algorithms and five regression trees structures. These methods are briefly described below.

MLR has been widely employed to estimate forest parameters by relating dependent variables, from field-work campaign, and independent variables, extracted from the ALS point cloud (García, Godino, & Mauro, 2012; Gonzalez-Ferreiro, Dieguez-Aranda, & Miranda, 2012; Lim, Treitz, Wulder, St-Onge, & Flood, 2003; Means et al., 1999; Næsset & Økland, 2002; Watt et al., 2013). As a first step following Montealegre et al. (2016) the Spearman's rank correlation coefficient ( $\rho$ ) was calculated in R software, in order to select the ALS variables that show the strongest correlation coefficient with field plot biomass data. The selection of the ALS metrics was made within a minimum  $p$  value of  $\pm 0.5$ . Then, the selected variables were included in a forward stepwise regression, in order to avoid overfitting by selecting the smallest possible number of predictor variables. The fitted model was selected according to measures of goodness of fit. Moreover, it was verified if the fitted model meets the basic assumptions of linear regression models according to García et al. (2012). Logarithmic transformation of dependent and independent variables was explored in the cases where statistical hypothesis of linear regression models could not be fulfilled (García et al., 2012; Means et al., 1999), as well as to verify whether the measures of goodness of fit of the models improve.

Support vector machine (SVM) is a supervised learning model which has associated learning algorithms that analyse and recognize patterns. This method assumes that input data are separable in space (Mountrakis, Jungo, & Caesar, 2011). SVM tries to find among multidimensional hyperplanes

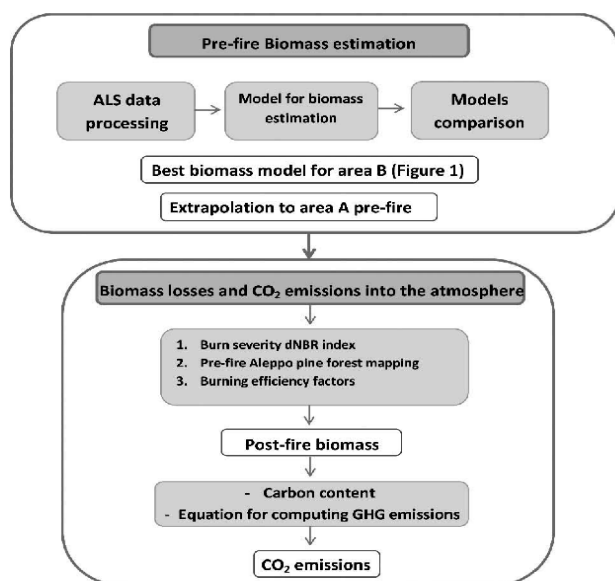


Figure 2. Methodology for biomass losses and CO<sub>2</sub> emissions estimation.



the optimal separation between classes, where the separability is a maximum. The data located in the hyperplane are the most difficult to classify since they have lower separability and they are called support vectors. SVM was implemented by using R package “e1071” and models with linear and radial kernels were computed. In both SVM models, the parameter cost was defined in the interval 1–1000, and the parameter gamma in the interval 0.01–1, applying the best parameters after tuning the model.

Random forest (RF) is an ensemble learning method that uses decision trees as base classifier. RF combines a decision tree that depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001). The algorithm adds randomness to bagging and increases the diversity of decision trees by growing them from different subsets. In each decision tree, RF divides the nodes by using the best variables from a random sample. RF was implemented through R package “randomForest” (Liaw & Wiener, 2002) and “caret” (Van Essen et al., 2001). The RF model was adjusted using two parameters: the number of trees to growth (ntrees) and the number of variables selected randomly at each split (mtry). They were in the intervals 1–1000 and 1–2, respectively.

The regression tree structures are nonparametric regression techniques based on “If-Then” rules. In this study three linear models and two non-linear local models are computed. The R package “CORElearn” has been used to perform the different regression trees. It should be added that the differences between them refer to the regression model considered in the leaf nodes.

Locally weighted linear regression (LWLR), or loess, is a method which fits a regression surface to data by smoothing the dependent variable as a function of the independent variables (Cleveland & Devlin, 1988). The coefficient of smoothness is fitted by computing weighted mean square error and considering a distance function.

Linear model with a minimum length principle (MDL) is based on the rule developed by Rissanen (1978) which considers that regularities in a set of data can be used to compress the data by using fewer symbols, from a finite alphabet, than needed to explain the data faithfully.

Reduced linear model (RLM) is a linear model computed by the least square method and, after that, simplified using an exhaustive search to remove those variables that contribute little to the model in order to minimize the estimated error, as in regression tree models like the so-called M5 in Waikato Environment for Knowledge Analysis (Weka) software (Quinlan, 1992).

K nearest neighbour (KNN) is a lazy learning method based on the KNN algorithm (Fix & Hodges, 1951), which includes two phases. First, the KNNs are

searched using the complete dataset and considering an established distance. Then, the mean of the k-most similar instances is used for the prediction.

Weighted k nearest neighbours (WKNNs) is a refinement of the KNN algorithm, which gives greater weight to the closer neighbours according to their distance to the observations. In this sense, the weighted mean of the KNNs is used for the prediction.

### Model validation and comparison

The algorithms were computed after applying a pre-processing phase which is based on the normalization of the data in values ranging from 0 to 1. The scaling of the data avoids weights saturation (Görgens et al., 2015) and may improve the performance of the models. In order to avoid overfitting of the model by selecting the smallest possible number of predictor variables, a forward stepwise regression was used.

Considering that fieldwork is a time-consuming task and increases the costs of the study, it was not possible to measure a high number of field plots. In this sense, the 46 measured plots, although may seem a low number, are enough to meet the statistical requirements. Accordingly, the models were validated using a leave-one-out cross-validation (LOOCV) technique (Maltamo et al., 2014), in order to do not further reduce the sample (Andersen, McGaughey, & Reutebuch, 2005). For those methods with randomness, LOOCV was executed 100 times so as to increase the robustness in the results (García-Gutiérrez et al., 2015).

The comparison between models was performed by analysing the results in terms of root mean square error (RMSE) and bias. Furthermore, Friedman non-parametric test was applied in order to compare the performance of the different models (Friedman, 1940). The test was carried out separately for each RMSE measure of each fold of the cross-validation (Stojanova, Panče, Valentin, Andrej, & Sašo, 2010). In those cases where the null hypothesis of Friedman test was rejected, which implies that the models were not equivalent, the Nemenyi (1963) *post-hoc* test was used to determine whether the differences between the models were statistically significant, with a significance level of 0.05.

### Estimates of biomass losses and conversion to CO<sub>2</sub> emissions into the atmosphere

The estimation of biomass losses was performed in three phases: (i) wildfire severity estimation, (ii) pre-fire Aleppo pine forest location mapping and (iii) selection of burning efficiency factors related to pre-fire vegetation (De Santis et al., 2010; Oliva & Chuvieco, 2011).

First, wildfire severity was estimated according to FIREMON methodology (Key & Benson, 2006). NBR was calculated for pre-fire and post-fire images

(Equation (2)). Then, the  $\Delta\text{NBR}$  was estimated by the subtraction of NBR post-fire from NBR pre-fire (Equation (3)). Subsequently, the burned area was delimited using this index in a GIS environment.

$$\text{NBR} = (\rho_{\text{NIR}} - \rho_{\text{SWIR}}) / (\rho_{\text{NIR}} + \rho_{\text{SWIR}}), \quad (2)$$

$$\Delta\text{NBR} = \text{NBR}_{\text{prefire}} - \text{NBR}_{\text{postfire}}, \quad (3)$$

where  $\rho_{\text{NIR}}$  (near infrared) and  $\rho_{\text{SWIR}}$  (short-wave infrared) refer to bands 5 and 7 Landsat 8 OLI reflectance, respectively.

In a second phase, the location of pre-fire Aleppo pine woodland was delimited using the Spanish National Forest Map (MAGRAMA, 2016b), and the canopy height model derived from the ALS data captured previous to fire. In order to improve the accuracy in forest location, stands less than 2 m high were excluded from the analysis.

Third, a thorough bibliographic search of burning efficiency values for Mediterranean conifer forests was conducted (Deeming, Burgan, & Cohen, 1977; Miranda et al., 2005). However, few approaches were suitable to our Mediterranean conifer forests and most of them assume that forest biomass is consumed completely (French et al., 2004). The goal of this study was to obtain spatialized coefficients related to different burn severity levels. Thus, following De Santis et al. (2010) methodology, three burning efficiency factors, related to pre-fire vegetation, were applied considering low, moderate and high severity levels. The four generic severity ranges proposed by Key and Benson (2006) (Table 1) were reclassified in three ranges so as to match with the three burning efficiency factors. In this regard, the low burning efficiency value denotes low consumption of the leaves and very low woody branches consumption; the moderate burning efficiency value indicates intermediate consumption of the leaves and moderate consumption of small branches; and the high burning efficiency value suggests a complete consumption of the leaves and high loss of small branches and twigs.

The conversion of biomass losses to  $\text{CO}_2$  emissions requires the estimation of the biomass carbon content and the application of an emission factor. The carbon content was computed using a conversion factor of 0.499 set by Montero et al. (2005) for Aleppo pine. With respect to the emission factors, several conversion factors have been proposed so as

to estimate different GHG emissions to the atmosphere. In this sense, the account of  $\text{CO}_2$  emissions to the atmosphere generated from forest biomass combustion was obtained according to Trozzi, Vaccaro, & Piscitello (2002) equation, which includes the same parameters as the equations established by IPCC (2006), Levine (2003) and Seiler and Crutzen (1980) (Equation (4)).

$$\text{CO}_2 = \varepsilon \cdot \delta \cdot C, \quad (4)$$

where  $\varepsilon$  is the fraction of total carbon emitted as  $\text{CO}_2$  (0.888);  $\delta$  is the factor of conversion from the emissions in ton of carbon to the emissions in ton of  $\text{CO}_2$  (44/12); and  $C$  is the carbon content.

## Results

A summary of the field plot characteristics is presented in Table 2. Inventoried trees present a variety of diameters, from 14.2 to 28.1 cm, and diverse heights, ranging from 7.2 to 17.2 m. This accounted for the variability of biomass in the study area.

All models included two ALS-derived variables: the percentage of first returns above 2 m ( $t$ -test: 8.3) and the 40th percentile of the return heights ( $t$ -test: 4.4), both variables showing a direct and coherent relation with AGB. The higher value of the variables, the higher biomass amount.

The regression models to estimate the AGB are summarized in Table 3. The MLR and the SVM with radial kernel (cost = 570 and gamma = 0.03) models presented the lowest RMSE with 6.1 and 7.3 ton/ha, respectively. LWLR regression tree performs slightly better than SVM with linear kernel (cost = 210 and gamma = 0.01), with RMSE of 8.3 and 8.5 ton/ha, respectively. Furthermore, the remaining regression trees as well as RF machine learning (ntrees = 500 and mtry = 1) show a lower accuracy. It should be added that most of the models present values of bias close to zero, except from SVM linear kernel, WKNN and KNN models that show a slight overestimation with values close to 1.

The performance comparison between the models, by using Friedman test, indicates that the models are not equivalent with a  $p$ -value of 0.000. However, the application of *post-hoc* Nemenyi test shows that only WKNN ( $p$ -value = 0.0) and KNN ( $p$ -value = 0.0) models presents differences statistically significant, with 95% of probability.

**Table 1.** Severity levels,  $\Delta\text{NBR}$  generic ranges defined by Key and Benson (2006), and burning efficiency factors used to estimate biomass losses following De Santis et al. (2010).

| Severity level         | $\Delta\text{NBR}$ range | Burning efficiency factors |
|------------------------|--------------------------|----------------------------|
| Unburned               | -100 to +99              | 0.00                       |
| Low severity           | +100 to +269             | 0.25                       |
| Moderate-low severity  | +270 to +439             | 0.42                       |
| Moderate-high severity | +440 to +659             |                            |
| High severity          | +660 to +1300            | 0.57                       |

**Table 2.** Summary of the field plots characteristics ( $n = 46$ ; inventoried trees = 1870).

|                        | Min.  | Max.   | Range | Mean  | SD    |
|------------------------|-------|--------|-------|-------|-------|
| Slope (degrees)        | 2.5   | 29.8   | 27.2  | 12.8  | 6.4   |
| Tree height (m)        | 7.2   | 17.2   | 10.0  | 9.6   | 1.8   |
| Tree density (tree/ha) | 282.9 | 1202.5 | 919.6 | 575.8 | 237.8 |
| dbh (cm)               | 14.2  | 28.1   | 13.9  | 18.7  | 2.8   |
| AGB (ton/ha)           | 24.4  | 130.9  | 106.4 | 55.4  | 20.4  |

**Table 3.** Summary of the models and the validation results for the estimated variable in terms of RMSE, relative RMSE (% RMSE) and bias.

| Predictive model  | Fitting phase |       |      | Cross-validation |       |      |       |
|-------------------|---------------|-------|------|------------------|-------|------|-------|
|                   | RMSE          | %RMSE | Bias | RMSE             | %RMSE | Bias | $R^2$ |
| MLR               | 6.2           | 11.1  | 0.1  | 6.2              | 11.1  | 0.0  | 0.8   |
| SVM radial kernel | 6.0           | 10.8  | 0.3  | 7.4              | 13.3  | 0.6  | 0.8   |
| SVM linear kernel | 8.0           | 14.3  | 1.1  | 8.5              | 15.4  | 1.4  | 0.8   |
| RF                | 4.7           | 8.5   | 0.2  | 9.3              | 16.7  | 0.5  | 0.8   |
| LWLR              | 5.8           | 10.5  | -0.2 | 8.3              | 15.0  | 0.2  | 0.8   |
| MDL               | 6.1           | 11.0  | -0.4 | 8.8              | 15.9  | 0.6  | 0.8   |
| RLM               | 7.3           | 13.2  | -0.1 | 9.3              | 16.7  | -0.1 | 0.8   |
| WKNN              | 10.7          | 19.3  | 0.9  | 12.7             | 23.0  | 1.1  | 0.6   |
| KNN               | 11.4          | 20.6  | 1.2  | 12.9             | 23.2  | 1.1  | 0.6   |

Figure 3 shows the scatter plots of the observed AGB against the model predictions for the different regression models. MLR and SVM with radial kernel show consistent results and high coefficient of determination (0.88 and 0.87, respectively). SVM with linear kernel and LWLR also present good coefficient of determination (0.84 and 0.83, respectively). Lower coefficients of determination as well as less stable results are evidenced in the scatter-plots for the remaining regression tress, especially WKNN and KNN.

The implementation of the MLR model (Equation (5)) in a GIS allowed estimating pre-fire AGB, which accounts 546,486.7 ton.

$$\text{PrefireAGB} = 1.007 \cdot 10689.32$$

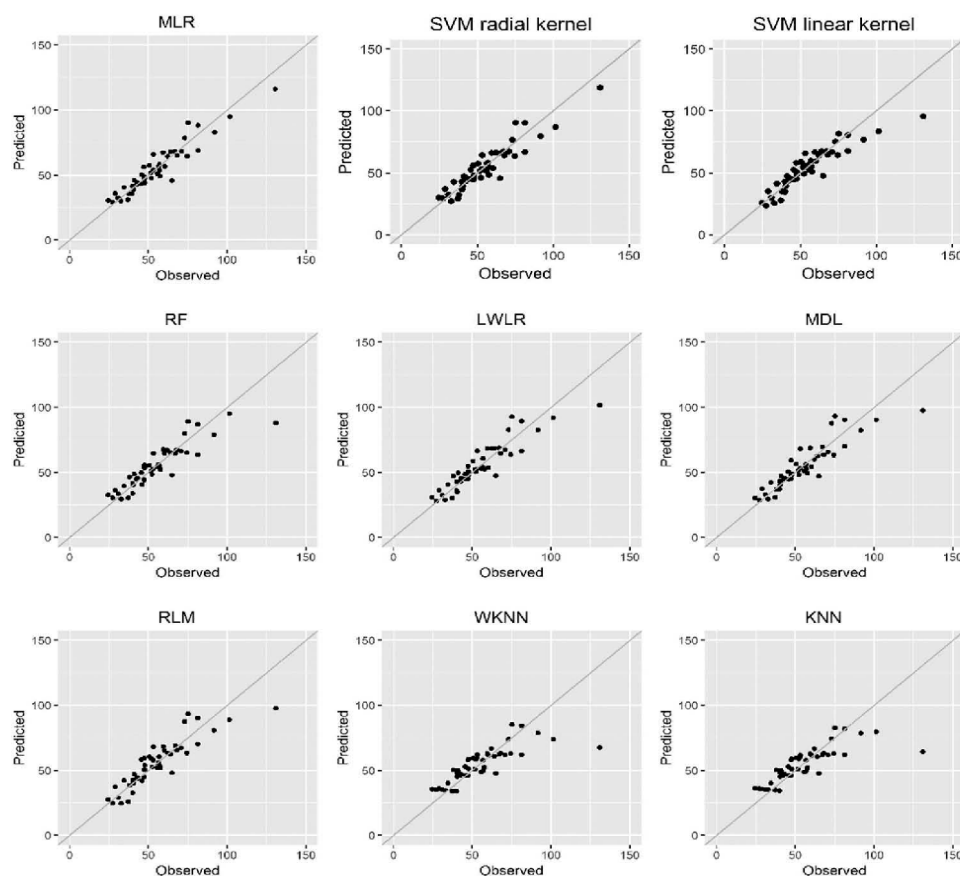
$$\cdot e^{(0.0158 \cdot \text{percentage of first returns above 2m})}$$

$$\cdot e^{(0.0713 \cdot 40\text{th percentile of height})} \quad (5)$$

The Aleppo pine forest was burned with a high severity in most part of the area, as can be observed in Figure 4 (a). The biomass losses range from 4 ton/ha to more than 12 ton/ha (Figure 4(b)). As can be observed in Table 4, high severity areas represent ~60% of Aleppo pine burned area, accounting ~70% of biomass losses. Finally, the combustion of Aleppo pine forest in Luna wildfire emitted 426,754.8 tons of CO<sub>2</sub> into the atmosphere.

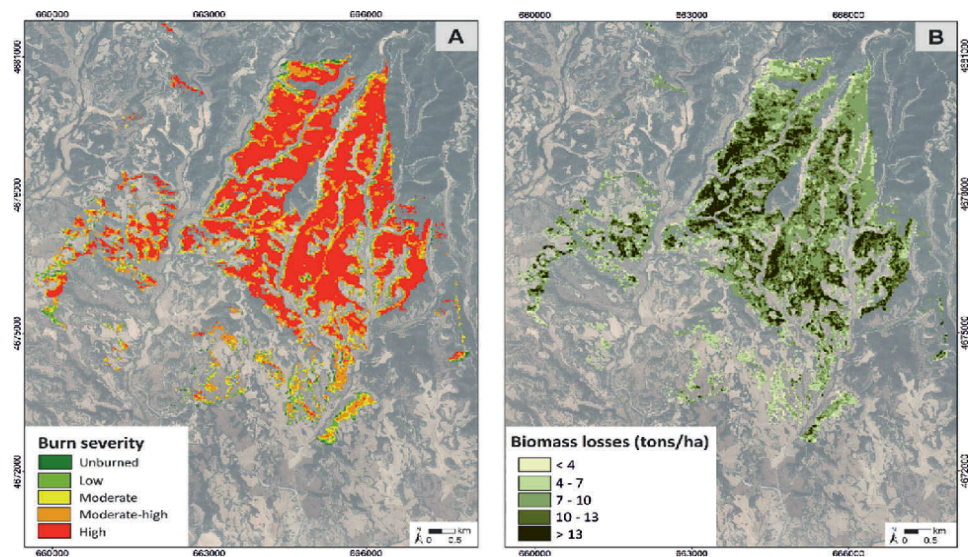
## Discussion

The use of GHG emissions equations is widely accepted for accounting forest biomass combustion by a wildfire (IPCC, 2006; Levine, 2003; Seiler &



**Figure 3.** Scatterplot of predicted values vs. observed values for the AGB using different regression methods.





**Figure 4.** Burn severity estimated in Aleppo pine forest burned in Luna wildfire using  $\Delta\text{NBR}$  index (a). Biomass losses estimation applying combustion factors (b). High spatial resolution orthophotography (PNOA-2012) (IGN, 2017a) is used as backdrop.

**Table 4.** Summary of results obtained for the burned area concerning to Aleppo pine forest affected by fire, pre-fire AGB, biomass losses, carbon content and  $\text{CO}_2$  emissions.

| Burn severity | Aleppo pine area (ha) | Pre-fire AGB (ton) | Biomass losses (ton) | Carbon content (ton) | $\text{CO}_2$ emissions (ton) |
|---------------|-----------------------|--------------------|----------------------|----------------------|-------------------------------|
| Unburned      | 26.8                  | 20,744.3           | 0.0                  | 0.0                  | 0.0                           |
| Low           | 117.0                 | 36,022.2           | 9005.5               | 4493.7               | 14,631.7                      |
| Medium        | 594.5                 | 169,908.7          | 71,361.6             | 35,609.4             | 115,944.4                     |
| High          | 1034.2                | 319,811.4          | 182,292.5            | 90,963.9             | 296,178.6                     |
| Total         | 1772.6                | 546,486.7          | 262,659.7            | 131,067.2            | 426,754.8                     |

Crutzen, 1980; Trozzi et al., 2002). Moreover, several conversion factors as well as emission factors, from global to regional scales, have been proposed to accurately estimate emissions to the atmosphere. However, one of the main uncertainties related to the use of these equations is the account of pre-fire biomass and biomass losses. In this sense, LiDAR technology has been proposed as the best technique to accurately estimate forest structural parameters, such as biomass, and artificial intelligence methods have been applied and compared to generate this variable. Nevertheless, little research has focused on comparing the performance of several regression models, including machine learning algorithms and regression trees, regarding traditional MLR models to estimate forest parameters.

This study proposes the use of low point density ALS data to improve the estimation of pre-fire AGB by comparing a set of state-of-the-art methods and traditional linear regression methods in a Mediterranean Aleppo pine forest, considering that biomass estimation is the key information to compute burning emissions.

The results demonstrate that low-density ALS data can be used to accurately estimate pre-fire biomass.

The two ALS-derived variables included in the models were analogous to those proposed by other authors (i.e. Guerra-Hernández et al., 2016b; Montagnoli et al., 2015). These variables concern the canopy cover distribution and the vertical distribution of the point cloud. The comparison between regression models shows that the MLR model has the lowest RMSE (6.1 ton/ha) and bias (0.0), matching with the values obtained by other authors (Gonzalez-Ferreiro et al., 2012; Montealegre, Lamelas, de la Riva, García-Martín, & Escribano, 2015c). Consequently, MLR slightly outperforms other nonparametric methods supporting Görgens et al. (2015) findings. However, no statistically significant differences between MLR and SVM with kernel radial were found. This suggests that the results partly agree with Gleason and Im (2012), Gagliasso et al. (2014) and García-Gutiérrez et al. (2015), who obtained lower estimation errors with nonparametric techniques, although the later authors included a relatively high number of independent variables in the models. In this sense, the use of a large number of variables tends to increase the performance of the models. Nevertheless, the selection of a reduced number of biologically representative variables, especially when computing non-linear regression models, might generate more understandable models for forest management purposes. This also might explain that MLR models outperform other nonparametric models, considering the number of variables included in Görgens et al. (2015) models. It is to notice that, as in the case of several previous studies (García, Riaño, Chuvieco, & Danson, 2010; Næsset & Gobakken, 2008; Næsset & Økland, 2002), it has been necessary to perform a logarithmic

transformation of the dependent variable in order to meet the assumptions of the linear regression model.

The three-phase approach performed is considered a suitable option for estimating biomass losses, which account for 262,659.7 ton. This methodology solves the lack of post-fire forest structure information derived from ALS data and constitutes an alternative to field estimation of burning efficiency, which is laborious, expensive and requires a detailed knowledge of the pre-fire scenario (De Santis et al., 2010). The use of conversion and emission factors for the Mediterranean basin, included in Trozzi et al. (2002) equation, enables to accurately estimate CO<sub>2</sub> emission at a regional scale, summing up a total of 426,754.8 ton.

When comparing nonparametric methods and linear regression models, discrepancies appear. Our findings show that not always the use of nonparametric methods ensures the best biomass estimations. In this sense, the generation of several models such as MLR, SVM with radial kernel, LWLR or SVM with linear kernel might be taken into account. Furthermore, the use of variable selection processes should be considered, in order to determine a limited number of variables which are biologically representative. Consequently, the use of new artificial intelligence models and nonparametric models, which have several advantages for example no need of normality, should be used within the forestry and environmental purposes of obtaining robust and understandable models. The improved estimation of CO<sub>2</sub> emissions from biomass burning, by including ALS data as relevant information for computing biomass, is considered to better understand the interactions between fire disturbances and the emissions to the atmosphere. Nevertheless, it is to notice that our model does not consider the emissions generated by combustion of litter, shrubs and young trees, implying an underestimation of the total CO<sub>2</sub> emissions during wildfire. It was not possible to estimate these fractions of biomass. In fact, the few studies developed to estimate shrub biomass were performed using high-density ALS data or full waveform LiDAR (Estornell, Ruiz, Velázquez-Martí, & Hermosilla, 2012; E. Greaves et al., 2016; Swatantran, Dubayah, Roberts, Hofton, & Bryan Blair, 2011). Moreover, the majority of them were developed in areas without tree cover due to the difficulty of the pulse to penetrate the canopy (Vosselman & Maas, 2010). In this sense, further research is needed on the estimation of shrub biomass using low point density ALS data in order to improve GHG emissions to the atmosphere.

Considering that the findings are site-dependent, the comparison of different biologically representative models for biomass estimation at regional scales, as well as alternative variable selection processes, may be considered. In this sense, the use of multi-

temporal ALS or multi-temporal series of remote sensing data might be useful to better understand the effect of wildfire disturbances to the atmosphere. It would also be desirable to focus on the account of CO<sub>2</sub> emissions or other GHG gasses generated by combustion of other Mediterranean species.

## Conclusions

This study verifies the usefulness of low-density ALS data to accurately estimate pre-fire AGB in a mono-specific Aleppo pine forest, which is relevant information to compute biomass losses caused by fire and CO<sub>2</sub> emissions. The comparison of the effectiveness of a set of state-of-the-art artificial intelligence methods and traditional linear regression methods is especially interesting to improve forest parameters modelling. The best model for pre-fire AGB estimation was the MLR, which included two ALS variables: the percentage of first returns above 2 m and 40th percentile of height, presenting an RMSE of 6.1 ton/ha and a bias of 0.0. No statistically significant differences between MLR and SVM with kernel radial, which is the second best model, were found. The three-phase approach used for biomass losses estimation and the subsequent transformation into CO<sub>2</sub> enable to quantify the emissions to the atmosphere by the combustion of Mediterranean Aleppo pine forest in Luna wildfire, summing up a total of 426,754.8 ton.

## Acknowledgements

This work was supported by the Spanish Ministry of Education Culture and Sports under Grant [FPU Grant BOE, 27/12/2014]. The ALS data were provided by the Spanish National Geographic Information Centre (CNIG). The authors are grateful to the material resources provided by the Centro Universitario de la Defensa (CUD) to develop the fieldwork.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Spanish Ministry of Education Culture and Sports under Grant [FPU Grant BOE, 27/12/2014].

## ORCID

Darío Domingo  <http://orcid.org/0000-0002-8362-7559>  
 María Teresa Lamelas-Gracia  <http://orcid.org/0000-0002-8954-7517>  
 Antonio Luis Montealegre-Gracia  <http://orcid.org/0000-0001-6288-2780>

Juan de la Riva-Fernández  <http://orcid.org/0000-0003-2615-270X>

## References

- Akagi, S.K., Yokelson, R.J., Burling, I.R., Meinardi, S., Simpson, I., Blake, D.R., ... Weise, D.R. (2013). Measurements of reactive trace gases and variable O<sub>3</sub> formation rates in some South Carolina biomass burning plumes. *Atmospheric Chemistry and Physics*, 13(3), 1141–1165. doi:10.5194/acp-13-1141-2013
- Andersen, H.E., McGaughey, R.J., & Reutebuch, S.E. (2005). Estimating forest canopy fuel parameters using LIDAR data. *Remote Sensing of Environment*, 94(4), 441–449. doi:10.1016/j.rse.2004.10.013
- Andreae, M.O., Browell, E.V., Garstang, M., Gregory, G.L., Harriss, R.C., Hill, G.F., & Wofsy, S.C. (1988). Biomass-burning emissions and associated haze layers over Amazonia. *Journal of Geophysical Research*, 93(D2), 1509–1527. doi:10.1029/JD093iD02p01509
- Breiman, L. (2001). Random forest. *Machine Learning*, 45, 5–32. doi:10.1023/A:1010933404324
- Call, P.T., & Albin, F.A. (1997). Aerial and surface fuel consumption in crown fires. *International Journal of Wildland Fire*, 7(3), 259–264. doi:10.1071/WF970259
- Chuvieco, E. (2009). Global impacts of fire. In E., Chuvieco Ed., *Earth observation of wildland fires in Mediterranean ecosystems* (pp. 1–10). Berlin Heidelberg: Springer. doi:10.1007/978-3-642-01754-4
- Cleveland, W.S., & Devlin, S.J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83 (403), 596–610. doi:10.1080/01621459.1988.10478639
- De Santis, A., Asner, G.P., Vaughan, P.J., & Knapp, D.E. (2010). Mapping burn severity and burning efficiency in California using simulation models and Landsat imagery. *Remote Sensing of Environment*, 114(7), 1535–1545. doi:10.1016/j.rse.2010.02.008
- Deeming, J., Burgan, R., & Cohen, J. (1977). General Technical Report INT-39, *The National Fire-Danger Rating System* — 1978. U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT, pp. 63.
- Estornell, J., Ruiz, L.A., Velázquez-Martí, B., & Hermosilla, T. (2012). Estimation of biomass and volumen of shrub vegetation using LiDAR and spectral data in a Mediterranean environment. *Biomass and Bioenergy*, 46, 710–721. doi:10.1016/j.biombioe.2012.06.023
- Evans, J.S., & Hudak, A.T. (2007). A multiscale curvature algorithm for classifying discrete return LiDAR in forested environments. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4), 1029–1038. doi:10.1109/TGRS.2006.890412
- Evans, J.S., Hudak, A.T., Faux, R., & Smith, A.M.S. (2009). Discrete return LiDAR in natural resources: Recommendations for project planning, data processing, and deliverables. *Remote Sensing*, 1(4), 776–794. doi:10.3390/rs1040776
- Fix, E., & Hodges, J. (1951). Texas, Report No. 4, Project No. 21-49-004, *Discriminatory analysis, nonparametric discrimination: Consistency properties*. USAF School of Aviation Medicine, Randolph Field
- French, N.H.F., Goovaerts, P., & Kasischke, E.S. (2004). Uncertainty in estimating carbon emissions from boreal forest fires. *Journal of Geophysical Research*, 109(D14), D14S08. doi:10.1029/2003JD003635
- Friedman, M. (1940). A comparison of alternative tests of significance for the problem rankings. *The Annals of Mathematical Statistics*, 11(1), 86–92. doi:10.1214/aoms/1177731944
- Gagliasso, D., Hummel, S., & Temesgen, H. (2014). A comparison of selected parametric and non-parametric imputation methods for estimating forest biomass and basal area. *Open Journal of Forestry*, 4(1), 42–48. doi:10.4236/ojfor.2014.41008
- García, D., Godino, M., & Mauro, F. (2012). LIDAR: Aplicación práctica al inventario forestal. *Editorial Académica Española*, 10, 3659009555.
- García, M., Riaño, D., Chuvieco, E., & Danson, F.M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sensing of Environment*, 114(4), 816830. doi:10.1016/j.rse.2009.11.021
- García-Gutiérrez, J., Martínez-Álvarez, F., Troncoso, A., & Riquelme, J.C. (2015). A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing*, 167, 24–31. doi:10.1016/j.neucom.2014.09.091
- Gleason, C.J., & Im, J. (2012). Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment*, 125, 80–91. doi:10.1016/j.rse.2012.07.006
- Gonzalez-Ferreiro, E., Dieguez-Aranda, U., & Miranda, D. (2012). Estimation of stand variables in *Pinus radiata* D. Don plantations using different LiDAR pulse densities. *Forestry*, 85(2), 281–292. doi:10.1093/forestry/cps002
- González-Ferreiro, E., Miranda, D., Barreiro-Fernandez, L., Bujan, S., García-Gutiérrez, J., & Dieguez-Aranda, U. (2013). Modelling stand biomass fractions in Galician *Eucalyptus globulus* plantations by use of different LiDAR pulse densities. *Forest Systems*, 22(3), 510–525. doi:10.5424/fs/2013223-03878
- Görgens, E.B., Montagni, A., & Rodriguez, L.C.E. (2015). A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Computers and Electronics in Agriculture*, 116, 221–227. doi:10.1016/j.compag.2015.07.004
- Greaves, H., Vierling, L., Eitel, J., Boelman, N., Magney, T., Prager, C., & Griffin, K. (2016). High-resolution mapping of aboveground shrub biomass in Arctic tundra using airborne lidar and imagery. *Remote Sensing of Environment*, 184, 361–373. doi:10.1016/j.rse.2016.07.026
- Guerra-Hernández, J., Görgens, E.B., García-Gutiérrez, J., Rodriguez, L.C.E., Tomé, M., & González-Ferreiro, E. (2016b). Comparison of ALS based models for estimating aboveground biomass in three types of Mediterranean forest. *European Journal of Remote Sensing*, 49, 185–204. doi:10.5721/EuJRS201649
- Guerra-Hernández, J., Tomé, M., & González-Ferreiro, E. (2016a). Using low density LiDAR data to map Mediterranean forest characteristics by means of an area-based approach and height threshold analysis. *Revista De Teledetección*, 46, 103–117. doi:10.4995/raet.2016.3980
- Hall, S.A., Burke, I.C., Box, D.O., Kaufmann, M.R., & Stoker, J.M. (2005). Estimating stand structure using discrete-return Lidar: An example from low density, fire prone ponderosa pine forests. *Forest Ecology and Management*, 208(1), 189–209. doi:10.1016/j.foreco.2004.12.001



- Hayashi, R., Weiskittel, A., & Sader, S. (2014). Assessing the feasibility of low-density LiDAR for stand inventory attribute predictions in complex and managed forests of Northern Maine, USA. *Forests*, 5(2), 363–383. doi:10.3390/f5020363
- Holopainen, M., Vastaranta, M., Rasinmäki, J., Kalliovirta, J., Mäkinen, A., Haapanen, R., ... Hyyppä, H. (2010). Estimation of timber assortments using low-density ALS data. In W. Wagner & B. Székely (Eds.), *ISPRS TC VII Symposium – 100 Years ISPRS: Vol. XXXVIII, Part 7A* (pp. 59–64). Vienna, Austria: IAPRS.
- IGN. (2017a). *PNOA Imagen*. Ministerio de Fomento, Instituto Geográfico Nacional, Madrid. Retrieved January 10, 2017, <http://pnoa.ign.es/presentacion-y-objetivo>.
- IGN. (2017b). *PNOA LiDAR*. Ministerio de Fomento, Instituto Geográfico Nacional, Madrid. Retrieved January 10, 2017, <http://pnoa.ign.es/presentacion>.
- IPCC. (2006). 2006 IPCC guidelines for national greenhouse gas inventories. Volume 4: Agriculture, forestry and other land use. In H.S. Eggleston, L. Buendia, K. Miwa, T. Ngara, & K. Tanabe Eds, *Prepared by the national greenhouse gas inventories programme*. Japan: IGES. ISBN 4-88788-032-4
- Key, C.H., & Benson, N. (2006). Landscape assessment (LA). Sampling and analysis methods. In L.J. Lutes, D. C. Keane, R.E. Caratti, J.F. Key, C.H. Benson, & N.C. Gangi (Eds.), *FIREMON: Fire effects monitoring and inventory system. Integration of standardized field data collection techniques and sampling design with remote sensing to assess fire effects* (pp. LA1–LA51). Fort Collins, CO: Rocky Mountain Research Station, U.S. Department of Agriculture, Forest Service.
- Lal, R. (2008). Sequestration of atmospheric CO<sub>2</sub> in global carbon pools. *Energy and Environmental Science*, 1, 86–100. doi:10.1039/B809492F
- Latifi, H., Nothdurft, A., & Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: Application of multiple optical/LiDAR-derived predictors. *Forestry*, 83(4), 395–407. doi:10.1093/forestry/cpq022
- Le Toan, T., Beaudoin, A., Riou, J., & Guyon, D. (1992). Relating forest biomass to SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 403–411. doi:10.1109/36.134089
- Leboeuf, A., Fournier, R.A., Luther, J.E., Beaudoin, A., & Guindon, L. (2012). Forest attribute estimation of north-eastern Canadian forests using QuickBird imagery and a shadow fraction method. *Forest Ecology and Management*, 266, 66–74. doi:10.1016/j.foreco.2011.11.008
- Lefsky, M.A., Cohen, W.B., Parker, G.G., & Harding, D.J. (2002). Lidar remote sensing for ecosystem studies. *BioScience*, 52(1), 19–30; (2002)052[0019: LRSFES]2.0. CO;2. doi:10.1641/0006-3568
- Levine, J.S. (2003). Biomass burning: The cycling of gases and particulates from the biosphere to the atmosphere. *Treatise on Geochemistry*, 4, 143–158. doi:10.1016/B0-08-043751-6/04143-8
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Lim, K., Treitz, P., Wulder, M., St-Onge, B., & Flood, M. (2003). LiDAR remote sensing of forest structure. *Progress in Physical Geography*, 27(1), 88–106. doi:10.1191/0309133303pp360ra
- MAGRAMA (2016a). *Estadísticas de incendios forestales*. Ministerio de Agricultura y Peces, Alimentación y Medio Ambiente, Madrid. Retrieved April 10, 2016, from [http://www.magrama.gob.es/es/desarrollo-rural/estadisticas/Incendios\\_default.aspx](http://www.magrama.gob.es/es/desarrollo-rural/estadisticas/Incendios_default.aspx)
- MAGRAMA (2016b). *Mapa Forestal de España 1:50.000 (MFE50)*. Ministerio de Agricultura y Peces, Alimentación y Medio Ambiente, Madrid. Retrieved May 4, 2016, from <http://www.magrama.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/informacion-disponible/mfe50.aspx>
- Maltamo, M., Næsset, E., & Vauhkonen, J. (2014). *Forestry applications of airborne laser scanning: Concepts and case studies*. Netherlands: Springer. doi:10.1007/978-94-017-8663-8
- McGaughey, R.J. (2008). *FUSION/LDV: Software for LIDAR data analysis and visualization*. USA: USDA Forest Service, Pacific Northwest Research Station.
- Means, J., Acker, S., Harding, D., Blair, J., Lefsky, M., Cohen, W., & McKee, W. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western Cascades of Oregon. *Remote Sensing of Environment*, 67, 298–308. doi:10.1016/S0034-4257(98)00091-1
- Mehtätalo, L., Virolainen, A., Tuomela, J., & Packalen, P. (2015). Estimating tree height distribution using low-density ALS data with and without training data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(4), 1432–1441. doi:10.1109/JSTARS.2015.2418675
- Mieville, A., Granier, C., Lioussé, C., Guillaume, B., Mouillot, F., Lamarque, J.-F., & Pétron, G. (2010). Emissions of gases and particles from biomass burning during the 20th century using satellite data and an historical reconstruction. *Atmospheric Environment*, 44 (11), 1469–1477. doi:10.1016/j.atmosenv.2010.01.011
- Miranda, A.I., Borrego, C., Sousa, M., Valente, J., Barbosa, P., & Carvalho, A. (2005). *Model of forest fire emissions to the atmosphere*. Aveiro, Portugal: Department of Environment and Planning, University of Aveiro AMB-QA-07/2005, Deliverable D252 of SPREAD Project (EVG1-CT-2001-00043).
- Mitsopoulos, I.D., & Dimitrakopoulos, A.P. (2007). Allometric equations for crown fuel biomass of Aleppo pine (*Pinus halepensis* Mill.) in Greece. *International Journal of Wildland Fire*, 16(5), 642–647. doi:10.1071/WF06038
- Montagnoli, A., Fusco, S., Terzagui, M., Kirschbaum, A., Pflugmacher, D., Cohen, W.B., ... Chiatante, D. (2015). Estimating forest aboveground biomass by low density lidar data in mixed broad-leaved forests in the Italian pre-alps. *Forest Ecosystems*, 2–10. doi:10.1186/s40663-015-0035-6
- Montealegre, A.L., Lamelas, M.T., & de la Riva, J. (2015a). A comparison of open-source LiDAR filtering algorithms in a Mediterranean forest environment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8), 4072–4085. doi:10.1109/JSTARS.2015.2436974
- Montealegre, A.L., Lamelas, M.T., & de la Riva, J. (2015b). Interpolation routines assessment in ALS-derived digital elevation models for forestry applications. *Remote Sensing*, 7, 8631–8654. doi:10.3390/rs70708631
- Montealegre, A.L., Lamelas, M.T., de la Riva, J., García-Martín, A., & Escibano, F. (2015c). Assessment of biomass and carbon content in a Mediterranean Aleppo pine forest using ALS data. In *Proceedings of 1st International Electronic Conference on Remote Sensing*. Basel, Switzerland: MDPI, 22 June - 5 July 2015, Vol. 1, d004. Sciforum Electronic Conference Series.

- Montealegre, A.L., Lamelas, M.T., de la Riva, J., García-Martín, A., & Escribano, F. (2016). Use of low point density ALS data to estimate stand-level structural variables in Mediterranean Aleppo pine forest. *Forestry*, 89 (4), 373–382. doi:10.1093/forestry/cpw008
- Montero, G., Ruiz-Peinado, R., & Muñoz, M. (2005). *Producción de biomasa y fijación de CO<sub>2</sub> por los bosques españoles* (p. 270). Madrid: Instituto Nacional de Investigación y Técnica Agraria y Alimentaria (INIA). ISBN: 8474985129
- Moriondo, M., Good, P., Durao, R., Bindi, M., Giannakopoulos, C., & Corte-Real, J. (2006). Potential impact of climate change on fire risk in the Mediterranean area. *Climate Research*, 31(1), 85–95. doi:10.3354/cr031085
- Mountrakis, G., Jungo, I., & Caesar, O. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. doi:10.1016/j.isprsjprs.2010.11.001
- Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80(1), 88–99. doi:10.1016/S0034-4257(01)00290-5
- Næsset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, 112(6), 3079–3090. doi:10.1016/j.rse.2008.03.004
- Næsset, E., & Økland, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, 79 (1), 105–115. doi:10.1016/S0034-4257(01)00243-7
- Narayan, C., Fernandes, P.M., Van Brusselen, J., & Schuck, A. (2007). Potential for CO<sub>2</sub> emissions mitigation in Europe through prescribed burning in the context of the Kyoto Protocol. *Forest Ecology and Management*, 251(3), 164–173. doi:10.1016/j.foreco.2007.06.042
- Nemenyi, P. (1963). *Distribution-free multiple comparisons*. Princeton University, Ph.D. thesis, Princeton, NY, USA.
- Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, 56(1), 1–7. doi:10.1016/0034-4257(95)00224-3
- Oliva, P., & Chuvieco, E. (2011). Towards a dynamic burning efficiency factor. In J. San-Miguel, I. Gitas, A. Camia, & S. Oliveira (Eds.), *Advances in remote sensing and GIS applications in Forest Fire Management: From local to global assessment* (pp. 51–56). Luxembourg: Publications Office of the European Union.
- Oliveira, S., Oehler, F., San-Miguel-Ayán, J., Camia, A., & Pereira, J.M.C. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. *Forest Ecology and Management*, 275, 117–129. doi:10.1016/j.foreco.2012.03.003
- Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., ... Hayes, D. (2011). A large and persistent carbon sink in the world's forests. *Science*, 333(6045), 988–993. doi:10.1126/science.1201609
- Pausas, J.G., Llovet, J., Rodrigo, A., & Vallejo, R. (2008). Are wildfires a disaster in the Mediterranean basin? – A review. *International Journal of Wildland Fire*, 17(6), 713–723. doi:10.1071/WF07151
- Pétron, G., Granier, C., Khattatov, B., Yudin, V., Lamarque, J.-F., Emmons, L., & Edwards, D.P. (2004). Monthly CO surface sources inventory based on the 2000–2001 MOPITT satellite data. *Geophysical Research Letters*, 31, L21107. doi:10.1029/2004GL020560
- Popescu, S.C., Randolph, H.W., & Ross, F.N. (2003). Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 29(5), 564–577. doi:10.5589/m03-027
- Prasad, V.K., Kant, Y., Gupta, P.K., Sharma, C., Mitra, A., & Badarinath, K.V. (2001). Biomass and combustion characteristics of secondary mixed deciduous forests in Eastern Ghats of India. *Atmospheric Environment*, 35 (18), 3085–3095. doi:10.1016/S1352-2310(01)00125-X
- Quinlan, J.R. (1992). Learning with continuous classes. In *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence* (pp. 343–348). Singapore, World Scientific.
- Renslow, M. (2013). *Manual of Airborne Topographic Lidar* (p. 528). Bethesda, MD: 99: The American Society for Photogrammetry and Remote Sensing. ISBN: 978-1570830976
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5), 465–471. doi:10.1016/0005-1098(78)90005-5
- Rodrigues, M., Ibarra, P., Echeverria, M., Perez-Cabello, F., & de la Riva, J. (2014). A method for regional-scale assessment of vegetation recovery time after high-severity wild-fires: Case study of Spain. *Progress in Physical Geography*, 38(5), 556–575. doi:10.1177/0309133314542956
- Roy, D.P., Jin, Y., Lewis, P.E., & Justice, C.O. (2005). Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sensing of Environment*, 97(2), 137–162. doi:10.1016/j.rse.2005.04.007
- Sá, A.C.L., Pereira, J.M.C., & Silva, J.M.N. (2005). Estimation of combustion completeness based on fire-induced spectral reflectance changes in a dambo grassland (Western Province, Zambia). *International Journal of Remote Sensing*, 26(19), 4185–4195. doi:10.1080/01431160500113468
- San-Miguel-Ayán, J., Rodrigues, M., De Oliveira, S.S., Pacheco, C.K., Moreira, F., Duguy, B., & Camia, A. (2012). Land cover change and fire regime in the European Mediterranean region. In F. Moreira, M. Arianoutsou, P. Corona, & J. de las Heras (Eds.), *Post-fire management and restoration of Southern European forests* (pp. 21–43). Netherlands: Springer. doi:10.1007/978-94-007-2208-8\_2
- Scott, J.H., & Reinhardt, E.D. (2001). Assessing crown fire potential by linking models of surface and crown fire behavior. Research Paper RMRS-29. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, Co.
- Sebastián-López, A., Salvador-Civil, R., Gonzalo-Jiménez, J., & SanMiguel-Ayán, J. (2008). Integration of socio-economic and environmental variables for modelling long-term fire danger in Southern Europe. *European Journal of Forest Research*, 127(2), 149–163. doi:10.1007/s10342-007-0191-5
- Seiler, W., & Crutzen, P.J. (1980). Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. *Climatic Change*, 2 (3), 207–247. doi:10.1007/BF00137988
- Shendryk, I., Margareta, H., Leif, K., & Natascha, K. (2014). Low-density LiDAR and optical imagery for biomass estimation over boreal forest in Sweden. *Forests*, 5(5), 992–1010. doi:10.3390/f5050992



- Singh, K.K., Gang, C., James, B.M., & Ross, K.M. (2015). Effects of LiDAR point density and landscape context on estimates of urban forest biomass. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 310–322. doi:10.1016/j.isprsjprs.2014.12.021
- Stojanova, D., Panče, P., Valentin, G., Andrej, K., & Sašo, D. (2010). Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecological Informatics*, 5(4), 256–266. doi:10.1016/j.ecoinf.2010.03.004
- Swatantran, A., Dubayah, R., Roberts, D., Hofton, M., & Bryan Blair, J. (2011). Mapping biomass and stress in the Sierra Nevada using lidar and hyperspectral data fusion. *Remote Sensing of Environment*, 115, 2917–2930. doi:10.1016/j.rse.2010.08.027
- Tanase, M., de la Riva, J., Santoro, M., Pérez-Cabello, F., & Kasischke, E. (2011). Sensitivity of SAR data to post-fire forest regrowth in Mediterranean and boreal forests. *Remote Sensing of Environment*, 115(8), 2075–2085. doi:10.1016/j.rse.2011.04.009
- Trozzi, C., Vaccaro, R., & Piscitello, R. (2002). Emissions estimate from forest fires: Methodology, software and European case studies. In *Proceedings of the 11th International Emission Inventory Conference*. Atlanta, GA.
- USGS. (2016). *USGS Earth explorer*. U.S. Department of the Interior, U.S. Geological Survey. Retrieved April 10, 2016 from <http://earthexplorer.usgs.gov/>
- Van Der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G. J., Kasibhatla, P.S., & Arellano, A.F. (2006). Interannual variability in global biomass burning emissions from 1997 to 2004. *Atmospheric Chemistry and Physics*, 6 (11), 3423–3441. doi:10.5194/acp-6-3423-2006
- Van Der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P.S., & Van Leeuwen, T.T. (2010). Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmospheric Chemistry and Physics*, 10(23), 11707–11735. doi:10.5194/acp-10-11707-2010
- Van Essen, D.C., Dickson, J., Harwell, J., Hanlon, D., Anderson, C.H., & Drury, H.A. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of the American Medical Informatics Association : JAMIA*, 8(5), 443–459. doi:10.1136/jamia.2001.0080443
- Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, 46–56. doi:10.1016/j.rse.2016.04.008
- Vosselman, G., & Maas, H.-G. (2010). *Airborne and terrestrial laser scanning*. p. 336. Caithness: Whittles Publishing. ISBN: 978-1904445-87-6.
- Ward, D.E., Hao, W.M., Susott, R.A., Babbitt, R.E., Shea, R. W., Kauffman, J.B., & Justice, C.O. (1996). Effect of fuel composition on combustion efficiency and emission factors for African savanna ecosystems. *Journal of Geophysical Research: Atmospheres*, 101(D19), 23569–23576. doi:10.1029/95JD02595
- Watt, M.S., Meredith, A., Watt, P., Gunn, A., Andersen, H., Reutebuch, S., & Nelson, R. (2013). Use of LiDAR to estimate stand characteristics for thinning operations in young Douglas-fir plantations. *New Zealand Journal of Forestry Science*, 43, 18. doi:10.1186/1179-5395-43-18
- Wiedinmyer, C., Akagi, S.K., Yokelson, R.J., Emmons, L.K., Al-Saadi, J.A., Orlando, J.J., & Soja, A.J. (2011). The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development*, 4(3), 625–641. doi:10.5194/gmd-4-625-2011
- Wiedinmyer, C., & Neff, J.C. (2007). Estimates of CO<sub>2</sub> from fires in the United States: Implications for carbon management. *Carbon Balance and Management*, 2–10. doi:10.1186/1750-0680-2-10



### **3.2. Estimation of total biomass in Aleppo pine forest stands applying parametric and nonparametric methods to low-density airborne laser scanning data**

Estimación de biomasa total en bosques de pino carrasco aplicando métodos paramétricos y no parámetros mediante datos de escáner láser aeroportado de baja densidad

#### **RESUMEN**

La cuantificación de la biomasa total es de utilidad para la evaluación de las políticas de regulación climáticas desde escalas locales a escalas globales. Esta investigación estima la biomasa total, incluyendo la biomasa arbórea y arbustiva, en bosques de Pino carrasco localizados en la región de Aragón (España), utilizando datos de escáner láser aeroportado (ALS) y trabajo de campo. La comparación de cinco métodos de selección y cinco modelos de regresión se realizó con objeto de relacionar la biomasa total, estimada en 83 parcelas de campo mediante ecuaciones alométricas, con diversas variables extraídas de la nube de puntos ALS. Para el cálculo de las variables ALS se utilizó un umbral de 0.2 m. La muestra se dividió en entrenamiento y validación componiéndose de 62 y 21 parcelas de campo, respectivamente. El modelo con menor error cuadrático medio después de la validación (15,14 tons ha<sup>-1</sup>) fue el modelo de regresión lineal multivariante. Dicho modelo incluyó tres variables ALS: el percentil 25 de la altura de retornos, la varianza y el porcentaje de primeros retornos sobre la media. El estudio confirma la utilidad de los datos ALS de baja densidad para estimar con exactitud la biomasa total, y por consiguiente mejorar la cuantificación de biomasa disponible y contenido de carbono en Pinares mediterráneos de Pino carrasco.





## Article

# Estimation of Total Biomass in Aleppo Pine Forest Stands Applying Parametric and Nonparametric Methods to Low-Density Airborne Laser Scanning Data

Darío Domingo <sup>1,\*</sup> , María Teresa Lamelas <sup>1,2</sup> , Antonio Luis Montealegre <sup>1</sup>,  
Alberto García-Martín <sup>1,2</sup> and Juan de la Riva <sup>1</sup>

<sup>1</sup> GEOFOREST-IUCA, Department of Geography, University of Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain; tlamelas@unizar.es (M.T.L.); monteale@unizar.es (A.L.M.); algarcia@unizar.es (A.G.-M.); delariva@unizar.es (J.d.l.R.)

<sup>2</sup> Centro Universitario de la Defensa de Zaragoza, Academia General Militar, Ctra. de Huesca s/n, 50090 Zaragoza, Spain

\* Correspondence: ddomingo@unizar.es; Tel.: +34-876-554-058

Received: 31 January 2018; Accepted: 19 March 2018; Published: 21 March 2018



**Abstract:** The account of total biomass can assist with the evaluation of climate regulation policies from local to global scales. This study estimates total biomass (TB), including tree and shrub biomass fractions, in *Pinus halepensis* Miller forest stands located in the Aragon Region (Spain) using Airborne Laser Scanning (ALS) data and fieldwork. A comparison of five selection methods and five regression models was performed to relate the TB, estimated in 83 field plots through allometric equations, to several independent variables extracted from ALS point cloud. A height threshold was used to include returns above 0.2 m when calculating ALS variables. The sample was divided into training and test sets composed of 62 and 21 plots, respectively. The model with the lower root mean square error (15.14 tons/ha) after validation was the multiple linear regression model including three ALS variables: the 25th percentile of the return heights, the variance, and the percentage of first returns above the mean. This study confirms the usefulness of low-density ALS data to accurately estimate total biomass, and thus better assess the availability of biomass and carbon content in a Mediterranean Aleppo pine forest.

**Keywords:** ALS; total biomass; shrub fraction; regression models; Aleppo pine

## 1. Introduction

Forest ecosystems and their associated understory act as important carbon sinks, providing habitats for wildlife [1,2] and promoting economic and social services to societies [3,4]. The sequestration of carbon and CO<sub>2</sub> by forest biomass and soils plays a major role in managing greenhouse gas emissions [5] providing a low-cost opportunity in climate policies [6]. The estimation of tree and shrub biomass in the Mediterranean Basin also contributes to better understanding fire behavior, which constitutes one of the most relevant disturbances as well as being a source of greenhouse gas emission to the atmosphere [7–11]. The inclusion of understory vegetation in carbon sequestration has been traditionally ignored [12] since it represents lower amounts of biomass compared with tree aboveground biomass (AGB). However, considering the global scale this amount constitutes an important pool for carbon sequestration [12,13].

Remote sensing technologies have demonstrated effectiveness for estimating forest resources such as biomass [14,15]. Particularly, Airborne Laser Scanning (ALS) is currently considered one of the best

tools due to its capability to provide 3-D information of vegetation structure. Vertical forest structure has been estimated with ALS data for several applications, such as forest inventory [16–18], forest structural heterogeneity [19–22], fuel type mapping [23,24] fuel modelling [23–26] or tree damage detection after natural disasters [27–29] for several height strata. However, few studies have focused on shrub biomass characterization with ALS data [30–33]. Some studies have used low density ALS data to estimate forest biomass [25,34–38], but little research has been performed including shrub vegetation because of the inherent difficulty in the estimation related to its low height and uniform surface [30]. Several studies state that ALS data tends to underestimate shrub vegetation [39–42]. Besides, when shrub and tree vegetation cover is high [43] and density of ALS data is low, the accuracy of digital elevation models (DEM) used to normalize return heights decreases [30]. The performed studies use an approach that combines ALS data and harvesting field measurements for biomass estimation [30,33]. In this sense, the lack of more studies to characterize shrub vegetation might have been associated with the necessary destructive sampling to generate forest structure equations, the assumption of simple geometric shapes [44,45], and the additional difficulty to estimate biomass at a regional scale using low-density ALS data. However, the presence of shrub biomass in the understory or the existence of shrubland areas constitutes an important land use in the Mediterranean basin. In this sense, the availability of shrub allometric equations for the main Spanish shrub species [46] have opened new opportunities. These equations allow the estimation of shrub biomass from simple field measurements, such as height and percentage cover per species or group of species. Thus, an area-based approach could be used to estimate biomass, using larger field plots (15 m radius) than traditional ones (1.5 m radius). This approach seems to be less affected by low-density ALS data, as has been proven in accurately estimating forest structure parameters in heterogeneous forests [47].

Few studies have compared different selection methods for ALS modelling [48,49]. In this sense, LASSO selection and the varimax rotation for Principal component analysis (PCA) selection are proposed as novel selection processes for biomass estimation. Some studies have compared different regression models to estimate forest parameters using high point cloud density [16,34,50–52], but only Li et al. [33] have compared two regression models to estimate shrub biomass.

The main objective of this study is to estimate the total biomass (TB) in heterogeneous *Pinus halepensis* Miller (hereinafter Aleppo pine) forest stands located in the Aragón Region (Spain). In this manuscript, TB refers to the dry weight of the plant material from trees, including roots, stems, bark, branches, and leaves from the ground to the apex, as well as the aboveground plant material from shrubs. The values of TB are expressed per unit area in terms of density, i.e., tons of TB per hectare (tons/ha). Specifically, we aim to (1) examine the relationship between biomass calculated at the field plot using allometric equations and ALS-derived variables; and (2) compare five variable selection methods and five different regression models, including regression trees and machine learning algorithms. This paper is organized in five sections. Section 2 describes the processing of ALS and field data, the selection methods, the regression models applied, as well as, the model comparison and validation methods used. Section 3 presents the results including the best TB model. Section 4 describes the discussion of the results and the paper's conclusions are included in Section 5.

## 2. Materials and Methods

### 2.1. Study Area

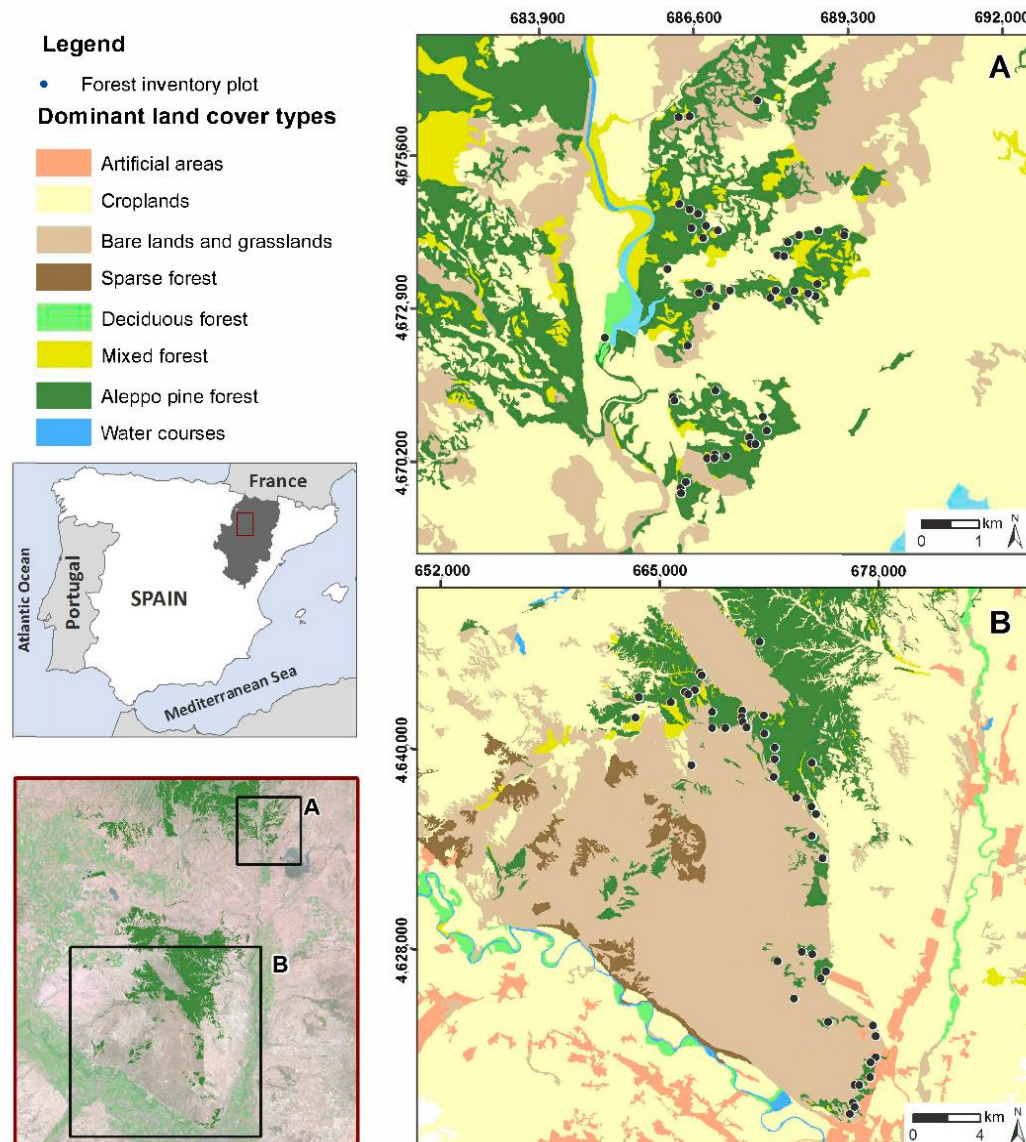
The Aleppo pine forests under study are located in the Ebro Basin (Figure 1), in Northeast Spain. This species represents almost 50% of the forested tree area in Aragón and is well adapted to Mediterranean environmental conditions. The two selected areas (Figure 1) are representative of the Aleppo pine forest in Central Ebro Valley, occupying 11,400 ha.

Climate of the region is Mediterranean with continental features. The average annual temperature is  $\approx 13$  °C, ranging from winter mean temperatures lower than 6 °C, to summer mean temperatures of



≈21 °C. Annual precipitation is medium-low and irregular, averaging 455 mm and mostly occurring in autumn and spring [53].

Aleppo pine forests are characterized by a hilly topography, with elevations ranging from 300 to 1150 m above sea level and slopes from 0° to 39°. The lithology of the study area varies from Miocene shales and sandstones, alternating with conglomerates in area A (Figure 1A) to Miocene carbonate and marl sediments in area B (Figure 1B).



**Figure 1.** Study area. Land cover types and location of 83 forest inventory plots. High spatial resolution orthophotography of the Spanish National Plan for Aerial Orthophotography [54] is used as backdrop. Coordinate System: ETRS89 UTM Zone 30 N.

Most of these pine stands are semi-natural, although some stands located in the south-eastern part of the area B (Figure 1B) were planted approximately forty years ago. The evergreen understory includes species such as *Quercus ilex* subsp. *rotundifolia*, *Quercus coccifera*, *Juniperus oxycedrus*, *Buxus sempervirens*, *Juniperus phoenicea*, *Rosmarinus officinalis*, and *Thymus vulgaris*.

## 2.2. ALS Data

The ALS data were collected by the Spanish National Plan for Aerial Orthophotography (PNOA) and captured in several surveys carried out between January and February 2011. An airborne Leica ALS60 discrete-return sensor with a small-footprint oscillating-mirror was used. The sensor was operated at a wavelength of 1.064  $\mu\text{m}$  and a scan angle of  $\pm 29^\circ$  from nadir. Data were provided by the National Geographic Information Centre (CNIG) in  $2 \times 2$  km tiles of raw data points in LAS binary files format v. 1.2. The geodetic reference system for x, y, and z coordinates was ETRS 1989 UTM Zone 30 N and heights were orthometrically corrected. The point cloud was captured with up to four returns measured per pulse. The point cloud density was 1.5 point/ $\text{m}^2$ , considering all returns with a vertical accuracy better than 0.2 m.

## 2.3. Field Plot Data and Conversion of Biomass Measurements to ALS Campaign Year

Field data were acquired in 83 circular plots, 15 m radius, at the two study areas (Figure 1). The sampling data fulfil the statistical requirements [55] and consider the size of the study area and the variability of pine forest in terms of canopy cover (CC), canopy height, and terrain slope [56]. The 45 field plots located in area A were collected from June to July 2015 and the 38 field plots located in area B were collected from July to September 2014. A stratified random sampling technique was selected to establish the location of the field plots, considering a representative sample of the tree density, forest structure [47], and terrain variability [57]. To perform this procedure, tree height (h), CC and terrain slopes of the study area were derived from ALS point cloud to define homogeneous areas.

The center of the designated circular plots was located in the field using a Leica VIVA<sup>®</sup> GS15 CS10 GNSS real-time kinematic Global Positioning System. The average accuracy of the planimetric coordinates was 0.16 m. The green crown height and h parameter were measured using a Haglöf Sweden<sup>®</sup> Vertex instrument and the tree diameter at breast height (dbh) was measured at 1.3 m, using a Haglöf Sweden<sup>®</sup> Mantax Precision Blue diameter calliper. Trees with a dbh larger than 7.5 cm were measured in each plot, and trees with a dbh lower than 7.5 cm were only counted. The percentage of canopy cover was estimated, as well as the average height of the different shrub species and the understory coverage.

The temporal lag between ALS acquisition data and field work campaign was 3 and 4 years in area B and area A, respectively. Although no significant changes took place in the study area in that period, field plot data were transformed to match the ALS campaign year. For this purpose, the values of dbh and height growth per diametric class provided by the National Forest Inventory (NFI) were used. These refer to the difference of years between the second NFI (NFI2) and the third one (NFI3), which is eleven years. A linear interpolation based on stand tables was made to estimate the variation between field data and ALS data acquisition (Table 1).

**Table 1.** Values of tree diameter at breast height (dbh) and height growth between the second National Forest Inventory (NFI2) and the third one (NFI3) (11 years) and subtractive values of dbh and height when applying linear interpolated degrowth to match Airborne Laser Scanning (ALS) year.

| Diametric Class (cm) | dbh Growth (mm) | Height Growth (m) | Area A   |            | Area B   |            |
|----------------------|-----------------|-------------------|----------|------------|----------|------------|
|                      |                 |                   | dbh (mm) | Height (m) | dbh (mm) | Height (m) |
| <10                  | 24              | 1.20              | −6.55    | −0.33      | −8.73    | −0.44      |
| 10–15                | 29              | 1.40              | −7.91    | −0.38      | −10.55   | −0.51      |
| 15–20                | 33              | 1.50              | −9.00    | −0.41      | −12.00   | −0.55      |
| 20–25                | 30              | 1.40              | −8.18    | −0.38      | −10.91   | −0.51      |
| 25–30                | 30              | 1.40              | −8.18    | −0.38      | −10.91   | −0.51      |
| 30–35                | 33              | 1.10              | −9.00    | −0.30      | −12.00   | −0.40      |
| 35–40                | 32              | 1.50              | −8.73    | −0.41      | −11.64   | −0.55      |
| 40–45                | 27              | 1.60              | −7.36    | −0.44      | −9.82    | −0.58      |
| 45–50                | 24              | 1.90              | 6.55     | 0.52       | 8.73     | 0.69       |
| 50–55                | 58              | 1.00              | −15.82   | −0.27      | −21.09   | −0.36      |
| 55–60                | 10              | 0.50              | −2.73    | −0.14      | −3.64    | −0.18      |



Total tree biomass fractions were calculated using the *Pinus halepensis* allometric equations according to Ruiz-Peinado et al. [58] (Equations (1)–(5)):

$$\begin{aligned} W_s(\text{kg}) &= 0.0139 \times \text{dbh}^2 \times h \\ W_{b7}(\text{kg}) &= 3.926 \times (\text{dbh} - 27.5) \times Z; \end{aligned} \quad (1)$$

$$\text{If dbh} \leq 27.5 \text{ cm then } Z = 0; \text{ If dbh} > 27.5 \text{ cm then } Z = 1 \quad (2)$$

$$W_{b2-7}(\text{kg}) = 4.257 + 0.00506 \times \text{dbh}^2 \times h - 0.0722 \times \text{dbh} \times h \quad (3)$$

$$W_{b2+n}(\text{kg}) = 6.197 + 0.00932 \times \text{dbh}^2 \times h - 0.0686 \times \text{dbh} \times h \quad (4)$$

$$W_r(\text{kg}) = 0.0785 \times \text{dbh}^2 \quad (5)$$

where  $W_s$  is the biomass weight of the stem fraction,  $W_{b7}$  is the biomass weight of the thick branch fraction (diameter larger than 7 cm),  $W_{b2-7}$  is the biomass weight of medium branch fraction (diameter between 2 and 7 cm),  $W_{b2+n}$  is the biomass weight of the thin branch fraction (diameter smaller than 2 cm) with needles, and  $W_r$  is the biomass weight of the roots.

Subsequently, the biomass fractions were summed up to calculate the total tree biomass of each individual tree. Then, these biomass values were added to obtain per plot biomass values that were later expressed in tons per hectare.

Shrub biomass was calculated at plot level using Montero et al. [46] equations for the different shrub formations of the study area according to the Spanish Forest Map (MFE) categories [59] (Equations (6)–(10)):

Shrub hedges, borders, galleries, etc.:

$$\ln(W) = 0.494 \times \ln(CC) \quad (6)$$

*Quercus coccifera* and *Pistacia lentiscus*:

$$\ln(W) = -2.892 + 1.505 \times \ln(h_m) + 0.462 \times \ln(CC) \quad (7)$$

*Leguminosae aulagoideas* and related shrubs:

$$\ln(W) = -2.464 + 0.808 \times \ln(h_m) + 0.761 \times \ln(CC) \quad (8)$$

*Labiatae* and *thymus* formations:

$$\ln(W) = -1.877 + 0.643 \times \ln(h_m) + 0.661 \times \ln(CC) \quad (9)$$

General shrub biomass:

$$\ln(W) = -2.560 + 1.006 \times \ln(h_m) + 0.672 \times \ln(C) \quad (10)$$

where  $W$  is the biomass weight for each species in tons/ha,  $h_m$  is the average shrub height at plot level and  $CC$  is the percentage of shrub canopy cover at plot level.

Equation (6) was applied for *Crataegus monogina*, *Rhamnus lycioides* and *Rosa canina*; Equation (7) was used for *Quercus coccifera*; Equation (8) was applied for *Genista scorpius*; Equation (9) was used for *Thymus* sp. Subsequently, general Equation (10) was used for all the other inventoried species: *Rosmarinus officinalis*, *Juniperus oxycedrus*, *Juniperus sabina*, *Buxus sempervirens*, *Genista florida*, and *Salsola vermiculata*. It should be noted that Equation (10) was also applied for *Quercus ilex* and *Pinus halepensis* with less than 7.5 cm of dbh, as it is not possible to use available tree allometric equations. Shrub biomass in every plot was calculated summing up the biomass values for each type of species.

The estimation of shrub biomass to match ALS year was carried out by applying shrub biomass growing equations according to Montero et al. [46] (Equations (11)–(14)). It should be noted that no growing equations were developed by Montero et al. [46] for shrub edges, borders, and gallerie formation. Consequently, the General shrub growing biomass (Equation (14)) was applied for *Crataegus monogina*, *Rhamnus lycioides*, and *Rosa canina*.

*Quercus coccifera* and *Pistacia lentiscus*:

$$\ln(W) = -4.955 + 1.150 \times \ln(h_m) + 0.463 \times \ln(CC) \quad (11)$$

*Leguminosae aulagoideas* and related shrubs:

$$\ln(W) = -4.479 + 0.715 \times \ln(h_m) + 0.701 \times \ln(CC) \quad (12)$$

*Labiatae* and *thymus* formations:

$$\ln(W) = -4.446 + 0.753 \times \ln(h_m) + 0.573 \times \ln(CC) \quad (13)$$

General shrub biomass:

$$\ln(W) = -4.771 + 0.814 \times \ln(h_m) + 0.676 \times \ln(CC) \quad (14)$$

where  $W$  is the biomass weight for each species in tons/ha,  $h_m$  is the average shrub height at plot level, and  $CC$  is the percentage of shrub canopy cover fraction at plot level. Then, the shrub biomass growing values per year in every plot were calculated summing up the biomass values for each species obtained from Equations (10)–(13). These annual growing values were subtracted from the measured ones considering the difference in years between ALS flight and field data acquisition.

Finally, TB density was calculated for each plot by adding up the tree biomass and the above ground shrub biomass expressed in tons of dry biomass per hectare.

## 2.4. ALS Data Processing

The noise point class was removed and ground points were classified according to Montealegre et al. [60] using the multiscale curvature classification algorithm [61]. This algorithm is implemented in MCC 2.1 command-line tool as a C++ application. The Point-TIN-Raster interpolation method [61], implemented in ArcGIS 10.5 software, was applied to generate a digital elevation model (DEM) with 1 m size grid [60]. Point heights were normalized with the DEM and ALS tiles were clipped to the spatial extent of each field plot. Furthermore, a full suite of statistical metrics commonly used as independent variables in forestry was calculated [62] using FUSION LDV 3.60 open source software [63] including variables related to height distribution: percentiles, mean elevation, kurtosis, skewness, etc.; and variables related to the percentage of canopy cover. According to Nilsson [64] and Næsset and Økland [57], a threshold value of 0.2 m height was applied to remove ground laser hits while considering the understory. The threshold is related to the ALS sensor precision in coordinate  $Z$ , which has a root mean square error (RMSE) below 0.2 m.

## 2.5. Selection of ALS Variables

Five selection processes were analyzed to select the ALS independent variables: (1) Spearman's rank correlation; (2) Stepwise selection; (3) Principal component analysis (PCA) and Varimax rotation; (4) LASSO selection; and (5) All subset selection. These methods are briefly described below.

Spearman's rank correlation coefficient ( $\rho$ ) was computed using R to determine the strength and direction of the relationship between field plot biomass data and ALS data. The selection of the ALS variables was made considering a minimum positive and negative  $\rho$  value of 0.5.

Stepwise selection method is based on dropping or adding variables at several steps. Backward, forward and bidirectional stepwise selection procedures were applied using R.

PCA allows the number of variables for regression purposes to be reduced. Although it has been widely used in many research fields, it is not a common approach for ALS metric selection [49]. PCA was computed using R package “lattice”. The obtained components with eigenvalues greater than 0.1 were retained according to the Kaiser Criterion [65]. A Varimax rotation, which maximizes the sum of the variance, was applied to better interpret the PCA results [66,67]. The three most representative metrics for all the PCs were selected.

LASSO (Last absolute shrinkage and selection operator) is a technique based on regularization methods. It generates interpretable models by minimizing the residual sum of squares regarding the sum of the absolute value of the coefficients that are less than a constant [68]. LASSO was computed in R using the “glmnet” package.

All subset selection methods allow a suitable group of metrics to be selected on which the models can focus their attention, while disregarding the rest of the variables. A wide variety of search approaches have been developed for subset selection. In this regard, four search approaches were implemented using R package “leaps”: exhaustive, backward, forward, and sequential replacement (Seqrep) [69]. The maximum size of subsets was set to three.

A maximum number of three predictor variables were selected using the different selection methods to avoid overfitting [70] and to generate more understandable models for forest management [38].

## 2.6. Parametric and Nonparametric Models for Estimating TB

The performance of five regression methods was compared to estimate TB, including a parametric model, the multiple linear regression model (MLR), and four nonparametric models: Support vector machine (SVM), Random forest (RF), Locally weighted linear regression (LWLR) and Linear model with a minimum length principle (MDL). Parametric models summarize data using a finite set of parameters while in nonparametric methods the number of parameters is potentially infinite. These methods are described below.

MLR is one of the most broadly applied methods for the estimation of forest structure variables using ALS data [38,57,71–75]. MLR is a linear approach for modeling the relationship between a scalar dependent variable and two or more explanatory variables. The selected variables, considering the different selection methods, were included in the models. The basic assumptions of linear regression models were verified for the fitted model: normality of the residuals, homoscedasticity and independence or no auto-correlation in the residuals [76] in order to make comparisons between suitable models. Dependent and independent variables were transformed logarithmically in those cases where statistical assumptions of linear regression were not fulfilled [71,74,77] to verify whether the measures of goodness of fit of the models improved.

SVM is a supervised learning model which has associated learning algorithms that analyze and recognize patterns. This method tries to discover among multidimensional hyperplanes the optimal separation between classes, where the separability is a maximum, assuming that input data are separable in space [38,77]. Data located in the hyperplane are called support vectors, being the most difficult to classify since they have a lower separability. SVM, with linear and radial kernels, were computed using R package “e1071”. The models were tuned applying a cost parameter within the interval 1–1000 and a gamma parameter in the interval 0.01–1.

RF is an ensemble learning method that adds randomness to bagging, increasing decision tree diversity by growing them from distinct subsets. RF combines a decision tree, considering the values of an independent random vector sample, with the same distribution for all trees in the forest [78]. RF divides the nodes of each decision tree using the best variables from a random sample. RF was computed using R “randomForest” [79] and “caret” packages [80]. The model was tuned by applying

a number of trees to growth (ntrees) within the interval 1–1000 and a number of variables selected between 1 and 2.

Two regression tree structures based on “If Then” rules were computed using the R package “CORElearn”. Those nonparametric techniques differ between them in the regression model applied in the leaf nodes. The first one, the LWLR, or loess, is a method which smoothes the dependent variable as a function of the independent ones to fit a regression surface to data [81]. The coefficient of smoothness is defined by computing the weighted mean square error considering a distance function. The second one, the MDL, considers that regularities in a set of data can be used to compress the data by using fewer symbols from a finite alphabet to explain the data faithfully. This method is based on the rule developed by Rissanen [82].

### 2.7. Model Comparison and Validation

Two types of models were compared: (i) models that include the same independent variables as the MLR one; and (ii) models that use two or three ALS variables, being different from those included in the first comparison.

The original sample was split into a training set of 75% of the cases (62 plots) and a test set of 25% of the cases (21 plots). Validation was performed for all the models, being executed 100 times for those methods with associated randomness as SVM, RF, LWLR, and MDL to increase robustness in the results [34]. In addition, data were normalized in values ranging from 0 to 1 in order to avoid weights saturation according to Görgens [52].

The model performance was compared in terms of root mean square error (RMSE), relative RMSE (%RMSE) and bias. Furthermore, the Friedman nonparametric test was applied using the RMSE of each plot in order to determine differences between models [83]. The Nemenyi post-hoc test was applied to determine whether differences were statistically significant, with a significance level of 0.05 [84]. This test was used only for those cases in which the null-hypothesis of the Friedman test was rejected, implying non-equivalence between models.

## 3. Results

Table 2 shows a summary of the field plot characteristics. The heights of the inventoried trees range from 3.3 to 16.8 m and they present a variety of diameters, from 8.0 to 27.5 cm. The shrubs present heights ranging from 0.3 to 2.0 m and a CC from 1% to 40%. The average tree biomass fraction is the most important fraction representing almost 75% of TB, and shrub biomass represents in average around 25% of TB.

**Table 2.** Summary of the field plot characteristics ( $n = 83$ ). CC is canopy cover; TB is total biomass.

|                         | Min. | Max.   | Range  | Mean  | Standard Deviation |
|-------------------------|------|--------|--------|-------|--------------------|
| Slope (degrees)         | 0.70 | 29.80  | 29.10  | 26.10 | 6.85               |
| Tree height (m)         | 3.32 | 16.77  | 13.44  | 8.04  | 2.27               |
| Tree dbh (cm)           | 8.02 | 27.54  | 19.53  | 16.42 | 4.11               |
| Tree biomass (tons/ha)  | 1.27 | 251.41 | 250.14 | 59.58 | 37.15              |
| Shrub height (m)        | 0.30 | 2.06   | 1.76   | 1.07  | 0.44               |
| Shrub CC (%)            | 1.00 | 40.00  | 39.00  | 11.52 | 6.09               |
| Shrub biomass (tons/ha) | 0.70 | 47.77  | 47.07  | 19.23 | 12.91              |
| TB (tons/ha)            | 7.66 | 253.14 | 245.48 | 78.81 | 42.34              |

The ALS metrics selected by the five selection processes are presented in Table 3. Considering the maximum number of selected variables established, the selection methods included generally one metric related to the canopy height (CHM), one metric associated with the canopy height variability (CHVM), and another expressing the canopy density (CDM). However, sequential replacement all subsets regression, forward all subsets regression, and PCA selected only two metrics.

Regarding CHM, low height percentiles were selected in some cases, but elevation mean and higher percentiles were chosen for the remaining methods. The most selected CHVM was variance. The percentage of first returns above mean and the percentage of all returns above a height break were the CDM metrics with the strongest correlation with TB. All selected metrics showed a logical relation with the dependent variable.

**Table 3.** ALS selected metrics using different variable selection methods, when applying a threshold value of 0.2 m to the point cloud data. Seqrep is sequential replacement; P refers to the percentile e.g., P05 is the 5th percentile; Elev. is Elevation; ret. is returns; Elev. AAD is average absolute deviation.

| Selection Method            |               | Metrics         |                          |   |
|-----------------------------|---------------|-----------------|--------------------------|---|
|                             |               | CHM             | CHVM                     | CDM   |
| Spearman's rank correlation |               | Elev. mean      | Elev. variance           | (All ret. above mean)/(Total first ret.) $\times$ 100 |
| Stepwise                    | Backward      | Elev. mean      | Elev. variance           | (All ret. above mean)/(Total first ret.) $\times$ 100 |
|                             | Forward       | P20             | Elev. variance           | % First ret. above mean                               |
|                             | Bidirectional | Elev. mean      | Elev. variance           | (All ret. above mean)/(Total first ret.) $\times$ 100 |
| PCA                         |               | P99             |                          | % All ret. above 0.2 m                                |
| LASSO selection             |               | P40             | Elev. standard deviation | % First ret. above 0.2 m                              |
| All subsets regression      | Seqrep        | P05, Elev. mean |                          | % First ret. above mean                               |
|                             | Exhaustive    | P25             | Elev. variance           | % First ret. above mean                               |
|                             | Backward      | P90             | Elev. AAD                | % First ret. above mean                               |
|                             | Forward       | P70             |                          | % First ret. above mean                               |

The regression models to estimate TB are summarized in Table 4. The MLR and SVM with radial kernel or linear kernel present the lowest RMSE. LWLR, MDL, and RF show lower accuracies. Although most of the models present values of bias close to zero, some of them show a slight overestimation with values close to 1. The selected MLR model including the 25th percentile of height, elevation variance, and the percentage of first returns above mean, present the lowest RMSE with 15.14 tons/ha, a %RMSE of 19.21, and a relative  $R^2$  after validation of 0.87. SVM with radial kernel including the 5th percentile of height, elevation mean, and percentage of first returns above mean is the second best model. SVM with linear kernel, including the same ALS metrics as the MLR method, presents also similar errors.

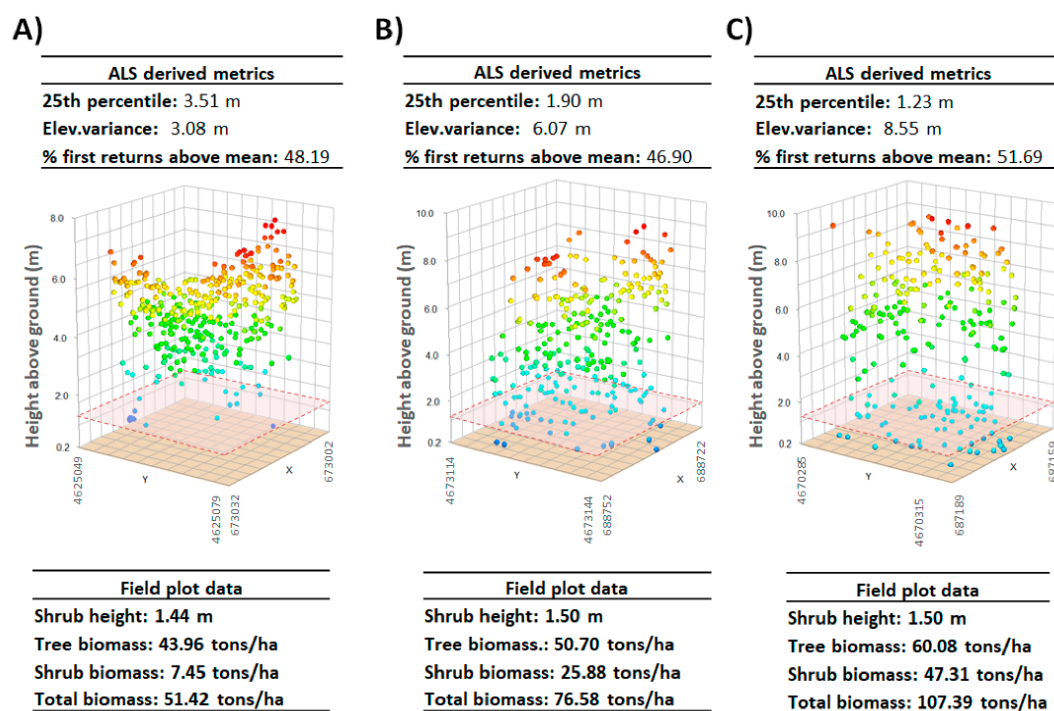
**Table 4.** Summary of the models and validation results in terms of RMSE (tons/ha), %RMSE, and bias (tons/ha). SVM r. refers to Support Vector Machine with radial kernel; SVM l. refers to Support Vector Machine with linear kernel; ret. is returns.

| ALS Metrics  | Model  | Fitting Phase |       |       | Cross-Validation |       |       |       |
|--|--------|---------------|-------|-------|------------------|-------|-------|-------|
|  |        | RMSE          | %RMSE | Bias  | RMSE             | %RMSE | Bias  | $R^2$ |
| Elev. Variance + P25 + % first ret. above mean                                     | MLR    | 14.53         | 18.44 | 0.00  | 15.14            | 19.21 | 0.01  | 0.87  |
|  | SVM r. | 13.94         | 17.68 | −0.36 | 16.39            | 20.79 | 0.00  | 0.86  |
|  | SVM l. | 14.56         | 18.48 | −0.74 | 15.56            | 19.74 | −0.83 | 0.86  |
|  | RF     | 10.01         | 12.70 | 0.39  | 19.64            | 24.93 | 0.75  | 0.80  |
|  | LWLR   | 12.11         | 15.37 | 0.20  | 19.59            | 24.86 | 1.35  | 0.78  |
|  | MDL    | 14.01         | 17.77 | 0.03  | 17.98            | 22.81 | 0.37  | 0.82  |
| P05 + Elev. mean + % first ret. above mean   | SVM r. | 13.88         | 17.61 | −0.28 | 15.50            | 19.66 | −0.26 | 0.86  |
| Elev. mean + Elev. variance + (All ret. above mean)/(total first ret. above 0.2 m) | RF     | 9.64          | 12.23 | 0.39  | 19.14            | 24.28 | 0.39  | 0.80  |
| P40 + Elev. Std.dev + % first ret. above 0.2 m                                     | LWLR   | 11.38         | 14.44 | −0.45 | 19.38            | 24.59 | −0.54 | 0.80  |

Figure 2 shows the independent variables included in the selected MLR model associated with the vertical distribution of ALS returns in three selected field plots representative of the variability of TB of the Aleppo pine forest under study. The TB of smaller pines with low shrub presence (Figure 2A) is lower than in taller pines with higher presence of understory (Figure 2C). Thereby, the TB presents a direct relationship with these variables (Figure 2B). The higher the variance, the higher the TB as



the dispersion of the data increases. Similarly, the increase of the percentage of first returns above the mean implies an increase in TB content. The 25th percentile presents an inverse correlation with the TB. A lower value of the 25th percentile implies a higher presence of shrubs in the understory.



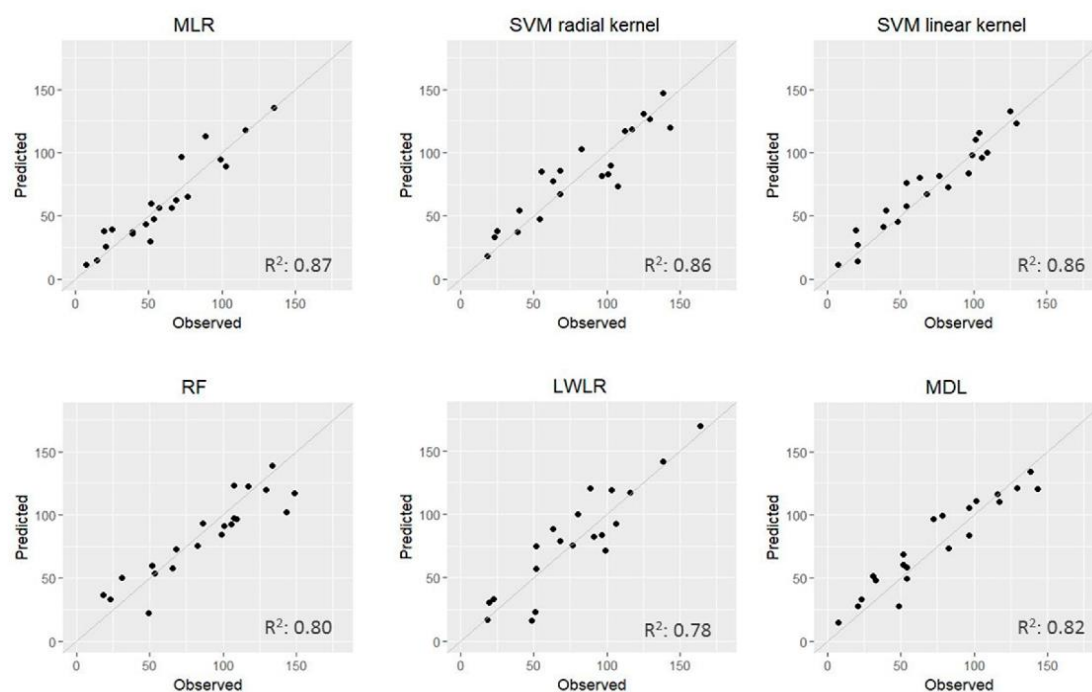
**Figure 2.** Airborne Laser Scanning (ALS) metrics associated with the vertical distribution of point cloud returns in three selected field plots representative of the study area: (A) smaller pines with low presence of shrubs (area B); (B) average height pines with an intermediate presence of understory (area A); (C) taller pines with high presence of shrubs (area A). The red square represents the average of the shrub height in each plot.

Figure 3 presents the scatter plots of observed and predicted TB for the models that include the same variables as the best selected MLR ones. It should be noted that, in this model, logarithmic transformation of the dependent or the independent variables was not required to meet the linear regression assumptions. MLR, SVM with linear kernel and SVM with radial kernel show high coefficients of determination (0.87, 0.86, and 0.86, respectively), while MDL (0.82), RF (0.80), and LWLR (0.78) present slightly lower coefficients of determination, 0.80 and 0.78, respectively. The implementation of the MLR model (Equation (15)) in ArcGIS allowed the estimation of TB.

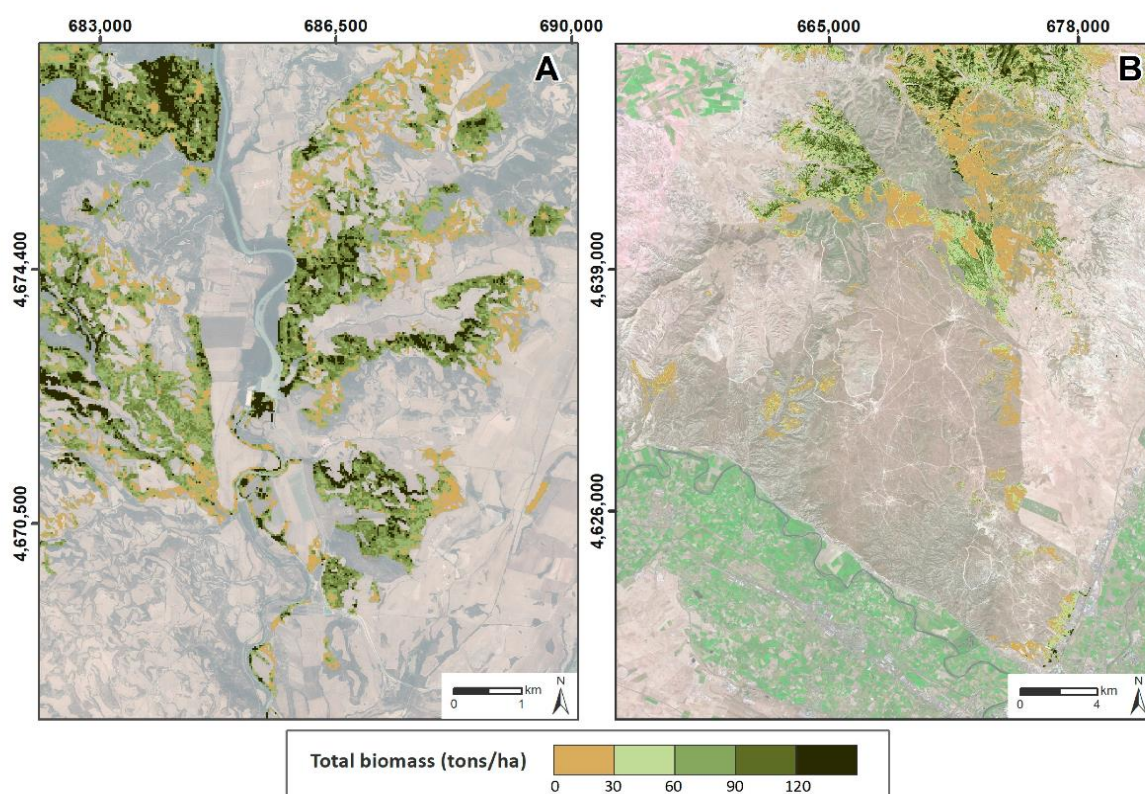
$$TB = -14195.1 + 4753.9 \times \text{Elevation variance} + 8992.0 \times P25 + 699.6 \times \% \text{ first returns above mean} \quad (15)$$

The performance comparison between models by using the Friedman test shows that the models are not equivalent with a  $p$ -value of 0.00. Thereby, the required application of the post-hoc Nemenyi test indicates that no statistically significant differences exist between models, with 95% of probability.

Figure 4 shows the TB mapping for the study area. The north part of the study area (Figure 4A) shows a heterogeneous Aleppo pine forest with higher TB values (above 150 tons/ha), which denote mature forested areas with the presence of shrubs and taller trees. In contrast, the southern part (Figure 4B) shows some homogeneous areas with lower TB values (below 10 tons/ha). These areas have been affected more recently by several wildfires, contrasting with the pines of area A or some north stands of area B that have higher ages, presenting TB values ranging from approximately 30 tons/ha to more than 120 tons/ha.



**Figure 3.** Scatterplot of predicted vs. observed values of total biomass (TB) (tons/ha) using different regression methods.



**Figure 4.** Maps of TB. High spatial resolution orthophotography from Spanish National Plan for Aerial Orthophotography (PNOA) [54] is used as backdrop. Coordinate System: ETRS89 UTM Zone 30 N.

#### 4. Discussion

The results demonstrate that low density ALS data can be used to accurately estimate TB. Regarding the selection method, all subsets regression and selection based on Spearman's rank was the most powerful technique for ALS metrics selection. LASSO selection was also a good technique. However, the use of Stepwise selection and PCA showed a restricted power in identifying the best subsets. These findings agree with Garcia-Gutierrez et al. [48], who found that Stepwise selection was the technique with less power, but all subsets regression obtained good results. Although PCA was used by Silva et al. [49] for stem volume modeling, it has been outperformed by other selection techniques for TB estimation.

The selection of a reduced number of biologically representative variables increases the understanding of models for forest management purposes. In this sense, the use of automatic methods for variable selection is considerably less time-demanding when modeling. The three ALS-derived variables included in the best model were coherent and analogous to those derived for characterizing forest vertical diversity [22] and aboveground biomass [15,36]. Variables derived from digital surface models (DSM) were not computed in this study. Nevertheless, they have been proposed in other studies because they seem more suitable for tree damage detection [29] or estimation of vertical diversity [19] when not using an area based approach (ABA). These variables concern the variability of different strata of the point cloud and canopy cover distribution.

The comparison between regression models shows that the MLR model had the lowest RMSE (15.14 tons/ha) and bias (0.01), matching with the values obtained by other authors for aboveground biomass estimation [73,85]. Thus, according to Görgens et al. [52] and Domingo et al. [38], MLR slightly outperforms other nonparametric methods. Although MLR was the best model, no statistically significant differences were found with other methods. These results partly agree with other approaches [34,36,48], which obtained lower estimation errors with nonparametric techniques. The use of selection methods is useful to better determine the best subset of variables and the comparison of models, such as MLR, SVM with radial kernel, SVM with linear kernel, and LWLR.

TB estimation by using trees and shrub allometric equations and low-density ALS data is considered to better account for biomass and carbon reservoirs in Mediterranean forest ecosystems. The comparison of different models for TB estimation including other tree species or biomass estimation in shrubland areas may be considered. In this sense, 35% of the forested land in Spain, which includes grasslands, shrubland, and trees, is covered by shrubs [83]. Furthermore, shrub species in the understory, with an average presence of 24%, constitute an important part of TB in Mediterranean Aleppo pine forest. This pattern is also common in the understory of other Mediterranean forests as for example *Pinus* and *Quercus ilex*. Shrubland areas are also especially important due to the abandonment of crops and wildfire disturbances [13], that might even increase the presence of these species in the near future. Consequently, further research is needed on the estimation of TB using tree and shrub allometric equations and low density ALS data in shrubland and burned areas to compare with previous higher point density studies. Moreover, the use of ALS technology with higher density, the use of terrestrial LiDAR, or the fusion with high spatial resolution images derived by unmanned aerial vehicles might be useful to better account for forest biomass.

#### 5. Conclusions

Allometric equations are one of the available methods for estimating forest structural parameters. This method has been widely applied to accurately estimate tree height, crown diameter, basal area, stem density, volume or aboveground biomass at plot level and to relate it to ALS data to estimate it at stand level [36,47,56,57,86]. Nevertheless, little research has focused on the estimation of biomass with ALS data including shrub fraction. The lack of knowledge of shrub vegetation associated with the traditionally necessary destructive sampling to generate forest structure equations might have been one of the drawbacks of including those species for biomass and carbon account. In this sense, the development of new allometric equations for shrub species, defined to estimate biomass from



simple field measurements and the generalization of ALS data have opened new opportunities for TB estimation.

This study assesses the usefulness of low point density ALS data to accurately estimate biomass, including tree and shrub fraction, in an Aleppo pine forest. The comparison of the effectiveness of five variable selection methods increased modeling efficiency. All subsets regression and manual selection based on Spearman's rank were the most powerful techniques. LASSO selection was also a good technique but, Stepwise selection and PCA showed less power to identify the best subsets. The comparison of parametric and nonparametric regression methods showed that the best model for TB estimation was the MLR. This model included three ALS metrics: the 25th percentile of the return heights, the variance, and the percentage of first returns above the mean, presenting an RMSE of 15.14 tons/ha and a bias of 0.01. No statistically significant differences between the generated models were found. The results allow TB mapping at the regional scale providing useful information for forest management purposes, with values ranging from below 10 tons/ha to areas above 150 tons/ha. This study moves a step forward to compute TB by using allometric tree and shrub equations and ALS data to better account for forest biomass as assets to manage greenhouse gases from a local to a global scale.

**Acknowledgments:** This work was supported by the Government of Spain, Department of Education Culture and Sports under Grant [FPU Grant BOE, 14/06250] and by the Department of Economy and Competitiveness under project SERGISAT [CGL2014-57013-C2-2-R]. The ALS data were provided by the Spanish National Geographic Information Centre (CNIG). The authors are grateful to the material resources provided by the Centro Universitario de la Defensa de Zaragoza (CUD). We express our appreciation to Lt. Col. Francisco Escribano for his help during the fieldwork.

**Author Contributions:** M.T.L. and A.L.M. had the original idea for the study. All co-authors conducted the fieldwork campaign and D.D., M.T.L., and A.L.M. developed the methodology and performed the analysis. The lead author wrote this paper, incorporating suggestions from all co-authors, who approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Martinuzzi, S.; Vierling, L.A.; Gould, W.A.; Falkowski, M.J.; Evans, J.S.; Hudak, A.T.; Vierling, K.T. Mapping snags and understory shrubs for LiDAR based assessment of wildlife habitat suitability. *Remote Sens. Environ.* **2009**, *113*, 2533–2546. [[CrossRef](#)]
2. Wing, B.M.; Ritchie, M.W.; Boston, K.; Cohen, W.B.; Gitelman, A.; Olsen, M.J. Prediction of understory vegetation cover with airborne LiDAR in an interior ponderosa pine forest. *Remote Sens. Environ.* **2012**, *124*, 730–741. [[CrossRef](#)]
3. Lal, R. Sequestration of atmospheric CO<sub>2</sub> in global carbon pools. *Energy Environ. Sci.* **2008**, *1*, 86–100. [[CrossRef](#)]
4. Pan, Y.; Birdsey, R.A.; Fang, J.; Houghton, R.; Kauppi, P.E.; Kurz, W.A.; Phillips, O.L.; Shvidenko, A.; Lewis, S.L.; Canadell, J.G.; et al. A large and persistent carbon sink in the world's forests. *Science* **2011**, *333*, 988–993. [[CrossRef](#)] [[PubMed](#)]
5. Hoen, H.F.; Solberg, B. Potential and Economic Efficiency of Carbon Sequestration in Forest Biomass through Silvicultural Management. *For. Sci.* **1994**, *40*, 429–451.
6. Gren, I.-M.; Aklilu, A.Z. Policy design for forest carbon sequestration: A review of the literature. *For. Policy Econ.* **2016**, *70*, 128–136. [[CrossRef](#)]
7. Seiler, W.; Crutzen, P.J. Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. *Clim. Chang.* **1980**, *2*, 207–247. [[CrossRef](#)]
8. Andreae, M.O.; Browell, E.V.; Garstang, M.; Gregory, G.L.; Harriss, R.C.; Hill, G.F.; Jacob, D.J.; Pereira, M.C.; Sachse, G.W.; Setzer, A.W.; et al. Biomass-burning emissions and associated haze layers over Amazonia. *J. Geophys. Res.* **1988**, *93*, 1509–1527. [[CrossRef](#)]
9. Van der Werf, G.R.; Randerson, J.T.; Giglio, L.; Collatz, G.J.; Mu, M.; Kasibhatla, P.S.; Morton, D.C.; DeFries, R.S.; Jin, Y.; van Leeuwen, T.T. Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmos. Chem. Phys.* **2010**, *10*, 11707–11735. [[CrossRef](#)]

10. Akagi, S.K.; Yokelson, R.J.; Burling, I.R.; Meinardi, S.; Simpson, I.; Blake, D.R.; McMeeking, G.R.; Sullivan, A.; Lee, T.; Kreidenweis, S.; et al. Measurements of reactive trace gases and variable O<sub>3</sub> formation rates in some South Carolina biomass burning plumes. *Atmos. Chem. Phys.* **2013**, *13*, 1141–1165. [[CrossRef](#)]
11. Wiedinmyer, C.; Akagi, S.K.; Yokelson, R.J.; Emmons, L.K.; Al-Saadi, J.A.; Orlando, J.J.; Soja, A.J. The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning. *Geosci. Model Dev.* **2011**, *4*, 625–641. [[CrossRef](#)]
12. Penman, J.; Gytarsky, M.; Hiraishi, T.; Krug, T.; Kruger, D.; Pipatti, R.; Buendia, L.; Miwa, K.; Ngara, T.; Tanabe, K.; et al. *Good Practice Guidance for Land Use, Land-Use Change and Forestry*; Intergovernmental Panel on Climate Change (IPCC); IGES: Hayama, Japan, 2003.
13. Pasalodos-Tato, M.; Ruiz-Peinado, R.; del Río, M.; Montero, G. Shrub biomass accumulation and growth rate models to quantify carbon stocks and fluxes for the Mediterranean region. *Eur. J. For. Res.* **2015**, *134*, 537–553. [[CrossRef](#)]
14. García, M.; Riaño, D.; Chuvieco, E.; Danson, F.M. Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sens. Environ.* **2010**, *114*, 816–830. [[CrossRef](#)]
15. Montealegre-Gracia, A.L.; Lamelas-Gracia, M.T.; García-Martín, A.; de la Riva-Fernández, J.; Escribano-Bernal, F. Using low-density discrete Airborne Laser Scanning data to assess the potential carbon dioxide emission in case of a fire event in a Mediterranean pine forest. *GISci. Remote Sens.* **2017**, *54*, 721–740. [[CrossRef](#)]
16. Latifi, H.; Nothdurft, A.; Koch, B. Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: Application of multiple optical/LiDAR-derived predictors. *Forestry* **2010**, *83*, 395–407. [[CrossRef](#)]
17. Vosselman, G.; Maas, H.G. *Airborne and Terrestrial Laser Scanning*; Whittles Publishing: Scotland, UK, 2010; ISBN 978-1904445-87-6.
18. Maltamo, M.; Næsset, E.; Vauhkonen, J. (Eds.) *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*; Springer: Dordrecht, The Netherlands, 2014; Volume 27, ISBN 978-94-017-8662-1.
19. Kane, V.R.; Lutz, J.A.; Roberts, S.L.; Smith, D.F.; McGaughey, R.J.; Povak, N.A.; Brooks, M.L. Landscape-scale effects of fire severity on mixed-conifer and red fir forest structure in Yosemite National Park. *For. Ecol. Manag.* **2013**, *287*, 17–31. [[CrossRef](#)]
20. Kane, V.R.; North, M.P.; Lutz, J.A.; Churchill, D.J.; Roberts, S.L.; Smith, D.F.; McGaughey, R.J.; Kane, J.T.; Brooks, M.L. Assessing fire effects on forest spatial structure using a fusion of Landsat and airborne LiDAR data in Yosemite National Park. *Remote Sens. Environ.* **2014**, *151*, 89–101. [[CrossRef](#)]
21. Listopad, C.M.C.S.; Masters, R.E.; Drake, J.; Weishampel, J.; Branquinho, C. Structural diversity indices based on airborne LiDAR as ecological indicators for managing highly dynamic landscapes. *Ecol. Indic.* **2015**, *57*, 268–279. [[CrossRef](#)]
22. Bottalico, F.; Chirici, G.; Giannini, R.; Mele, S.; Mura, M.; Puxeddu, M.; Mcroberts, R.E.; Valbuena, R.; Travaglini, D. Modeling Mediterranean forest structure using airborne laser scanning data. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *57*, 145–153. [[CrossRef](#)]
23. García, M.; Riaño, D.; Chuvieco, E.; Salas, J.; Danson, F.M. Multispectral and LiDAR data fusion for fuel type mapping using Support Vector Machine and decision rules. *Remote Sens. Environ.* **2011**, *115*, 1369–1379. [[CrossRef](#)]
24. Rosa, M.F.; Stow, D.A. Mapping fuels at the wildland-urban interface using colour ortho-images and LiDAR data. *Geocarto Int.* **2014**, *29*, 570–588. [[CrossRef](#)]
25. Hall, S.A.; Burke, I.C.; Box, D.O.; Kaufmann, M.R.; Stoker, J.M. Estimating stand structure using discrete-return lidar: An example from low density, fire prone ponderosa pine forests. *For. Ecol. Manag.* **2005**, *208*, 189–209. [[CrossRef](#)]
26. Riaño, D.; Meier, E.; Allgöwer, B.; Chuvieco, E.; Ustin, S.L. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sens. Environ.* **2003**, *86*, 177–186. [[CrossRef](#)]
27. Ranson, K.J.; Sun, G.; Kovacs, K.; Kharuk, V.I. Use of ICESat GLAS data for forest disturbance studies in central Siberia. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004; Volume 3, pp. 1936–1939. [[CrossRef](#)]

28. Honkavaara, E.; Litkey, P.; Nurminen, K. Automatic Storm Damage Detection in Forests Using High-Altitude Photogrammetric Imagery. *Remote Sens.* **2013**, *5*, 1405–1424. [\[CrossRef\]](#)
29. Rahman, M.T.; Rashed, T. Urban tree damage estimation using airborne laser scanner data and geographic information systems: An example from 2007 Oklahoma ice storm. *Urban For. Urban Green.* **2015**, *14*, 562–572. [\[CrossRef\]](#)
30. Estornell, J.; Ruiz, L.A.; Velázquez-Martí, B.; Hermosilla, T. Analysis of the factors affecting LiDAR DTM accuracy in a steep shrub area. *Int. J. Digit. Earth* **2011**, *4*, 521–538. [\[CrossRef\]](#)
31. Estornell, J.; Ruiz, L.A.; Velázquez-Martí, B.; Hermosilla, T. Estimation of biomass and volume of shrub vegetation using LiDAR and spectral data in a Mediterranean environment. *Biomass Bioenergy* **2012**, *46*, 710–721. [\[CrossRef\]](#)
32. Greaves, H.E.; Vierling, L.A.; Eitel, J.U.H.; Boelman, N.T.; Magney, T.S.; Prager, C.M.; Griffin, K.L. High-resolution mapping of aboveground shrub biomass in Arctic tundra using airborne lidar and imagery. *Remote Sens. Environ.* **2016**, *184*, 361–373. [\[CrossRef\]](#)
33. Li, A.; Dhakal, S.; Glenn, N.; Spaete, L.; Shinneman, D.; Pilliod, D.; Arkle, R.; McIlroy, S. Lidar Aboveground Vegetation Biomass Estimates in Shrublands: Prediction, Uncertainties and Application to Coarser Scales. *Remote Sens.* **2017**, *9*, 903. [\[CrossRef\]](#)
34. García-Gutiérrez, J.; Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing* **2015**, *167*, 24–31. [\[CrossRef\]](#)
35. Shendryk, I.; Hellström, M.; Klemetsson, L.; Kljun, N. Low-Density LiDAR and Optical Imagery for Biomass Estimation over Boreal Forest in Sweden. *Forests* **2014**, *5*, 992–1010. [\[CrossRef\]](#)
36. Guerra-Hernández, J.; Görgens, E.B.; García-Gutiérrez, J.; Carlos, L.; Rodriguez, E.; Tomé, M.; González-Ferreiro, E. European Journal of Remote Sensing Comparison of ALS based models for estimating aboveground biomass in three types of Mediterranean forest. *Eur. J. Remote Sens.* **2016**, *49*, 185–204. [\[CrossRef\]](#)
37. Montagnoli, A.; Fusco, S.; Terzaghi, M.; Kirschbaum, A.; Pflugmacher, D.; Cohen, W.B.; Scippa, G.S.; Chiatante, D. Estimating forest aboveground biomass by low density lidar data in mixed broad-leaved forests in the Italian Pre-Alps. *For. Ecosyst.* **2015**, *2*. [\[CrossRef\]](#)
38. Domingo, D.; Lamelas-Gracia, M.T.; Montealegre-Gracia, A.L.; de la Riva-Fernández, J. Comparison of regression models to estimate biomass losses and CO<sub>2</sub> emissions using low-density airborne laser scanning data in a burnt Aleppo pine forest. *Eur. J. Remote Sens.* **2017**, *50*, 384–396. [\[CrossRef\]](#)
39. Gaveau, D.L.A.; Hill, R.A. Quantifying canopy height underestimation by laser pulse penetration in small-footprint airborne laser scanning data. *Can. J. Remote Sens.* **2003**, *29*, 650–657. [\[CrossRef\]](#)
40. Hill, R.A.; Thomson, A.G. Mapping woodland species composition and structure using airborne spectral and LiDAR data. *Int. J. Remote Sens.* **2005**, *26*, 3763–3779. [\[CrossRef\]](#)
41. Hopkinson, C.; Chasmer, L.E.; Sass, G.; Creed, I.F.; Sitar, M.; Kalbfleisch, W.; Treitz, P. Vegetation class dependent errors in lidar ground elevation and canopy height estimates in a boreal wetland environment. *Can. J. Remote Sens.* **2005**, *31*, 191–206. [\[CrossRef\]](#)
42. Riaño, D.; Chuvieco, E.; Ustin, S.L.; Salas, J.; Rodríguez-Pérez, J.R.; Ribeiro, L.M.; Viegas, D.X.; Moreno, J.M.; Fernández, H. Estimation of shrub height for fuel-type mapping combining airborne LiDAR and simultaneous color infrared ortho imaging. *Int. J. Wildl. Fire* **2007**, *16*, 341. [\[CrossRef\]](#)
43. Meng, X.; Currit, N.; Zhao, K. Ground Filtering Algorithms for Airborne LiDAR Data: A Review of Critical Issues. *Remote Sens.* **2010**, *2*, 833–860. [\[CrossRef\]](#)
44. Loudermilk, E.L.; Hiers, J.K.; O'Brien, J.J.; Mitchell, R.J.; Singhanian, A.; Fernandez, J.C.; Cropper, W.P.; Slatton, K.C. Ground-based LIDAR: A novel approach to quantify fine-scale fuelbed characteristics. *Int. J. Wildl. Fire* **2009**, *18*, 676. [\[CrossRef\]](#)
45. Velázquez-Martí, B.; Fernández-González, E.; Estornell, J.; Ruiz, L.A. Dendrometric and dasometric analysis of the bushy biomass in Mediterranean forests. *For. Ecol. Manag.* **2010**, *259*, 875–882. [\[CrossRef\]](#)
46. González, G.M.; Pasalodos-Tato, M.; López-Senespiedra, E.; Onrubia, R.; Bravo-Oviedo, A.; Ruiz-Peinado, R. Contenido de Carbono en la biomasa de las principales especies de matorral y arbustados de España. In Proceedings of the 6th Congreso Forestal Español, Pontevedra, Spain, 10 June–14 June 2013; Sociedad Española de Ciencias Forestales: Pontevedra, Spain, 2013.

47. Montealegre, A.L.; Lamelas, M.T.; de la Riva, J.; García-Martín, A.; Escribano, F. Use of low point density ALS data to estimate stand-level structural variables in Mediterranean Aleppo pine forest. *Forestry* **2016**, *89*, 373–382. [[CrossRef](#)]
48. Garcia-Gutierrez, J.; Gonzalez-Ferreiro, E.; Riquelme-Santos, J.C.; Miranda, D.; Dieguez-Aranda, U.; Navarro-Cerrillo, R.M. Evolutionary feature selection to estimate forest stand variables using LiDAR. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 119–131. [[CrossRef](#)]
49. Silva, C.A.; Klauber, C.; Hudak, A.T.; Vierling, L.A.; Liesenberg, V.; Carvalho, S.P.C.E.; Rodriguez, L.C.E. A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. *Forestry* **2016**, *89*, 422–433. [[CrossRef](#)]
50. Gagliasso, D.; Hummel, S.; Temesgen, H. A Comparison of Selected Parametric and Non-Parametric Imputation Methods for Estimating Forest Biomass and Basal Area. *Open J. For.* **2014**, *4*, 42–48. [[CrossRef](#)]
51. Gleason, C.J.; Im, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote Sens. Environ.* **2012**, *125*, 80–91. [[CrossRef](#)]
52. Görgens, E.B.; Montagni, A.; Rodriguez, L.C.E. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Comput. Electron. Agric.* **2015**, *116*, 221–227. [[CrossRef](#)]
53. Cuadrat, J.M.; Saz, M.A.; Vicente-Serrano, S.M. *Atlas Climático de Aragón*; Gobierno de Aragón: Aragón, Spain, 2007.
54. Plan Nacional de Ortofotografía Aérea LiDAR. Available online: <http://pnoa.ign.es/presentacion> (accessed on 16 January 2018).
55. Hair, J.F.; Prentice, E.; Cano, D. *Análisis Multivariante*; Prentice-Hall: Madrid, Spain, 1999; ISBN 9788483220351.
56. Næsset, E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* **2002**, *80*, 88–99. [[CrossRef](#)]
57. Næsset, E.; Økland, T. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sens. Environ.* **2002**, *79*, 105–115. [[CrossRef](#)]
58. Ruiz-Peinado, R.; Del Rio, M.; Montero, G. New models for estimating the carbon sink capacity of Spanish softwood species. *For. Syst.* **2011**, *20*, 176. [[CrossRef](#)]
59. MAGRAMA Mapa Forestal de España 1:50.000 (MFE50). Available online: <http://www.magrama.gob.es/es/biodiversidad/servicios/banco-datos-naturaleza/informacion-disponible/mfe50.aspx> (accessed on 4 May 2016).
60. Montealegre, A.L.; Lamelas, M.T.; de la Riva, J. Interpolation Routines Assessment in ALS-Derived Digital Elevation Models for Forestry Applications. *Remote Sens.* **2015**, *7*, 8631–8654. [[CrossRef](#)]
61. Evans, J.S.; Hudak, A.T. A Multiscale Curvature Algorithm for Classifying Discrete Return LiDAR in Forested Environments. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1029–1038. [[CrossRef](#)]
62. Evans, J.S.; Hudak, A.T.; Faux, R.; Smith, A.M.S. Discrete Return Lidar in Natural Resources: Recommendations for Project Planning, Data Processing, and Deliverables. *Remote Sens.* **2009**, *1*, 776–794. [[CrossRef](#)]
63. FUSION Version Check. Available online: <http://forsys.sefs.uw.edu/fusion/fusionlatest.html> (accessed on 16 January 2018).
64. Nilsson, M. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sens. Environ.* **1996**, *56*, 1–7. [[CrossRef](#)]
65. Kaiser, H.F. The Application of Electronic Computers to Factor Analysis. *Educ. Psychol. Meas.* **1960**, *20*, 141–151. [[CrossRef](#)]
66. Horst, P. *Factor Analysis of Data Matrices*, 1st ed.; Holt, Rinehart and Winston: New York, NY, USA, 1965. [[CrossRef](#)]
67. Kaiser, H.F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **1958**, *23*, 187–200. [[CrossRef](#)]
68. Tibshirani, R. Regression shrinkage and selection via the lasso: A retrospective. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2011**, *73*, 273–282. [[CrossRef](#)]
69. Miller, A.J. *Subset Selection in Regression*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2002; ISBN 9781584881711.



70. Hermosilla, T.; Ruiz, L.A.; Kazakova, A.N.; Coops, N.C.; Moskal, L.M. Estimation of forest structure and canopy fuel parameters from small-footprint full-waveform LiDAR data. *Int. J. Wildl. Fire* **2014**, *23*, 224. [\[CrossRef\]](#)
71. Means, J.; Acker, S.; Harding, D.; Blair, J.; Lefsky, M.; Cohen, W.; Harmon, M.; McKee, W. Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western Cascades of Oregon. *Remote Sens. Environ.* **1999**, *67*, 298–308. [\[CrossRef\]](#)
72. Lim, K.; Treitz, P.; Wulder, M.; St-Onge, B.; Flood, M. LiDAR remote sensing of forest structure. *Prog. Phys. Geogr.* **2003**, *27*, 88–106. [\[CrossRef\]](#)
73. Gonzalez-Ferreiro, E.; Dieguez-Aranda, U.; Miranda, D. Estimation of stand variables in *Pinus radiata* D. Don plantations using different LiDAR pulse densities. *Forestry* **2012**, *85*, 281–292. [\[CrossRef\]](#)
74. García, D.; Godino, M.; Mauro, F. *Lidar: Aplicación Práctica Al Inventario Forestal*; Académica Española: Madrid, Spain, 2012; ISBN 3659009555.
75. Watt, M.S.; Meredith, A.; Watt, P.; Gunn, A.; Andersen, H.; Reutebuch, S.; Nelson, R. Use of LiDAR to estimate stand characteristics for thinning operations in young Douglas-fir plantations. *N. Z. J. For. Sci.* **2013**, *43*, 1–10. [\[CrossRef\]](#)
76. González-Olabarria, J.-R.; Rodríguez, F.; Fernández-Landa, A.; Mola-Yudego, B. Mapping fire risk in the Model Forest of Urbión (Spain) based on airborne LiDAR measurements. *For. Ecol. Manag.* **2012**, *282*, 149–156. [\[CrossRef\]](#)
77. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [\[CrossRef\]](#)
78. Breiman, L. Random Forest. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
79. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
80. Van Essen, D.C.; Drury, H.A.; Dickson, J.; Harwell, J.; Hanlon, D.; Anderson, C.H. An integrated software suite for surface-based analyses of cerebral cortex. *J. Am. Med. Inform. Assoc.* **2001**, *8*, 443–459. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Cleveland, W.S.; Devlin, S.J. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *J. Am. Stat. Assoc.* **1988**, *83*, 596–610. [\[CrossRef\]](#)
82. Rissanen, J. Modeling by shortest data description. *Automatica* **1978**, *14*, 465–471. [\[CrossRef\]](#)
83. Stojanova, D.; Panov, P.; Gjorgjioski, V.; Kobler, A.; Džeroski, S. Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecol. Inform.* **2010**, *5*, 256–266. [\[CrossRef\]](#)
84. Nemenyi, P. Distribution-Free Multiple Comparisons. Ph.D. Thesis, Princeton University, Princeton, NY, USA, 1963.
85. Montealegre, A.L.; Lamelas, M.T.; de la Riva, J.; García-Martín, A.; Escibano, F. Assessment of Biomass and Carbon Content in a Mediterranean Aleppo Pine Forest Using ALS Data. In Proceedings of the 1st International Electronic Conference on Remote Sensing, Basel, Switzerland, 22 June–5 July 2005; MDPI: Basel, Switzerland, 2015; Volume 1, p. d004.
86. Popescu, S.C.; Wynne, R.H.; Nelson, R.F. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Can. J. Remote Sens.* **2003**, *29*, 564–577. [\[CrossRef\]](#)



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



### 3.3. Quantifying forest residual biomass in *Pinus halepensis* Miller stands using Airborne Laser Scanning data

Cuantificación de biomasa residual forestal en masas de *Pinus halepensis* Miller. utilizando datos de escáner láser aeroportado de baja densidad







#### RESUMEN

La estimación de biomasa residual forestal es de gran interés en la evaluación del potencial de las energías renovables o verdes. Este estudio relaciona la biomasa residual forestal de Pino carrasco, estimada en 192 parcelas de campo, con diversas variables independientes obtenidas de datos de escáner láser aeroportado (ALS) en Aragón (España). Con objeto de estimar la biomasa residual forestal se compararon cinco métodos de selección y cuatro modelos de regresión no paramétricos. La muestra fue dividida en entrenamiento y validación, estando compuesta de 144 y 48 parcelas de campo, respectivamente. Los mejores modelos se obtuvieron utilizando el método support vector machine con kernel radial. El modelo incluyó tres métricas ALS: el percentil 70 de los retornos de la nube de puntos, la varianza y el porcentaje de primeros retornos sobre la media. El error cuadrático medio tras la validación fue de 8,85 ton ha<sup>-1</sup>. La influencia de la densidad de puntos, el ángulo de escaneo, la capacidad de penetración del pulso en el dosel vegetal, así como la pendiente y la presencia de arbustos en la precisión del modelo fue evaluada mediante enfoques gráficos y la aplicación de test estadísticos. Densidades de puntos mayores a 1 punto m<sup>-2</sup>, ángulos de escaneo menores a 15°, una penetración del pulso en el dosel vegetal superior a 25% y la presencia de pendientes menores al 30% generó una menor variabilidad en el error medio, a la par que incrementó la precisión del modelo en 0,56, 1,94, 1,44, y 5,47 ton ha<sup>-1</sup>, respectivamente. La presencia de arbustos generó una mayor variabilidad en el error medio pero la pérdida de exactitud en el modelo fue leve (0,10 ton ha<sup>-1</sup>). No se encontraron diferencias estadísticamente significativas entre las categorías establecidas para las variables analizadas, exceptuando para el caso de la penetración del pulso en el dosel vegetal. La cartografía de biomasa residual forestal para masas de Pino carrasco utilizando el mejor modelo, determinó que en las masas analizadas se contabilizan 3.627.021,25 ton, lo que equivale a 1.584,91 miles de toneladas de petróleo (ktoe).





## Quantifying forest residual biomass in *Pinus halepensis* Miller stands using Airborne Laser Scanning data

Darío Domingo <sup>\*a</sup>, Antonio Luis Montealegre <sup>a</sup>, María Teresa Lamelas <sup>a,b</sup>,  
 Alberto García-Martín <sup>a,b</sup>, Juan de la Riva <sup>a</sup>, Francisco Rodríguez <sup>c,d,e</sup>  
 and Rafael Alonso <sup>c,d,e</sup>

<sup>a</sup>GEOFOREST-IUCA, Department of Geography, University of Zaragoza, Zaragoza, Spain;

<sup>b</sup>Centro Universitario de la Defensa de Zaragoza, Academia General Militar, Zaragoza, Spain;

<sup>c</sup>föra forest technologies sll, Campus Duques de Soria s/n, Soria, Spain; <sup>d</sup>EU Ingenierías Agrarias, Universidad de Valladolid, Campus Duques de Soria s/n, Soria, Spain; <sup>e</sup>Sustainable Forest Management Research Institute, University of Valladolid-INIA, Soria, Spain

(Received 13 March 2019; accepted 1 July 2019)

The estimation of forest residual biomass is of interest to assess the availability of green energy resources. This study relates the *Pinus halepensis* Miller forest residual biomass (FRB), estimated in 192 field plots, to several independent variables extracted from Airborne Laser Scanner (ALS) data in Aragón region (Spain). Five selection approaches and four non-parametric regression methods were compared to estimate FRB. The sample was divided into training and validation sets, composed of 144 and 48 plots, respectively. The best-fitted model was obtained using the Support Vector Machine method with the radial kernel. The model included three ALS metrics: the 70th percentile, the variance of the return heights, and the percentage of first returns above mean height. The root-mean-square error (RMSE) after validation was 8.85 tons ha<sup>-1</sup>. The influence of point density, scan angle, canopy pulse penetration, terrain slope, and shrub presence in model performance was assessed using graphical and statistical approaches. Point densities higher than 1 point m<sup>-2</sup>, scan angles lower than 15°, canopy pulse penetration over 25%, and terrain slopes under 30% generated a smaller variability in mean predictive error (MPE) values, thus increasing model accuracy in 0.56, 1.94, 1.44, and 5.47 tons ha<sup>-1</sup>, respectively. Shrub vegetation caused greater variability in MPE values but slightly decreased model accuracy (0.10 tons ha<sup>-1</sup>). No statistically significant differences were found between the categories in the influencing variables, except for canopy pulse penetration. The mapping of *Pinus halepensis* Miller FRB using the best-fitted model summed up a total of 3,627,021.25 tons, which equals to 1,584.91 thousand tonnes of oil (ktoe).

**Keywords:** forest residual biomass; bioenergy; Mediterranean forest; ALS-PNOA; model performance

### 1. Introduction

The use of bioenergy sources and the reduction of CO<sub>2</sub> emissions to the atmosphere is one of the climate and energy targets of the European Energy Roadmap for 2050 (Hamelin et al. 2019). Forest ecosystems provide alternative sources of bioenergy derived from logging residues in timber exploitation. Most countries leave branches and treetops in the forest (Hauglin et al. 2012). These biomass fractions, defined as forest

\*Corresponding author. Email: [ddomingo@unizar.es](mailto:ddomingo@unizar.es)

residual biomass (FRB), can be used as a source of energy in heating systems and in the generation of electricity by replacing fossil fuels in power plants (Richardson et al. 2002). FRB refers to foliage, branches and unmerchantable stem tops generated in timber exploitation or forest management (Velázquez Martí 2006). Harvesting logging residues is already operational in some countries such as Finland (Heinimo 2018). The treatments traditionally applied to these residues in Spain, generated mainly in forest management procedures, include stacking within the forest, controlled burning, and, less commonly, splintering to improve their incorporation into the soil (Velázquez Martí 2006). However, several initiatives such as the Interreg Europe, Wood E3 or BIOmasud support a change in these protocols to harvest these resources as a source of energy.

FRB uses have important environmental and socio-economic benefits. It is a renewable energy and its management can reduce fire risk in forested Mediterranean ecosystems. These environments, characterized by dry summers, a rugged topography, and heterogeneous formations from the structural point of view, are affected by wildfires. FRB is related with canopy bulk density and canopy fuel weight (Andersen, McGaughey, and Reutebuch 2005), two critical fuel metrics whose control through thinning or burning procedures may reduce active crown fire potential (Roccaforte, Fulé, and Covington 2008). Wildfires are a socio-environmental hazard in Mediterranean ecosystems, being *Pinus halepensis* Miller. (hereinafter *P. halepensis*) forests the most affected in Spain with 32,482 ha burned per year between 2001 and 2010 (Anuario de Estadística Forestal, 2013). Furthermore, the exploitation of FRB has socio-economic benefits as it increases rural development (García-Martín et al. 2012), constituting new business possibilities for forestland owners (Hauglin et al. 2012).

The use of remote sensing techniques (Gleason and Im 2011) such as multispectral (García-Martín et al. 2008), radar (Austin, Mackey, and Van Niel 2003), or Airborne Laser Scanning (ALS) sensors have been widely used for estimating biomass. ALS is an active remote sensing system that sends laser beams to the earth. Light Detection and Ranging (LiDAR) sensors capture the radiation scattered by the objects through a photodiode and measure the distance to the object through a telemeter generating a georeferenced three-dimensional point cloud. ALS provides a reliable three-dimensional representation of the earth's surface, being considered the best technology for mapping vegetation structure (Zhao et al. 2018). Several studies have used small-footprint LiDAR systems to estimate biomass using an area-based approach or at a single-tree level (Dalponte et al. 2018; Domingo et al. 2018). The use of LiDAR technology for estimating logging residues has been addressed using a single-tree approach in boreal forests (Hauglin et al. 2013, 2014; Kankare et al. 2013) and in the pine forest of eastern Texas (Popescu 2007). The area-based approach has also been explored in boreal forests (Hauglin et al. 2012; Straub and Koch 2011) and in western China (He et al. 2013). Most of the studies utilized high point density LiDAR data, except for the ones performed by Hauglin et al. (2012), He et al. (2013) and Popescu (2007). However, to the best of our knowledge, neither of these studies has been conducted in Mediterranean ecosystems, which are characterized by a high heterogeneity.

ALS data characteristics, such as flight altitude (Yu et al. 2004), footprint size (Roussel et al. 2017), scan angle (Liu et al. 2018a), point density (Gobakken and Næsset 2008), or pulse frequency (Næsset 2009), as well as environmental conditions such as slope, have been analyzed as sources of error in the estimation of forest attributes. The influence of point density in the estimation of different forest inventory attributes was analyzed in several studies (Roussel et al. 2017; Gobakken and Erik 2008). Fewer studies have considered point density effect on biomass estimation (García et al.

2017; Singh et al. 2015; Ruiz et al. 2014). All of them concluded that a reduction in point density does not introduce significant errors in above ground biomass. Accordingly, a reduction in point density is a viable solution to reduce the cost at regional scales. The effect of the scan angle in the estimation of forestry attributes such as tree height has been also analyzed (Disney et al. 2010; Holmgren, Nilsson, and Olsson 2003; Lovell et al. 2005). Holmgren, Nilsson, and Olsson (2003) studied the effect of the scan angle on canopy closure and Liu et al. (2018a) on gap fraction estimations. Furthermore, Montaghi (2013) analyzed the effect on ALS metrics and found that scan angles higher than 20° have a great effect in forest parameters estimations. Canopy pulse penetration (CPP) varies with canopy structural characteristics and ALS flight configuration (Gaveau and Hill 2003; Hopkinson 2006), affecting Digital Terrain Models (DTM) accuracy (Hollaus et al. 2006; Hyypä et al. 2000; Cowen et al. 2000). Leaf off conditions in deciduous forest improves CPP (Wasser et al. 2013; Hill and Broughton 2009) while lower rates of CPP are found in dense conifer forests (Hollaus et al. 2006) and in the bottom parts of the canopy or lower strata (Wasser et al. 2013; Chasmer, Hopkinson, and Treitz 2006). Furthermore, the effect of slope has been studied when estimating tree height (Breidenbach et al. 2008; Clark, Clark, and Roberts 2004; Ørka et al. 2018), treetop detection (Khosravipour et al. 2015), tree diameter, basal area, number of stems and volume (Ørka et al. 2018). In this sense, although ALS estimations were affected by the increase of slope, the effects were not severe. However, the influence of the sensor scan angle, the terrain slope, and the presence of shrubs species in the understory has been less analyzed when estimating biomass using ALS data.

Accordingly, the main goal of this study is to estimate the FRB of *P. halepensis* forests at a regional scale in the Aragón region (northeast Spain) using ALS point clouds and field data. Specifically, we aim to (1) assess the relationship between FRB and ALS-derived independent variables using an area-based approach; (2) compare five variable selection methods (Spearman's rank correlation, Stepwise selection, Principal component analysis, Least absolute shrinkage and selection operator, and All subset selection) and five regression methods (Support Vector Machine with radial and linear kernels, Random Forest, locally weighted linear regression, and regression tree based on a minimum length principle); (3) analyze the influence of point density, scan angle, canopy pulse penetration, terrain slope, and shrub presence in model performance; (4) map and quantify FRB using the most accurate model.

## 2. Materials and methods

Figure 1 describes the methodological process followed for field data and ALS processing, FRB estimation and the analysis of the influence of ALS and environmental characteristics in model accuracy.

### 2.1. Study area

The *P. halepensis* forests under study are located in the Aragón region (Figure 2) in Spain. This area is characterized by a hilly topography, with elevations ranging from 300 to 1150 m above sea level and slopes from up to 39°. These forests represent the 18.7% of the forested area in Aragón region, including semi-natural and reforested stands. The lithology of the study area varies from Miocene carbonate and marl sediments to Mesozoic and Eocene limestone platforms. The climate of the region is the Mediterranean



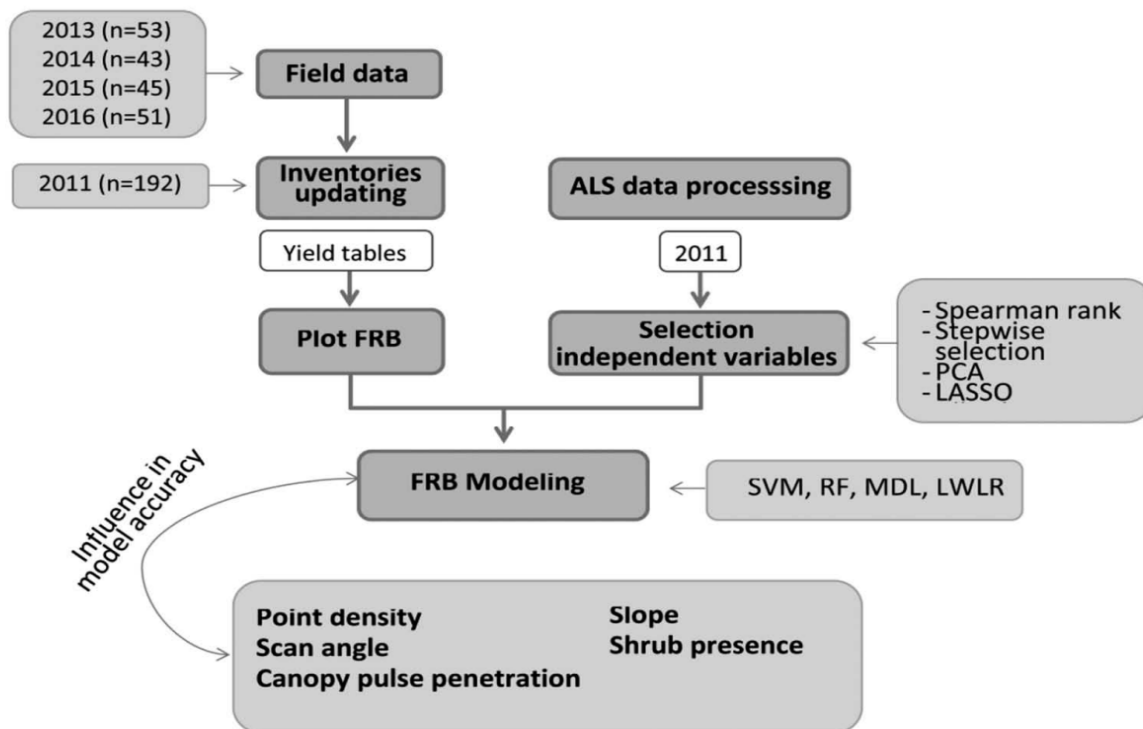


Figure 1. Methodological steps.

with continental features. Precipitation is irregular and mostly occurs during the spring and autumn. The average annual precipitation ranges from areas with less than 350 mm to areas that may reach 1,000 mm. The average temperature varies from 5°C in the winter to 22°C in the summer.

Some of the pine stands were afforested 60 to 40 years ago and usually present a lower presence of understory, less structural complexity and poor biodiversity (Granados et al. 2016) compared with the natural stands. The understory includes Mediterranean evergreen species, such as *Quercus ilex subsp. ballota*, *Quercus coccifera*, *Juniperus oxycedrus*, *Buxus sempervirens*, *Juniperus phoenicea*, *Rosmarinus officinalis* and *Thymus vulgaris*.

## 2.2. ALS data and processing

The ALS data were acquired by the PNOA project between January and February 2011 using a Leica ALS60 discrete-return sensor operating at a wavelength of 1,064 nm. Average flying altitude was 3,000 m above sea level, with a maximum scan angle up to 29° from nadir. Point clouds with up to four returns per pulse were provided in  $2 \times 2$  km tiles in LAS format and ETRS 1989 UTM coordinates.

The first processing step was the removal of noise points and the overlapping class, in the case of the existence of vertical (z) and horizontal (x, y) displacements between lines of flights. In this sense, overlapping returns were removed from 1637 tiles. Then, 3,800 tiles that cover the study area were filtered using the multiscale curvature classification algorithm (Evans and Hudak 2007). This filter is suitable for this environment according to Montealegre, Lamelas, and de la Riva. (2015a), being implemented in

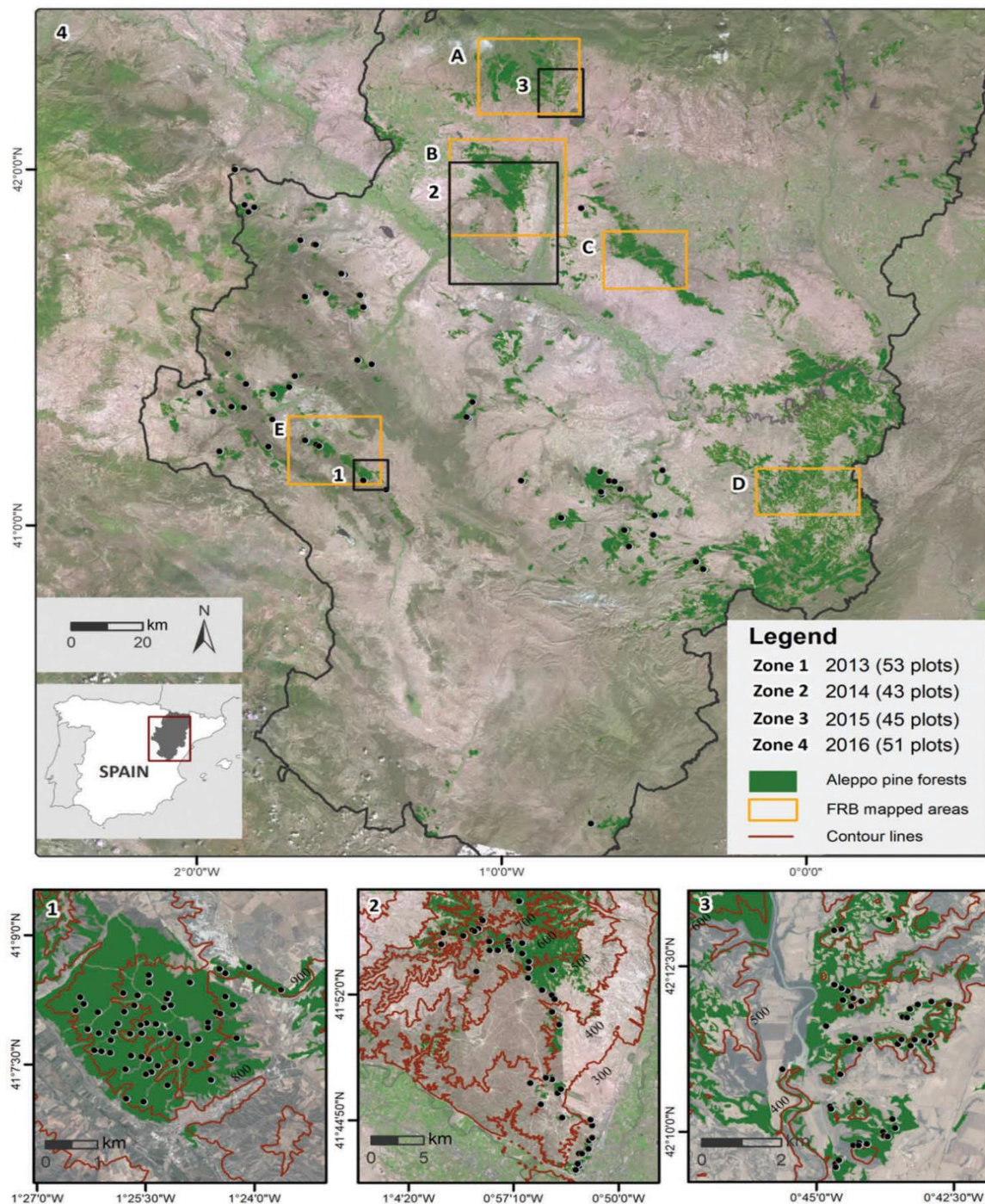


Figure 2. Study area and location of 192 forest inventory plots, including two images of representative plots, and location of FRB mapping areas in Figure 5 (A to E). High spatial resolution orthophotography from Spanish national plan for aerial orthophotography (PNOA) Spanish spatial data infrastructure (IDEE) service is used as a backdrop. Coordinate system: ETRS89 UTM Zone 30 N.



Multiscale Curvature Classification (MCC) 2.1 command-line tool. Ground points were interpolated with the Point-TIN-Raster method, implemented in ArcGIS 10.5 software to obtain a digital elevation model (DEM) with 1 m resolution (Montealegre, Lamelas, and de la Riva. 2015b).

Point clouds were clipped with the geometry of each field plot and the return heights were normalized with the DEM using FUSION LDV 3.60 open source software (McGaughey 2009). Furthermore, FUSION was used to generate a full suite of statistical metrics related to height distribution and canopy cover. Canopy height variables include minimum, maximum, mean, mode (*Elev.min*, *Elev.max*, *Elev.mean*, *Elev.mode*), percentiles at different intervals (*P01*, *P05*, *P10*, *P20*, *P30*, *P40*, *P50*, *P60*, *P70*, *P75*, *P80*, *P90*, *P95*, *P99*), elevation quadratic mean (*Elev. SQRT mean SQ*), Elevation cubic mean (*Elev. CUR mean CUBE*) and L moments (*Elev. L1*, *Elev. L2*, *Elev. L3*). Canopy height variability variables include standard deviation (*Elev. SD*), variance (*Elev.variance*), coefficient of variation (*Elev.CV*), interquartile distance (*Elev.IQ*), skewness (*Elev.skewness*) and kurtosis (*Elev.kurtosis*). Following Næsset and Økland (2002,) a threshold value of 2 m height was applied to remove ground and understory returns before generating the ALS-derived variables. Canopy density metrics include canopy relief ratio (*CRR*), percentage of first or all returns above a threshold, the mean or the mode (e.g.: % *first ret. Above 2.00*) as well as the ratio of all returns respect to the number of total returns (e.g.: (*All ret. Above 2.00*)/(total *first ret.*) by 100).

### 2.3. Field plot data and FRB calculation

Field data were acquired in 192 plots from four campaigns performed during 2013, 2014, 2015, and 2016 (Figure 2), hereinafter first, second, third, and fourth campaign, respectively. The sampling considered the variability of *P. halepensis* forests in terms of terrain slope, canopy height, and canopy cover.

Field data from the first campaign were acquired from June to July 2013 in 53 circular plots with 15 m radius. The center of each plot was positioned using a Leica VIVA® GS15 CS10 real-time kinematic Global Navigation Satellite System (GNSS) with a planimetric accuracy of 0.30 m. A diameter tape, with millimeter precision, was used for measuring tree diameter at breast height (dbh) for those trees with a dbh larger than 7.5 cm, which is the standard dbh for inventoried trees in Spain. A Suunto® hypsometer, with centimeter precision, was used for measuring green crown height and tree height of up to 4 randomly selected trees within each plot. The selection of the sample trees, from 7.5 cm up to 42.5 cm, was performed following the diametric classes defined as representative for the study area in the third national forest inventory. The height for those trees not measured in the field plots was predicted by using a height-diameter model developed from the sampled trees.

Field data from the second campaign were acquired from July to September 2014 in 43 circular plots and data from the third field campaign were gathered from June to July 2015 in 45 plots. In both cases, the 30 m diameter plots were positioned using the same GNSS instrument as in the first campaign obtaining a planimetric accuracy of 0.15 m in 2014 and 0.18 m in 2015. A Haglöf Sweden® Mantax Precision Blue diameter caliper, with a millimetric precision, was used for measuring the dbh of those trees with a dbh larger than 7.5 cm. A Haglöf Sweden® Vertex instrument, with a centimetric precision, was used to measure the green crown height and the height for all trees in the plot.

The 51 field plots from the fourth campaign were collected in April 2016. A Trimble submetric GNSS was used to position the center of each plot with a submetric accuracy

in planimetry. In this case, the radius was variable (5.6 m, 8.5 m, 11.3 m, and 14.10 m), depending on stand density, in order to obtain data from a similar number of trees in each plot. A Hagl f Sweden® Mantax Precision Blue diameter caliper was used for measuring those trees with a dbh larger than 7.5 cm. A Hagl f Sweden® Vertex was used for measuring the green crown height and the height of up to six trees, the nearest to the plot center. The sample was completed to achieve 100 dominant stems  $\text{ha}^{-1}$  considering those with larger dbh. The height for those trees not measured in the field plots was predicted by using a height-diameter model developed from the sampled trees.

Given the difference in time between the point cloud acquisition, in 2011, and the field campaigns, the field measurements were updated using the Spanish National Forest Inventory (NFI). The yield tables of dbh and height growth values for *P. halepensis* between the second NFI (NFI2) and the third one (NFI3) (11 years) were used. A linear interpolation based on these tables was made to obtain a subtractive value of dbh and height per each diametric class proposed by the NFI (Table 1).

FRB was calculated for each inventoried tree using the *P. halepensis* allometric equations proposed by Ruiz-Peinado, Del Rio, and Montero (2011). These authors propose different equations for the thick branch fraction, with a diameter larger than 7 cm, the medium branch fraction, with diameter between 2 and 7 cm, and the thin branch fraction, with a diameter smaller than 2 cm, including needles. The sum of the biomass in the different fractions gave the biomass at tree level. Then, FRB of each tree was summed up to obtain the FRB at plot level expressed in tons per hectare.

#### 2.4. FRB modeling

In order to select the most suitable independent variables for FRB modeling, five different approaches, with a total of 10 variants, described in Domingo et al. (2018) were applied (Figure 1): (i) Spearman's rank correlation coefficient determine the strength and direction of the relationship between field plot forest residual biomass and ALS metrics. A minimum positive and negative rho value of 0.5 and a subsequent manual variable selection were performed in R; (ii) Stepwise selection, which considers dropping or adding variables at several steps, was performed using backward, forward and bidirectional approaches; (iii) Principal component analysis (PCA) is a technique used to reduce the number of variables in multiple regression analysis. In this case, the PCAs with eigenvalues greater than 0.1 were retained, considering Kaiser Criterion, and a Varimax rotation was applied to maximize the sum of the variances and to better interpret the results; (iv) Least absolute shrinkage and selection operator (LASSO) allows to generate interpretable models by minimizing the residual sum of squares. This technique was computed in R with the "glmnet" package; and (v) All subset selection provides a suitable group of metrics on which the models should focus their attention. Previous selection of the number of metrics is required. In this study, exhaustive, backward, forward, and sequential replacement approaches were used.

The selection of ALS metrics was restricted to a combination of up to four independent variables, with the aim of generating more parsimonious models for forest management and reducing variable multicollinearity.

Five different regression methods were compared to estimate FRB. The multivariate linear regression model (MLR) was initially tested, but the normality of the residuals, homoscedasticity and independence, and no auto-correlation of the residuals were not fulfilled even transforming the dependent and independent variables. Consequently, this

Table 1. Values of tree diameter at breast height (dbh) and height growth between the NF12 and the NF13, and subtractive values of dbh and height when applying linear interpolated degrowth.

| Diameteric class (cm) | dbh growth (mm) | Height growth (mm) | Field campaign |            |                   |          |            |                   |          |            |                   |          |            |                   |
|-----------------------|-----------------|--------------------|----------------|------------|-------------------|----------|------------|-------------------|----------|------------|-------------------|----------|------------|-------------------|
|                       |                 |                    | 2013           |            |                   | 2014     |            |                   | 2015     |            |                   | 2016     |            |                   |
|                       |                 |                    | dbh (mm)       | Height (m) | Height growth (m) | dbh (mm) | Height (m) | Height growth (m) | dbh (mm) | Height (m) | Height growth (m) | dbh (mm) | Height (m) | Height growth (m) |
| <10                   | 24              | 1.20               | -4.36          | -0.22      |                   | -6.55    | -0.33      |                   | -8.73    | -0.44      |                   | -10.91   | -0.55      |                   |
| 10-15                 | 29              | 1.40               | -5.27          | -0.25      |                   | -7.91    | -0.38      |                   | -10.55   | -0.51      |                   | -13.18   | -0.64      |                   |
| 15-20                 | 33              | 1.50               | -6.00          | -0.27      |                   | -9.00    | -0.41      |                   | -12.00   | -0.55      |                   | -15.00   | -0.68      |                   |
| 20-25                 | 30              | 1.40               | -5.45          | -0.25      |                   | -8.18    | -0.39      |                   | -10.91   | -0.51      |                   | -13.64   | -0.64      |                   |
| 25-30                 | 30              | 1.40               | -5.45          | -0.25      |                   | -8.18    | -0.39      |                   | -10.91   | -0.51      |                   | -13.64   | -0.64      |                   |
| 30-35                 | 33              | 1.10               | -6.00          | -0.20      |                   | -9.00    | -0.30      |                   | -12.00   | -0.40      |                   | -15.00   | -0.50      |                   |
| 35-40                 | 32              | 1.50               | -5.82          | -0.27      |                   | -8.73    | -0.41      |                   | -11.64   | -0.55      |                   | -14.55   | -0.68      |                   |
| 40-45                 | 27              | 1.60               | -4.91          | -0.29      |                   | -7.36    | -0.44      |                   | -9.82    | -0.58      |                   | -12.27   | -0.73      |                   |
| 45-50                 | 24              | 1.90               | -4.36          | -0.35      |                   | -6.55    | -0.52      |                   | -8.73    | -0.69      |                   | -10.91   | -0.86      |                   |
| 50-55                 | 58              | 1.00               | -10.55         | -0.18      |                   | -15.82   | -0.28      |                   | -21.09   | -0.36      |                   | -26.36   | -0.45      |                   |
| 55-60                 | 10              | 0.50               | -1.82          | -0.09      |                   | -2.73    | -0.14      |                   | -3.64    | -0.18      |                   | -4.55    | -0.23      |                   |



parametric method was not considered for further analysis. Two variants of the non-parametric supervised learning Support Vector Machine (SVM) model were computed by applying linear and radial kernels. Following previous studies (Domingo et al. 2018), a cost parameter within the interval 1–1,000 and a gamma parameter within the interval 0.01–1 were tested. Random Forest (RF) method was implemented using the R packages “randomForest” and “caret,” including “corr.bias” parameter. The number of trees to growth was fitted between 1 and 3,000 and the number of variables to divide the nodes between 1 and 3. Finally, two non-parametric regression tree structures, based on “If Then” rules, were computed: locally weighted linear regression (LWLR) and linear model with a minimum length principle (MDL).

Model validation was performed by splitting the original sample into a training set of 75% of the plots (144 cases) and a test set of 25% of the plots (48 cases). The robustness of random models was assessed executing 100 times this process (García-Gutiérrez et al. 2015). For each computed model, the root-mean-square error (RMSE) (equation 1), relative RMSE respect to the mean (%RMSE) (equation 2), and bias were calculated. Friedman non-parametric test was used to determine whether there were differences between the computed models, considering the RMSE values (Stojanova et al. 2010). Post-hoc Nemenyi test was applied in order to determine whether differences were statistically significant at the level of 0.05 (Nemenyi 1963). This test was applied only when the null-hypothesis of the Friedman test was rejected, as implies non-equivalence between models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

$$\%RMSE = \frac{RMSE}{\bar{y}} \times 100 \quad (2)$$

where  $y_i$  is the observed FRB value for plot  $i$ ;  $\hat{y}_i$  is the predicted FRB for sample plot  $i$ ;  $n$  is the number of plots and  $\bar{y}$  is mean FRB for all plots.

### 2.5. Influence of ALS characteristics and environmental conditions in model accuracy

The effect of three ALS characteristics (point density, scan angle, and canopy pulse penetration) and two environmental variables (terrain slope and presence of shrubs) in model performance were tested. In this sense, the mean predicted errors (MPE) obtained for each field plot were grouped according to the established categories for each of the above-mentioned variables as described below (Table 2).

$$MPE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (3)$$

$$\%MPE = \frac{MPE}{\bar{y}} \times 100 \quad (4)$$

where  $y_i$  is the observed FRB value for plot  $i$ ;  $\hat{y}_i$  is the predicted FRB for sample plot  $i$ ;  $n$  is the number of plots and  $\bar{y}$  is mean FRB for all plots.

Table 2. Number of plots per category established for the ALS characteristics and environmental conditions.

| ALS characteristics                          | N° of plots | Environmental conditions | N° of plots |
|--|-------------|--------------------------|-------------|
| <i>Point density (points m<sup>-2</sup>)</i> |             | <i>Slope (%)</i>         |             |
| < 1  | 150         | < 15                     | 65          |
| > 1  | 42          | > 15                     | 127         |
| <i>Scan Angle (°)</i>                        |             | <i>Shrub presence</i>    |             |
| < 5  | 108         | Yes                      | 86          |
| 5–15   | 73          | No                       | 106         |
| > 15   | 11          |                          |             |
| <i>Canopy pulse penetration (%)</i>          |             |                          |             |
| < 25   | 77          |                          |             |
| 25–50  | 90          |                          |             |
| 50–75  | 21          |                          |             |
| > 75   | 4           |                          |             |

Following Montealegre, Lamelas, and de la Riva. (2015b), plots were categorized into two classes of point density: up to 1 point m<sup>-2</sup> and higher than 1 point m<sup>-2</sup>. This threshold allows to have relatively high accuracies when modeling forest parameters according to Jakubowski, Guo, and Kelly (2013) and García et al. (2017). Three categories of scan angle were determined: plots with an average scan angle of up to 5°, plots with an average scan angle between 5° and 15° and plots with an average scan angle higher than 15°. The first class was set close to the nadir. The breakpoint between second and third classes was defined by the maximum average scan angle established by the PNOA mission in order to reach the minimum nominal point density specified by the project. Following Montealegre, Lamelas, and de la Riva. (2015b), the influence of canopy pulse penetration was analyzed determining four categories: 0%-25%, 25%-50%, 50%-75%, and 75%-100%. The proportion of pulses that penetrate the canopy and reach the ground was calculated using a ground tolerance of 2 m. The effect of the terrain slope was assessed using two categories: plots with smooth slopes of up to 15%, and steep slopes higher than 15%. Finally, shrub presence was determined as plots with and without understory.

A graphical assessment using boxplots, including the average mean MPE per classes, was applied. In addition, normality and homogeneity tests, considering logarithmic and square root transformation, were analyzed. For those variables with non-normal distributions and two categories, Mann–Whitney and median tests were applied. Kruskal Wallis test was applied for those with more than two classes.

### 3. Results

#### 3.1. Field plot computation

In the case of the first campaign, equation 5 is applied for the estimation of tree height for those trees not measured in the fieldwork. Model performance of the ht model reported an RMSE of 1.36 m and R<sup>2</sup> of 0.63. Normality of the residuals, homoscedasticity, and independence or no autocorrelation in the residuals were verified for the fitted model.

$$ht = 0.776 \cdot G^{0.179} \cdot dbh_i^{0.660} \cdot 1.009 \quad (5)$$

where  $ht$  is tree height (m),  $dbh_i$  is the diameter at breast height (cm) and  $G$  is field plot basal area ( $\text{m}^2 \text{ ha}^{-1}$ ).

Equation 6 shows the model used for the estimation of tree height for those trees not measured in the fieldwork in the fourth campaign. Model performance of the  $ht$  model reported an RMSE of 0.80 m and  $R^2$  of 0.93.

$$ht = \left( 1.3^{2.5511} + (H_o^{2.5511} - 1.3^{2.5511}) \cdot \frac{1 - \exp(-0.025687 \cdot dbh)}{1 - \exp(-0.025687 \cdot D_o)} \right)^{1/12.5511} \quad (6)$$

where  $ht$  is tree height (m),  $dbh$  is the diameter at breast height (cm),  $H_o$  is the Assmann dominant height (m) and  $D_o$  is the Assman dominant diameter (cm).

Table 3 shows a summary of the field plot characteristics. Tree height of the inventoried plots range from 3.66 to 19.10 m with an average value of 10.06 m, and  $dbh$  values range from 8.64 up to 43.22 cm. The average FRB is 33.55 tons  $\text{ha}^{-1}$ , but the variability within plots is notorious with a range value of 134.23. In addition, the high range and the standard deviation values confirm the variability of the sample that characterizes *P. halepensis* forest in the Aragón region.

### 3.2. Variable selection and FRB modeling

Two selection methods determined the most suitable ALS metrics for the analyzed regression models (Table 4). Spearman rank correlation with subsequent manual variable selection was the most powerful method when computing SVM with radial kernel, RF and LWLR, while all subset selection exhaustive determined the best models when using SVM with linear kernel and MDL regression methods. As mentioned before, the maximum number of selected variables was set to four. In this sense, the generated models range from two variables, when computing RF, to four variables, when using SVM with linear kernel or MDL. The selection methods frequently include one or two metrics associated with canopy height, one metric related to canopy height variability and another one related to canopy density. Regarding canopy height metrics, higher percentiles were frequently selected as *P70* of height and *Elev. L2*. Variance was the most selected canopy height variability metric. In addition, the percentage of first returns above mean height, and the percentage of returns above mean height over the total first returns showed the strongest correlation with FRB.

The best regression methods to estimate FRB are summarized in Table 4. SVM with radial kernel presents the lowest RMSE, followed by MDL method. In comparison with those methods, SVM with linear kernel, RF, and LWLR shows lower accuracies. Most of the models show values of bias close to zero, but RF shows a slight overestimation with

Table 3. Summary of the field plot characteristics for 2011.

|                              | Min. | Max.   | Range  | Mean  | Standard deviation |
|------------------------------|------|--------|--------|-------|--------------------|
| Terrain slope (degrees)      | 1.10 | 64.65  | 63.56  | 23.65 | 16.98              |
| Tree height (m)              | 3.66 | 19.10  | 15.44  | 10.06 | 3.14               |
| Tree $dbh$ (cm)              | 8.64 | 43.22  | 34.58  | 20.74 | 7.89               |
| FRB (tons $\text{ha}^{-1}$ ) | 1.23 | 135.46 | 134.23 | 33.55 | 20.89              |

Table 4. Summary of the models and validation results in terms of RMSE (tons ha<sup>-1</sup>), % RMSE respect to the mean and bias (tons ha<sup>-1</sup>). SVM r. refers to Support Vector Machine with radial kernel; SVM l. refers to Support Vector Machine with linear kernel; RF refers to random forest; MDL refers to linear model with a minimum length principle; LWLR refers to locally weighted linear regression; SM is selection method; Rho is Spearman rank correlation; ASSe is All subset selection exhaustive; Elev. is elevation; ret. is returns; P is percentile.

| ALS metrics   | Model  | Fitting phase |      |       |       |       | Validation |       |      |   |                |
|---|--------|---------------|------|-------|-------|-------|------------|-------|------|---|----------------|
|   |        | SM            | RMSE | RMSE  | Bias  | %     | RMSE       | RMSE  | Bias | % | R <sup>2</sup> |
| $P70 + Elev. Variance + \% first\ ret. above\ mean$   | SVM r. | Rho           | 6.89 | 20.55 | 0.15  | 8.85  | 26.38      | 0.26  | 0.82 |   |                |
| $Elev. L2 + Elev. CV + Elev. CUR\ mean\ CUBE + (All\ ret. Above\ mean)/$<br>$(total\ first\ ret) \cdot 100$ | MDL    | ASSe          | 8.23 | 24.51 | 0.17  | 9.47  | 28.22      | -0.15 | 0.80 |   |                |
| $Elev. L2 + Elev. Variance + Elev. CUR\ mean\ CUBE + (All\ ret. Above\ mean)/(total\ first\ ret) \cdot 100$ | SVM l. | ASSe          | 9.45 | 28.15 | 0.33  | 9.65  | 28.76      | 0.19  | 0.78 |   |                |
| $Elev. SORT\ mean\ SQ + (All\ ret. above\ 2.00)/(total\ first\ ret) \cdot 100$                              | RF     | Rho           | 5.13 | 15.28 | -1.10 | 9.88  | 29.45      | -1.16 | 0.79 |   |                |
| $P70 + Elev. Variance + \% first\ ret. above\ mean$   | LWLR   | Rho           | 6.38 | 19.01 | -0.06 | 10.21 | 30.42      | 0.01  | 0.77 |   |                |

values close to 1. The selected SVM with radial kernel includes three ALS metrics: the 70<sup>th</sup> percentile of height, the variance of the height of the returns and the percentage of the first returns above mean height. The validation shows an RMSE of 8.85 tons ha<sup>-1</sup> with a % RMSE of 26.38, and a relative R<sup>2</sup> of 0.82.

The scatter plots of FRB observed and predicted values for the different regression methods are presented in Figure 3. According to Friedman test, the models are not equivalent to a p-value of 0.00. In addition, the post-hoc Nemenyi test indicates that no statistically significant differences exist between the RMSE values obtained from the different regression models, with 95% of probability.

### 3.3. Assessment of the influence of ALS characteristics and environmental conditions in model accuracy

The influence of ALS and environmental variables in model performance is graphically summarized in Table 5 and Figure 4. All the analyzed variables show a coherent relationship with MPE values. It can be observed that the increase in point density reduces MPE, ranging from an average of 5.21 tons ha<sup>-1</sup>, for those plots with a point density lower than 1 point m<sup>-2</sup>, to 4.65 tons ha<sup>-1</sup>, for those ones with a density higher

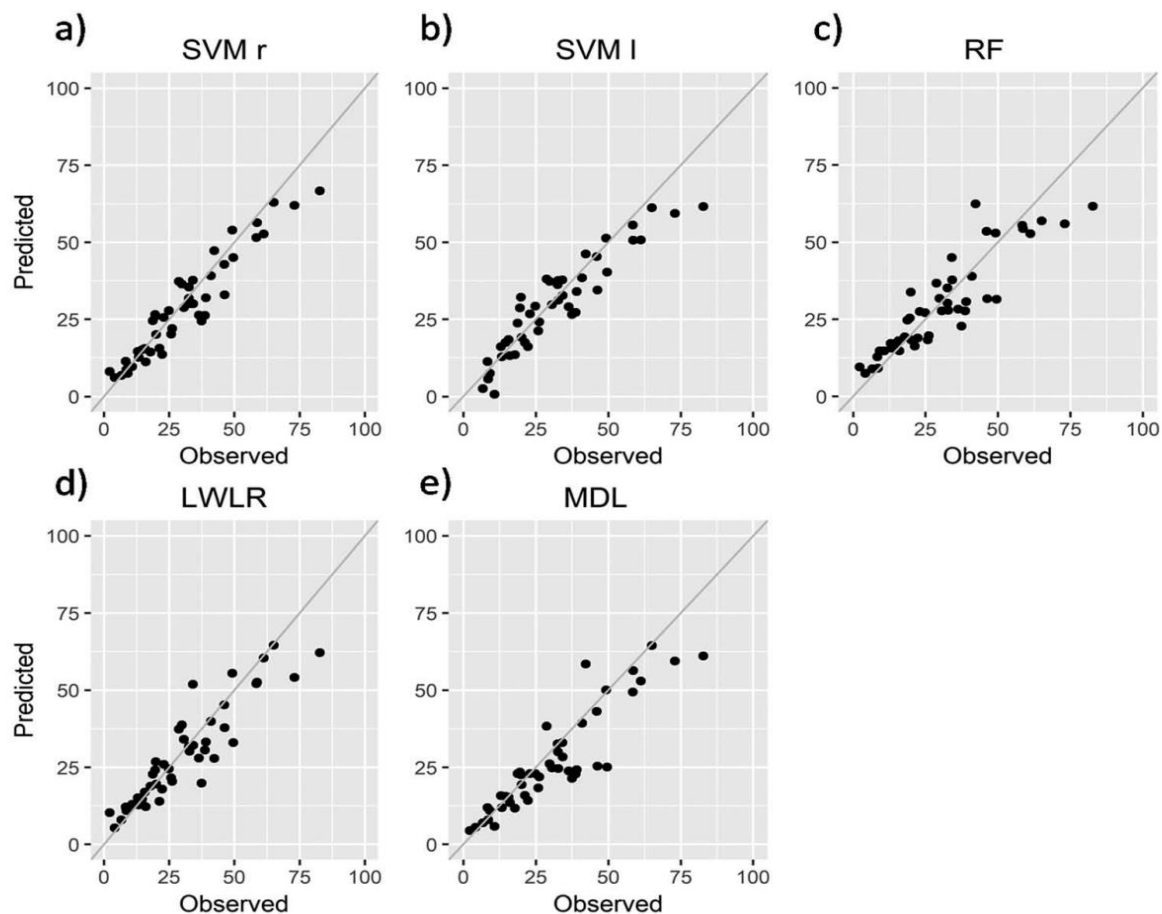


Figure 3. Scatterplot of predicted vs. observed values of FRB (tons ha<sup>-1</sup>) using four different regression methods.



Table 5. Summary of MPE values per category established for the ALS characteristics and environmental conditions.

|                                   | Classes  | MPE    |      |                    |
|-----------------------------------|----------|--------|------|--------------------|
|                                   |          | Median | Mean | Standard deviation |
| Density (points m <sup>-2</sup> ) | <1       | 3.11   | 5.21 | 5.95               |
|                                   | >1       | 3.41   | 4.65 | 4.22               |
| Scan Angle (°)                    | <5       | 2.97   | 4.58 | 4.51               |
|                                   | 5–15     | 3.19   | 5.62 | 6.99               |
|                                   | >15      | 6.18   | 6.52 | 5.05               |
| CPP (%)                           | <25      | 5.04   | 6.69 | 7.43               |
|                                   | 25–50    | 3.21   | 4.40 | 3.44               |
|                                   | 50–75    | 2.06   | 2.93 | 4.32               |
|                                   | >75      | 1.24   | 1.22 | 0.95               |
| Slope (%)                         | <15      | 2.95   | 4.14 | 3.71               |
|                                   | >15      | 3.35   | 5.58 | 6.33               |
| Shrub presence                    | Absence  | 3.45   | 5.04 | 4.40               |
|                                   | Presence | 3.02   | 5.15 | 6.85               |

than 1 point m<sup>-2</sup> (Table 2). Similarly, the increment of scan angle produces higher errors when modeling, with values of 4.58 tons ha<sup>-1</sup>, for those plots close to nadir, and values of up to 6.52 tons ha<sup>-1</sup>, for those ones with a scan angle higher than 15°. The decrease in CPP generates an increase in the MPE values of the models, ranging from 1.22 tons ha<sup>-1</sup> in plots with CPP higher than 75% to 6.69 tons ha<sup>-1</sup> in plots with CPP lower than 25%. The plots with slopes higher than 15% show higher MPE, with an average value of 5.58 tons ha<sup>-1</sup>. Finally, the presence of shrub slightly decreases model accuracy when modeling FRB, ranging from 5.15 tons ha<sup>-1</sup>, in plots without understory, up to 5.40 tons ha<sup>-1</sup>, for those with shrub presence. A general observed pattern is a greater dispersion in MPE values, decreasing model performance, with lower point densities, higher scan angles, lower CPP, steep slopes and shrub presence. However, plots with scan angle higher than 15° do not follow this pattern. This fact may be explained by the relative low amount of plots that reach these conditions.

The assessment of differences between classes, performed using Mann-Whitney and median tests, shows no statistically significant differences between classes, with a 95% of probability, in the case of point density, terrain slope, and shrub presence. Similar results present the analysis of the scan angle using the Kruskal Wallis test. However, statistically significant differences were found between CPP classes when applying ANOVA test with a p value of 0.01.

### 3.4. Mapping and quantification of FRB

The SVM r. model, which includes the *P70*, *Elev. variance* and the *% first ret. above mean*, was selected to estimate the FRB in the study area. This model was parametrized with a gamma value of 0.21 and a cost value of 10. Figure 5 shows the FRB mapping for the complete study area and five representative areas of the *P. halepensis* forest in this region: one in the foothills of the Pyrenees (Figure 5 A), two more located in the central Ebro valley (Figure 5 B,C) and two in the Iberian System (Figure 5 D,E). Selected areas present values of FRB ranging from less than 10 tons ha<sup>-1</sup> up to more than 100 tons ha<sup>-1</sup> in some specific patches, confirming the heterogeneity of the Mediterranean *P. halepensis* forest. Generally,

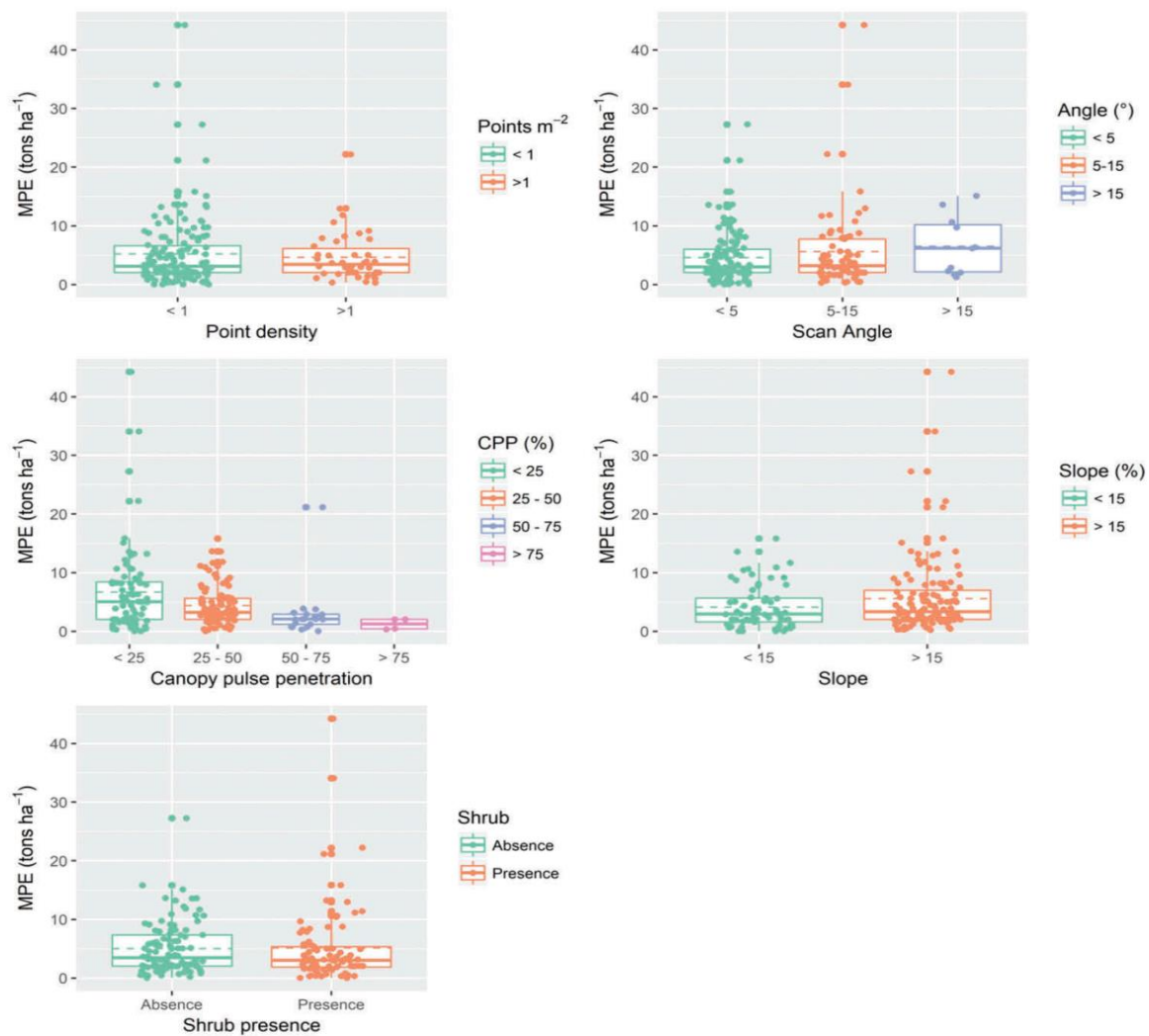


Figure 4. Boxplots of point density, scan angle, CPP, slope, and shrub presence classes related to MPE values. Solid lines depict the median MPE per class and dashed lines depict the mean MPE per class.

lower values of FRB tend to concentrate in the forested areas of the Ebro Valley, characterized by a dry climate and poor soils. In contrast, higher values of FRB are normally associated with mature forested areas with better climatic conditions, which are close to mountain ranges. Aspect also plays an important role at local scales increasing FRB values in those oriented to the north, with better climatic conditions and soils. The area mapped considers a polygon mask that covers the forested *P. halepensis* areas from the Forest Map of Spain (FMS) in scale 1:50,000. The isolated patches of up to 20 ha were excluded from mapping. In addition, the model did not consider those areas mapped as other land covers, e.g.: croplands, denoting good model performance, in the FMS. Furthermore, the borders of some small FMS polygons have not been included in the FRB map as pixel size was set to 25 m. Accordingly, 197,951.24 ha were mapped, representing 87.66% of the *P. halepensis* forested area estimated for the Aragón region by Cabanillas (2010). FRB summed up a total of 3,627,021.25 tons of FRB, which equals to 1,584.91 thousand tonnes of oil (ktoe).



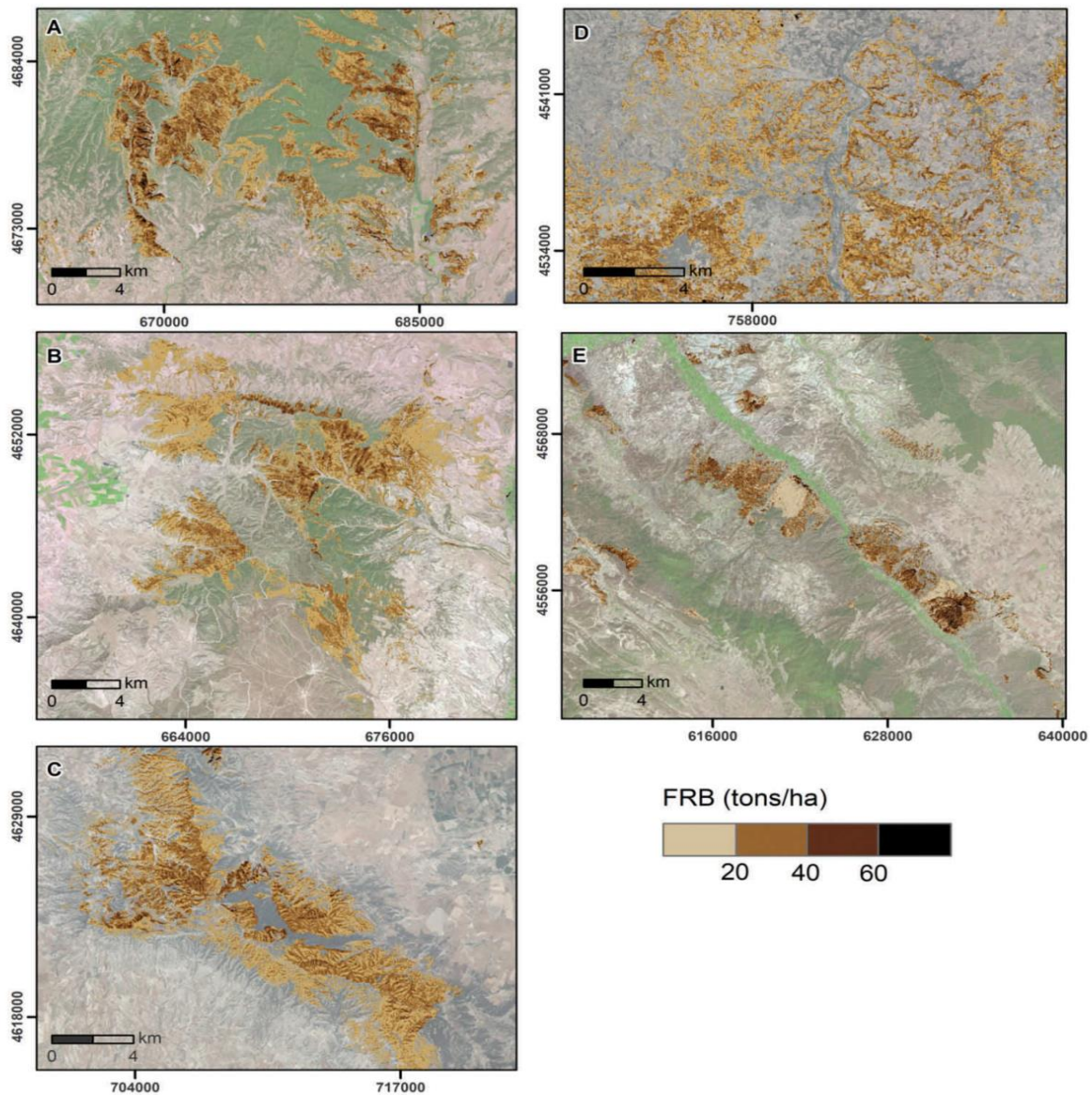


Figure 5. *P. halepensis* FRB in Aragón region. Several subsets of *P. halepensis* forest are shown to better visualize FRB results. Subsets location is included in Figure 2. High spatial resolution orthophotography from Spanish national plan for aerial orthophotography IDEE service is used as backdrop.

## 4. Discussion

### 4.1. Variable selection and FRB modeling

Results reveal that low-density ALS data can be used to accurately estimate FRB at a regional scale in Mediterranean environments characterized by a high heterogeneity. The use of an appropriate selection method reduces variable collinearity, increasing model understanding. In this sense, Kristensen et al. (2015) proposed Spearman rank as a good tool for determining the relationships between ALS and field metrics. In accordance with these authors, this method, with subsequent manual selection,



determined the best independent variables when computing SVM with radial kernel, RF and LWLR. All subset selection exhaustively determined the best models when using SVM with linear kernel and MDL regression methods. These results agree with Hansen et al. (2015) who also used similar best subset regression procedures to estimate biomass with ALS data. Furthermore, our results agree with García-Gutiérrez et al. (2015), who found that Stepwise was the powerless technique. In agreement with Næsset (2002) and Cao et al. (2016), PCA may be considered as a first attempt to reduce collinearity, but an additional selection approach should be considered before modeling.

The comparison between non-parametric regression methods shows that SVM with radial kernel outperforms the analyzed approaches, presenting the lowest RMSE (8.85 tons ha<sup>-1</sup>) and bias (0.26) and matching with the values obtained by Jakubowski, Guo, and Kelly (2013) and García-Gutiérrez et al. (2015). The better performance of SVM with radial kernel respect to Random forest and LWLR may be caused by overfitting of RF and LWLR in the training phase. In the case of SVM with linear kernel, the decrease in performance might be influenced by the dispersion in the distribution of values. Moreover, no significant differences were found between the models, agreeing with García-Gutiérrez et al. (2015) and Lee et al. (2018). Comparison between regression methods should be considered, especially when working with big datasets in heterogeneous forest stands as the Mediterranean ones. The generated models, specifically the regression model and the type of ALS selected metrics, might be applied when working at larger areas in Mediterranean Pine forests. However, the generalization of the generated models for other forest types (e.g.: deciduous) or coniferous forest in other environments may require further testing. The use of SVM when working with large field datasets is recommended. In this sense, the application of the proposed selection methods to identify the most suitable ALS metrics is advisable.

#### **4.2. Influence of ALS characteristics and environmental conditions in model accuracy**

A decrease in point density increases the MPE up to 0.56 tons ha<sup>-1</sup> in agreement with García et al. (2017), Singh et al. (2015) and Ruiz et al. (2014). Furthermore, the accuracy remained relatively high when using low-point densities. No statistically significant differences were found between groups (Jakubowski, Guo, and Kelly 2013). Our results confirm the conclusions from previous studies based on low-density ALS-PNOA data (Domingo et al. 2018). This public information is an accurate alternative to estimate forestry variables when higher point density data are not available. An increment in scan angle, from angles close to nadir (<5°) to angles higher than 15°, increases 1.94 tons ha<sup>-1</sup> models MPE. These results agree with Disney et al. (2010) who proposed to minimize the use of data collected at scan angles greater than ~15°. Liu et al. (2018) also proposed to avoid off-nadir angles from 23° to 38°. However, no statistically significant differences were found between those ranges agreeing with Montaghi (2013), who found that the estimation of forestry metrics using an area-based approach were relatively unaffected by scan angles up to 20°. In agreement with Hollaus et al. (2006) a CPP decrease generates an increase in the MPE of the models of up to 5.47 tons ha<sup>-1</sup>, when comparing plots with a penetration higher than 75% to plots with penetration lower than 25%. Secondary effects of densely covered areas are a decrease in CPP in lower strata (Chasmer, Hopkinson, and Treitz 2006; Wasser et al. 2013) as well as a decrease in DTM accuracy (Cowen et al. 2000; Hollaus et al. 2006; Hyypä et al. 2000). Steep slopes reduce model accuracy, but not significant changes were found, which agrees with Breidenbach et al. (2008), Clark, Clark, and Roberts (2004), and Ørka et al. (2018).

Slope effect might be partially explained by the lower accuracy of DTMs in these areas, considering that filters have more difficulties on determining ground points on steep slopes (Montealegre, Lamelas, and de la Riva. 2015a). Consequently, the filtering errors could be transferred to the estimation of forest parameters. The shrub presence slightly decreases FRB model accuracy, increasing the MPE of the models in  $0.11 \text{ tons ha}^{-1}$ . This fact may have an effect on DTM accuracy. However, these findings should be further analyzed before generalizing them for the estimation of other forestry variables.

### 4.3. Future research

Overall, the use of low-point density ALS data allows generating accurate maps of FRB potential and quantifying bioenergy forest resources in *P. halepensis* forest at a regional scale. Furthermore, it provides information for forest managers related to the canopy bulk density and canopy fuel weight spatial distribution. These are two fuel metrics critical to assess wildfire risk in pyrophytes forest frequently affected by fires.

Future research should focus on developing more automatic selection methods, using Spearman's rank correlation, and different types of ALS metrics, in order to decrease modeling time. Furthermore, the use of multi-temporal ALS data has great potential for the analysis of changes in FRB. In this sense, the combination with multi-temporal passive remote sensing series could assist in the characterization of the evolution of this resource over the last decades. Finally, the capture of point clouds with Unmanned Aerial Vehicles would be of interest to provide 3D point clouds at lower expenses for small to medium areas.

## 5. Conclusions

This study assesses the usefulness of low-point density ALS data to accurately estimate FRB at a regional scale in a Mediterranean environment. Furthermore, it analyses the influence of environmental and ALS variables in model performance. Spearman's rank coefficient has been the most powerful selection method to generate representative and meaningful models at a large regional scale, followed by All subsets regression exhaustive method. The SVM with radial kernel method produced the most accurate model for FRB, which included three ALS metrics: the *P70*, the *Elev. variance*, and the *% first ret. above mean*. In general, no statistically significant differences were found between the generated models in terms of RMSE. The use of machine-learning regression methods may boost model performance when working in heterogeneous forest environments with a high number of field plots. Except for canopy pulse penetration, no statistically significant differences were found for the analyzed ALS characteristics and environmental variables. Point densities lower than  $1 \text{ point m}^{-2}$ , high scan angles, lower CPP, steep slopes and shrub presence tend to decrease model performance in  $0.56$ ,  $1.94$ ,  $5.47$ ,  $1.44$  and  $0.10 \text{ tons ha}^{-1}$ , respectively. The large-scale mapping of  $197,951.24 \text{ ha}$  representing  $87.66\%$  of the *P. halepensis* forested area in Aragón region sums up a total of  $3,627,021.25 \text{ tons}$  of FRB, being equal to  $1,584.91 \text{ ktoe}$ .

## Acknowledgements

This work was supported by the Government of Spain, Department of Education, Culture and Sports under Grant [FPU Grant BOE, 14/06250]. The authors are grateful to the material resources provided by the Centro Universitario de la Defensa de Zaragoza (CUD). We express our appreciation to Dr. Francisco Escribano and Jesús Cabrera for their help during the fieldwork.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the Government of Spain, Department of Education, Culture and Sports under Grant [FPU Grant BOE, 14/06250]. This work was supported by the Centro Universitario de la Defensa de Zaragoza [2017-09];

## ORCID

Dario Domingo  <http://orcid.org/0000-0002-8362-7559>

Antonio Luis Montealegre  <http://orcid.org/0000-0001-6288-2780>

María Teresa Lamelas  <http://orcid.org/0000-0002-8954-7517>

Alberto García-Martín  <http://orcid.org/0000-0003-2610-7749>

Juan de la Riva  <http://orcid.org/0000-0003-2615-270X>

Francisco Rodríguez  <http://orcid.org/0000-0002-4844-1759>

## References

- Andersen, H.-E., R. J. McGaughey, and S. E. Reutebuch. 2005. "Estimating Forest Canopy Fuel Parameters Using LIDAR Data." *Remote Sensing of Environment* 94 (4): 441–449. doi:10.1016/J.RSE.2004.10.013. Elsevier.
- "Anuario de Estadística Forestal 2013." 2018. Accessed December 6. [https://www.mapa.gob.es/es/ desarrollo-rural/estadisticas/forestal\\_anuario\\_2013.aspx](https://www.mapa.gob.es/es/ desarrollo-rural/estadisticas/forestal_anuario_2013.aspx)
- Austin, J. M., B. G. Mackey, and K. P. Van Niel. 2003. "Estimating Forest Biomass Using Satellite Radar: An Exploratory Study in a Temperate Australian Eucalyptus Forest." *Forest Ecology and Management* 176 (1–3): 575–583. doi:10.1016/S0378-1127(02)00314-6. Elsevier.
- Breidenbach, J., B. Koch, G. Kändler, and A. Kleusberg. 2008. "Quantifying the Influence of Slope, Aspect, Crown Shape and Stem Density on the Estimation of Tree Height at Plot Level Using Lidar and InSAR Data." *International Journal of Remote Sensing* 29 (5): 1511–1536. doi:10.1080/01431160701736364. Taylor & Francis .
- Cao, L., N. C. Coops, J. L. Innes, R. J. Stephen, L. F. Sheppard, H. Ruan, and G. She. 2016. "Estimation of Forest Biomass Dynamics in Subtropical Forests Using Multi-Temporal Airborne LiDAR Data." *Remote Sensing of Environment* 178: 158–171. doi:10.1016/j.rse.2016.03.012.
- Chasmer, L., C. Hopkinson, and P. Treitz. 2006. "Investigating Laser Pulse Penetration through a Conifer Canopy by Integrating Airborne and Terrestrial Lidar." *Canadian Journal of Remote Sensing* 32 (2): 116–125. doi:10.5589/m06-011.
- Clark, M. L., D. B. Clark, and D. A. Roberts. 2004. "Small-Footprint Lidar Estimation of Sub-Canopy Elevation and Tree Height in a Tropical Rain Forest Landscape." *Remote Sensing of Environment* 91 (1): 68–89. doi:10.1016/J.RSE.2004.02.008. Elsevier.
- Cowen, D. J., J. R. Jensen, C. Hendrix, M. E. Hodgson, and S. R. Schili. 2000. "A GIS-Assisted Rail Construction Econometric Model that Incorporates LIDAR Data." *Photogrammetric Engineering & Remote Sensing* 66 (11): 1323–1328.
- Dalponete, M., L. Frizzera, H. O. Ørka, T. Gobakken, E. Næsset, and D. Gianelle. 2018. "Predicting Stem Diameters and Aboveground Biomass of Individual Trees Using Remote Sensing Data." *Ecological Indicators* 85 (February): 367–376. doi:10.1016/J.ECOLIND.2017.10.066. Elsevier.
- Disney, M. I., V. Kalogerou, P. Lewis, A. Prieto-Blanco, S. Hancock, and M. Pfeifer. 2010. "Simulating the Impact of Discrete-Return Lidar System and Survey Characteristics over Young Conifer and Broadleaf Forests." *Remote Sensing of Environment* 114 (7): 1546–1560. doi:10.1016/J.RSE.2010.02.009. Elsevier.



- Domingo, D., M. T. Lamelas, A. Montealegre, A. García-Martín, and J. de la Riva. 2018. "Estimation of Total Biomass in Aleppo Pine Forest Stands Applying Parametric and Nonparametric Methods to Low-Density Airborne Laser Scanning Data." *Forests* 9 (4): 158. doi:10.3390/f9040158. Multidisciplinary Digital Publishing Institute.
- Evans, Jeffrey S., and Andrew T. Hudak. 2007. "A Multiscale Curvature Algorithm for Classifying Discrete Return LiDAR in Forested Environments." *IEEE Transactions on Geoscience and Remote Sensing* 45 (4): 1029–1038.
- García, M., S. Saatchi, A. Ferraz, C. A. Silva, S. Ustin, A. Koltunov, and H. Balzter. 2017. "Impact of Data Model and Point Density on Aboveground Forest Biomass Estimation from Airborne LiDAR." *Carbon Balance and Management* 12 (1): 4. doi:10.1186/s13021-017-0073-1. Nature Publishing Group.
- García-Gutiérrez, J., F. Martínez-Álvarez, A. Troncoso, and J. C. Riquelme. 2015. "A Comparison of Machine Learning Regression Techniques for LiDAR-Derived Estimation of Forest Variables." *Neurocomputing* 167: 24–31. doi:10.1016/j.neucom.2014.09.091.
- García-Martín, A., J. de la Riva, F. Pérez Cabello, R. Montorio, D. García Galindo, and J. P. Puigdevall. 2008. "Evaluation of the Effect of Temporality on Forest Residual Biomass Estimation Using Summer Landsat TM Imagery." In *16th European Biomass Conference. From Research to Industry and Markets*, (ISBN 978-88-89407-58-1 J. Schmid, J.-P. Grimm, P. Gelm, and A. Grassi edited by, 254–262. Florence, Italy: ETA-Florence Renewable Energies.
- García-Martín, A., J. de la Riva, F. Pérez-Cabello, and R. Montorio. 2012. "Using Remote Sensing to Estimate a Renewable Resource: Forest Residual Biomass." In *Remote Sensing of Biomass - Principles and Applications*. InTech. doi:10.5772/17353.
- Gaveau, D. L. A., and R. A. Hill. 2003. "Quantifying Canopy Height Underestimation by Laser Pulse Penetration in Small-Footprint Airborne Laser Scanning Data." *Canadian Journal of Remote Sensing* 29 (5): 650–657. doi:10.5589/m03-023. Taylor & Francis.
- Gleason, C. J., and I. Jungho. 2011. "A Review of Remote Sensing of Forest Biomass and Biofuel: Options for Small-Area Applications." *GIScience & Remote Sensing* 48 (2): 141–170. doi:10.2747/1548-1603.48.2.141. Taylor & Francis Group .
- Gobakken, T., and N. Erik. 2008. "Assessing Effects of Laser Point Density, Ground Sampling Intensity, and Field Sample Plot Size on Biophysical Stand Properties Derived from Airborne Laser Scanner Data." *Canadian Journal of Forest Research* 38 (5): 1095–1109. doi:10.1139/X07-219.
- Granados, M. E., A. Vilagrosa, E. Chirino, and V. R. Vallejo. 2016. "Reforestation with Resprouter Species to Increase Diversity and Resilience in Mediterranean Pine Forests." *Forest Ecology and Management* 362 (February): 231–240. doi:10.1016/J.FORECO.2015.12.020. Elsevier.
- Hamelin, L., M. Borzęcka, M. Kozak, and P. Rafał. 2019. "A Spatial Approach to Bioeconomy: Quantifying the Residual Biomass Potential in the EU-27." *Renewable and Sustainable Energy Reviews* 100 (February): 127–142. doi:10.1016/j.rser.2018.10.017.
- Hansen, E., T. Gobakken, E. Næsset, E. H. Hansen, T. Gobakken, and N. Erik. 2015. "Effects of Pulse Density on Digital Terrain Models and Canopy Metrics Using Airborne Laser Scanning in a Tropical Rainforest." *Remote Sensing* 7 (7): 8453–8468. doi:10.3390/rs70708453. Multidisciplinary Digital Publishing Institute.
- Hauglin, M., J. Dibdiakova, T. Gobakken, and E. Næsset. 2013. "Estimating Single-Tree Branch Biomass of Norway Spruce by Airborne Laser Scanning." *ISPRS Journal of Photogrammetry and Remote Sensing* 79: 147–156. doi:10.1016/j.isprsjprs.2013.02.013. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).
- Hauglin, M., T. Gobakken, R. Astrup, L. Ene, and N. Erik. 2014. "Estimating Single-Tree Crown Biomass of Norway Spruce by Airborne Laser Scanning: A Comparison of Methods with and without the Use of Terrestrial Laser Scanning to Obtain the Ground Reference Data." *Forests* 5 (3): 384–403. doi:10.3390/f5030384.
- Hauglin, M., T. Gobakken, V. Lien, O. M. Bollandssås, and N. Erik. 2012. "Estimating Potential Logging Residues in a Boreal Forest by Airborne Laser Scanning." *Biomass and Bioenergy* 36: 356–365. doi:10.1016/j.biombioe.2011.11.004.
- He, Q., E. Chen, A. Ru, L. Yong, H. Qisheng, E. Chen, A. Ru, and L. Yong. 2013. "Above-Ground Biomass and Biomass Components Estimation Using LiDAR Data in a Coniferous Forest." *Forests* 4 (4): 984–1002. doi:10.3390/f4040984. Multidisciplinary Digital Publishing Institute.
- Heinimo, J. 2018. "Solid and Liquid Biofuels Markets in Finland - a Study on International Biofuels Trade." Accessed October 28. <http://www.doria.fi/handle/10024/31077>

- Hill, R. A., and R. K. Broughton. 2009. "Mapping the Understorey of Deciduous Woodland from Leaf-on and Leaf-off Airborne LiDAR Data: A Case Study in Lowland Britain." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (2): 223–233. doi:10.1016/j.isprsjprs.2008.12.004. Elsevier B.V.
- Hirata, Y. 2004. "The Effects of Footprint Size and Sampling Density in Airborne Laser Scanning to Extract Individual Trees in Mountainous Terrain." *International Archives for Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (8): 102–107. <http://www.isprs.org/proceedings/XXXVI/8-W2/HIRATA.pdf>
- Hollaus, M., W. Wagner, C. Eberhöfer, and W. Karel. 2006. "Accuracy of Large-Scale Canopy Heights Derived from LiDAR Data under Operational Constraints in a Complex Alpine Environment." *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (5): 323–338. doi:10.1016/j.isprsjprs.2006.05.002.
- Holmgren, J., M. Nilsson, and H. Olsson. 2003. "Simulating the Effects of Lidar Scanning Angle for Estimation of Mean Tree Height and Canopy Closure." *Canadian Journal of Remote Sensing* 29 (5): 623–632. doi:10.5589/m03-030. Taylor & Francis.
- Hopkinson, C. 2006. "The Influence of Lidar Acquisition Settings on Canopy Penetration and Laser Pulse Return Characteristics." In *2006 IEEE International Symposium on Geoscience and Remote Sensing*, 2420–2423. IEEE. doi:10.1109/IGARSS.2006.627.
- Hyypä, J., H. Hyypä, P. Litkey, Y. Xiaowei, H. Haggren, P. Rönholm, U. Pyysalo, J. Pitkänen, and M. Maltamo. 2000. "Algorithms and Methods of Airborne Laser Scanning for Forest Measurements." *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (8): 82–89.
- Jakubowski, M. K., Q. Guo, and M. Kelly. 2013. "Tradeoffs between Lidar Pulse Density and Forest Measurement Accuracy." *Remote Sensing of Environment* 130 (March): 245–253. doi:10.1016/J.RSE.2012.11.024. Elsevier.
- Kankare, V., M. Rätty, X. Yu, M. Holopainen, M. Vastaranta, T. Kantola, J. Hyypä, H. Hyypä, P. Alho, and R. Viitala. 2013. "Single Tree Biomass Modelling Using Airborne Laser Scanning." *ISPRS Journal of Photogrammetry and Remote Sensing* 85: 66–73. doi:10.1016/j.isprsjprs.2013.08.008. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).
- Khosravipour, A., A. K. Skidmore, T. Wang, M. Isenburg, and K. Khoshelham. 2015. "Effect of Slope on Treetop Detection Using a LiDAR Canopy Height Model." *ISPRS Journal of Photogrammetry and Remote Sensing* 104 (June): 44–52. doi:10.1016/J.ISPRSJPRS.2015.02.013. Elsevier.
- Kristensen, T., E. Næsset, M. Ohlson, P. V. Bolstad, and R. Kolka. 2015. "Mapping Above- and Below-Ground Carbon Pools in Boreal Forests: The Case for Airborne Lidar." Edited by Krishna Prasad Vadrevu. *PloS One* 10 (10): e0138450. doi:10.1371/journal.pone.0138450. Public Library of Science.
- Lee, J., J. Im, K. Kim, L. Quackenbush, J. Lee, I. Jungho, K. Kim, and L. J. Quackenbush. 2018. "Machine Learning Approaches for Estimating Forest Stand Height Using Plot-Based Observations and Airborne LiDAR Data." *Forests* 9 (5): 268. doi:10.3390/f9050268. Multidisciplinary Digital Publishing Institute.
- Liu, J., A. K. Skidmore, S. Jones, T. Wang, M. Heurich, X. Zhu, and Y. Shi. 2018. "Large Off-Nadir Scan Angle of Airborne LiDAR Can Severely Affect the Estimates of Forest Structure Metrics." *ISPRS Journal of Photogrammetry and Remote Sensing*. doi:10.1016/j.isprsjprs.2017.12.004.
- Lovell, J. L., D. L. B. Jupp, G. J. Newnham, N. C. Coops, and D. S. Culvenor. 2005. "Simulation Study for Finding Optimal Lidar Acquisition Parameters for Forest Height Retrieval." *Forest Ecology and Management* 214 (1–3): 398–412. doi:10.1016/J.FORECO.2004.07.077. Elsevier.
- McGaughey, R. 2009. "FUSION/LDV: Software for LIDAR Data Analysis and Visualization." US Department of Agriculture, Forest Service, Pacific Northwest Research Station.
- Montaghi, A. 2013. "Effect of Scanning Angle on Vegetation Metrics Derived from a Nationwide Airborne Laser Scanning Acquisition." *Canadian Journal of Remote Sensing* 39 (sup1): S152–S173. doi:10.5589/m13-052. Taylor & Francis.
- Montealegre, A., M. T. Lamelas, and J. de la Riva. 2015b. "Interpolation Routines Assessment in ALS-Derived Digital Elevation Models for Forestry Applications." *Remote Sensing* 7 (7): 8631–8654. doi:10.3390/rs70708631. Multidisciplinary Digital Publishing Institute.



- Montealegre, A. L., M. T. Lamelas, and J. de la Riva. 2015a. "A Comparison of Open-Source LiDAR Filtering Algorithms in A Mediterranean Forest Environment." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (8): 4072–4085. doi:10.1109/JSTARS.2015.2436974. IEEE.
- Næsset, E. 2002. "Predicting Forest Stand Characteristics with Airborne Scanning Laser Using a Practical Two-Stage Procedure and Field Data." *Remote Sensing of Environment* 80 (1): 88–99. doi:10.1016/S0034-4257(01)00290-5.
- Næsset, E. 2009. "Influence of Terrain Model Smoothing and Flight and Sensor Configurations on Detection of Small Pioneer Trees in the Boreal–Alpine Transition Zone Utilizing Height Metrics Derived from Airborne Scanning Lasers." *Remote Sensing of Environment* 113 (10): 2210–2223. doi:10.1016/J.RSE.2009.06.003. Elsevier.
- Næsset, E., and Ø. Tonje. 2002. "Estimating Tree Height and Tree Crown Properties Using Airborne Scanning Laser in a Boreal Nature Reserve." *Remote Sensing of Environment* 79 (1): 105–115. doi:10.1016/S0034-4257(01)00243-7.
- Nemenyi, P. 1963. "Distribution-Free Multiple Comparisons." Ph.D. Thesis, Princeton University, Princeton, NY, USA.
- Ørka, H. O., O. M. Bollandsås, E. H. Hansen, E. Næsset, and T. Gobakken. 2018. "Effects of Terrain Slope and Aspect on the Error of ALS-Based Predictions of Forest Attributes." *Forestry: an International Journal of Forest Research* 91 (2): 225–237. doi:10.1093/forestry/cpx058. Oxford University Press.
- Plan Nacional de Ortofotografía Aérea LiDAR. Accessed December 12. <http://pnoa.ign.es/presentacion>
- Popescu, S. C. 2007. "Estimating Biomass of Individual Pine Trees Using Airborne Lidar." *Biomass and Bioenergy* 31 (9): 646–655. doi:10.1016/j.biombioe.2007.06.022.
- Richardson, J., R. Björheden, P. Hakkila, A. T. Lowe, and C. T. Smith. 2002. *Bioenergy from Sustainable Forestry*. Vol. 71. Forestry Sciences. Dordrecht: Kluwer Academic Publishers. doi:10.1007/0-306-47519-7.
- Roccaforte, J. P., P. Z. Fulé, and W. Wallace Covington. 2008. "Landscape-Scale Changes in Canopy Fuels and Potential Fire Behaviour following Ponderosa Pine Restoration Treatments." *International Journal of Wildland Fire* 17 (2): 293. doi:10.1071/WF06120. CSIRO PUBLISHING.
- Roussel, J.-R., J. Caspersen, M. Béland, S. Thomas, and A. Achim. 2017. "Removing Bias from LiDAR-Based Estimates of Canopy Height: Accounting for the Effects of Pulse Density and Footprint Size." *Remote Sensing of Environment* 198 (September): 1–16. doi:10.1016/J.RSE.2017.05.032. Elsevier.
- Ruiz, L. A., T. Hermosilla, F. Mauro, and M. Godino. 2014. "Analysis of the Influence of Plot Size and LiDAR Density on Forest Structure Attribute Estimates." *Forests* 5 (5): 936–951. doi:10.3390/f5050936. Multidisciplinary Digital Publishing Institute.
- Ruiz-Peinado, R., M. Del Rio, and G. Montero. 2011. "New Models for Estimating the Carbon Sink Capacity of Spanish Softwood Species." *Forest Systems* 20 (1): 176. doi:10.5424/fs/2011201-11643.
- Saldaña, C., and M. Ana 2010. "Bases Para La Gestión De Masas Naturales De Pinus Halepensis Mill. En El Valle Del Ebro." E.T.S.I. Montes (UPM). <http://oa.upm.es/4960/>
- Singh, K. K., G. Chen, J. B. McCarter, and R. K. Meentemeyer. 2015. "Effects of LiDAR Point Density and Landscape Context on Estimates of Urban Forest Biomass." *ISPRS Journal of Photogrammetry and Remote Sensing* 101: 310–322. doi:10.1016/j.isprsjprs.2014.12.021.
- Stojanova, D., P. Panov, V. Gjorgjioski, A. Kobler, and D. Sašo. 2010. "Estimating Vegetation Height and Canopy Cover from Remotely Sensed Data with Machine Learning." *Ecological Informatics* 5 (4): 256–266. doi:10.1016/j.ecoinf.2010.03.004.
- Straub, C., and B. Koch. 2011. "Enhancement of Bioenergy Estimations within Forests Using Airborne Laser Scanning and Multispectral Line Scanner Data." *Biomass and Bioenergy* 35 (8): 3561–3574. doi:10.1016/j.biombioe.2011.05.017. Elsevier Ltd.
- Velázquez Martí, B. 2006. "Situación De Los Sistemas De Aprovechamiento De Los Residuos Forestales Para Su Utilización Energética." *Ecosistemas* 15 (1): 77–86. doi:10.1080/026432900750038317.
- Wasser, L., R. Day, L. Chasmer, and A. Taylor. 2013. "Influence of Vegetation Structure on Lidar-Derived Canopy Height and Fractional Cover in Forested Riparian Buffers during

- Leaf-off and Leaf-on Conditions.” *PloS One* 8 (1): e54776. doi:[10.1371/journal.pone.0054776](https://doi.org/10.1371/journal.pone.0054776). Public Library of Science.
- Yu, X., J. Hyypä, H. Hyypä, and M. Maltamo. 2004. “Effects of Flight Altitude on Tree Height Estimation Using Airborne Laser Scanning.” *International Archives Of Photogrammetry Remote Sensing And Spatial Information Sciences* XXXVI (8–W2): 96–101.
- Zhao, K., J. C. Suárez, M. García, H. Tongxi, C. Wang, and A. Londo. 2018. “Utility of Multitemporal Lidar for Forest and Carbon Monitoring: Tree Growth, Biomass Dynamics, and Carbon Flux.” *Remote Sensing of Environment* 204 (January): 883–897. doi:[10.1016/J.RSE.2017.09.007](https://doi.org/10.1016/J.RSE.2017.09.007). Elsevier.





### 3.4. Temporal Transferability of Pine Forest Attributes Modelling Using Low-Density Airborne Laser Scanning Data

Transferibilidad temporal de modelos de variables forestales utilizando datos de escáner láser aeroportado de baja densidad






#### RESUMEN

Este estudio evalúa la transferibilidad temporal de modelos generados utilizando datos de escáner láser aeroportado (ALS) y adquiridos en dos fechas diferentes. La estimación de siete variables forestales (densidad de pies, área basal, diámetro cuadrático medio, diámetro dominante, altura dominante, volumen maderable y biomasa arbórea) se realizó utilizando un enfoque basado en áreas en masas forestales mediterráneas de Pino carrasco. Los datos ALS de baja densidad se adquirieron en 2011 y en 2016, mientras que las 147 parcelas de campo se muestrearon en 2013, 2014 y 2016. La generación de datos de campo para las fechas de los vuelos ALS se realizó mediante la aplicación de modelos de crecimiento de árbol individual. Cinco métodos de selección y cinco modelos de regresión fueron comparados para relacionar las observaciones tomadas en campo respecto a las métricas ALS. La selección de los mejores modelos de regresión ajustados para cada variable forestal, y separadamente para los años 2011 y 2016, se realizó utilizando un enfoque indirecto. El ajuste de los modelos y la transferibilidad temporal de los mismos se analizó extrapolando los mejores modelos ajustados para 2011 a 2016, e inversamente de 2016 a 2011. El modelo de regresión no paramétrico support vector machine con kernel radial mostró los mejores resultados. La diferencia en el error cuadrático medio en porcentaje de los modelos ajustados y extrapolados fue de 2,13% para los modelos de 2011 y 1,58% para los de 2016.



## Article

# Temporal Transferability of Pine Forest Attributes Modeling Using Low-Density Airborne Laser Scanning Data

Darío Domingo <sup>1,\*</sup> , Rafael Alonso <sup>2,3</sup>, María Teresa Lamelas <sup>1,4</sup> ,  
Antonio Luis Montealegre <sup>1</sup> , Francisco Rodríguez <sup>2,3,5</sup>  and Juan de la Riva <sup>1</sup> 

<sup>1</sup> GEOFOREST-IUCA, Department of Geography, University of Zaragoza, Pedro Cerbuna 12, 50009 Zaragoza, Spain; tlamelas@unizar.es (M.T.L.); monteale@unizar.es (A.L.M.); delariva@unizar.es (J.d.I.R.)

<sup>2</sup> föra forest technologies sll, Campus Duques de Soria s/n, 42004 Soria, Spain; rafa.alonso@fora.es (R.A.); paco.rodriguez@fora.es (F.R.)

<sup>3</sup> Sustainable Forest Management Research Institute University of Valladolid-INIA, Campus Duques de Soria s/n, 42004 Soria, Spain

<sup>4</sup> Centro Universitario de la Defensa de Zaragoza, Academia General Militar, Ctra. de Huesca s/n, 50090 Zaragoza, Spain

<sup>5</sup> EU Ingenierías Agrarias, Campus Duques de Soria s/n, Universidad de Valladolid, 42004 Soria, Spain

\* Correspondence: ddomingo@unizar.es; Tel.: +34-876-554-058

Received: 6 November 2018; Accepted: 21 January 2019; Published: 28 January 2019



**Abstract:** This study assesses model temporal transferability using airborne laser scanning (ALS) data acquired over two different dates. Seven forest attributes (i.e. stand density, basal area, squared mean diameter, dominant diameter, tree dominant height, timber volume, and total tree biomass) were estimated using an area-based approach in Mediterranean Aleppo pine forests. Low-density ALS data were acquired in 2011 and 2016 while 147 forest inventory plots were measured in 2013, 2014, and 2016. Single-tree growth models were used to generate concomitant field data for 2011 and 2016. A comparison of five selection techniques and five regression methods were performed to regress field observations against ALS metrics. The selection of the best regression models fitted for each stand attribute, and separately for both 2011 and 2016, was performed following an indirect approach. Model performance and temporal transferability were analyzed by extrapolating the best fitted models from 2011 to 2016 and inversely from 2016 to 2011 using the direct approach. Non-parametric support vector machine with radial kernel was the best regression method with average relative % root mean square error differences of 2.13% for 2011 models and 1.58% for 2016 ones.

**Keywords:** model temporal transferability; ALS; forest inventory; backdating; Mediterranean forest

## 1. Introduction

Forest ecosystems provide economic and social benefits to humankind [1,2] requiring accurate tools to monitor their dynamics over time [3]. Over the last decades, optical remote sensing techniques have been used for monitoring forest changes at regional scales with the support of field surveys (e.g., [4,5]). However, airborne laser scanning (ALS) is better adapted to characterize forest structure [6] and estimate forest inventory parameters, providing accurate information to perform forest management and planning [3]. Furthermore, costs of ALS-based inventories are comparable to those associated with traditional ground-based ones [7,8]. Despite the great potential of this technology, multi-temporal ALS data have been utilized less, as the availability of two or more surveys in the same area has been limited by acquisition costs as well as by the need of temporal-concomitant field data (e.g., [3,6,9,10]). Recently, organizations, companies, and countries have made an effort to gather multi-temporal

datasets in different years (e.g., [11–13]) allowing the estimation of biophysical properties in forested areas over time. As a result, height growth has been estimated using the single tree or the area-based approach [14–20]. Biomass and carbon dynamics has also been analyzed [3,9,21–27], while less studies have estimated volume [17,18,21], basal area [17], and site index [14,28]. Multi-temporal data has also been used for quantifying fire-induced changes to forest structure [29], gaps presence [20,30,31] or detecting defoliation [32]. However, to the best of our knowledge, some relevant forest inventory attributes, such as stand density, squared mean diameter, and dominant diameter, have not been estimated using multi-temporal ALS data.

The use of low-density ALS data has been successfully used for estimating forestry attributes in different forest ecosystems (e.g., [33–35]), being also the case in Mediterranean forests of Spain (e.g., [36–38]). The analysis of the influence of point density on model predictions have been analyzed by several authors (e.g., [39–41]), who established that point density has little or no effect on predictions as statistical metrics remain stable [42]. Furthermore, Garcia et al. [43], Singh et al. [44], and Ruiz et al. [45] pointed out that low-density datasets were a viable solution at regional scales. Furthermore, the use of multi-temporal, low-point density data has only been explored in boreal ecosystems [17,28] and in temperate forests [24,27] but not in other ecosystems, such as Mediterranean ones, which are characterized by a higher heterogeneity in terms of forest structure.

Direct and indirect approaches have been proposed to model forest attributes using multi-temporal ALS data over time [26]. The direct approach adjusts one model for one point in time and estimates the inventory attribute for another point in time [3,9]. Previously, the model to be extrapolated was generated through regression methods that related a suite of ALS-metrics with ground-truth data. This approach allowed the temporal transferability of models reducing modeling time and fieldwork, as it was not needed to revisit them when the time between the ALS surveys was not large [28]. In contrast, the indirect approach fits two different models and estimates the variables for each point in time [3,9,17,24,25,27]. Several examples of the evaluation of these two approaches can be found in the literature. For example, when estimating biomass and carbon fluxes, Zhao et al. [3] and Meyer et al. [25] achieved better results with the indirect approach while Cao et al. [9], Skowronski et al. [24], and Bollandsås et al. [26] found slightly better performance of the direct approach.

These aforementioned modeling strategies sometimes face a challenge when lacking temporally-concomitant field data to calibrate forest stand models [3]. To this end, forest growth models can serve as useful tools to calibrate forest stand variables in a specific point in time. Thus, empirical growth models have received particular attention since the beginning due to their usefulness. Nowadays state-space stand-level models [46], distribution-based models, and both individual-tree models and complex process-based eco-physiological models [47] have dramatically increased flexibility and realism to forest simulations. In this sense, individual-tree growth models are powerful tools to update stand variables to the Light Detection and Ranging (LiDAR) mission date. For example, the use of diameter at breast height (dbh) and the height growth values from general yield tables of the Spanish National Forest Inventory have been applied for estimating total tree biomass in Aleppo pine forests [36]. However, specific single-tree growth models calibrated with tree rings are more accurate, particularly for improving model consistency when working at regional scale.

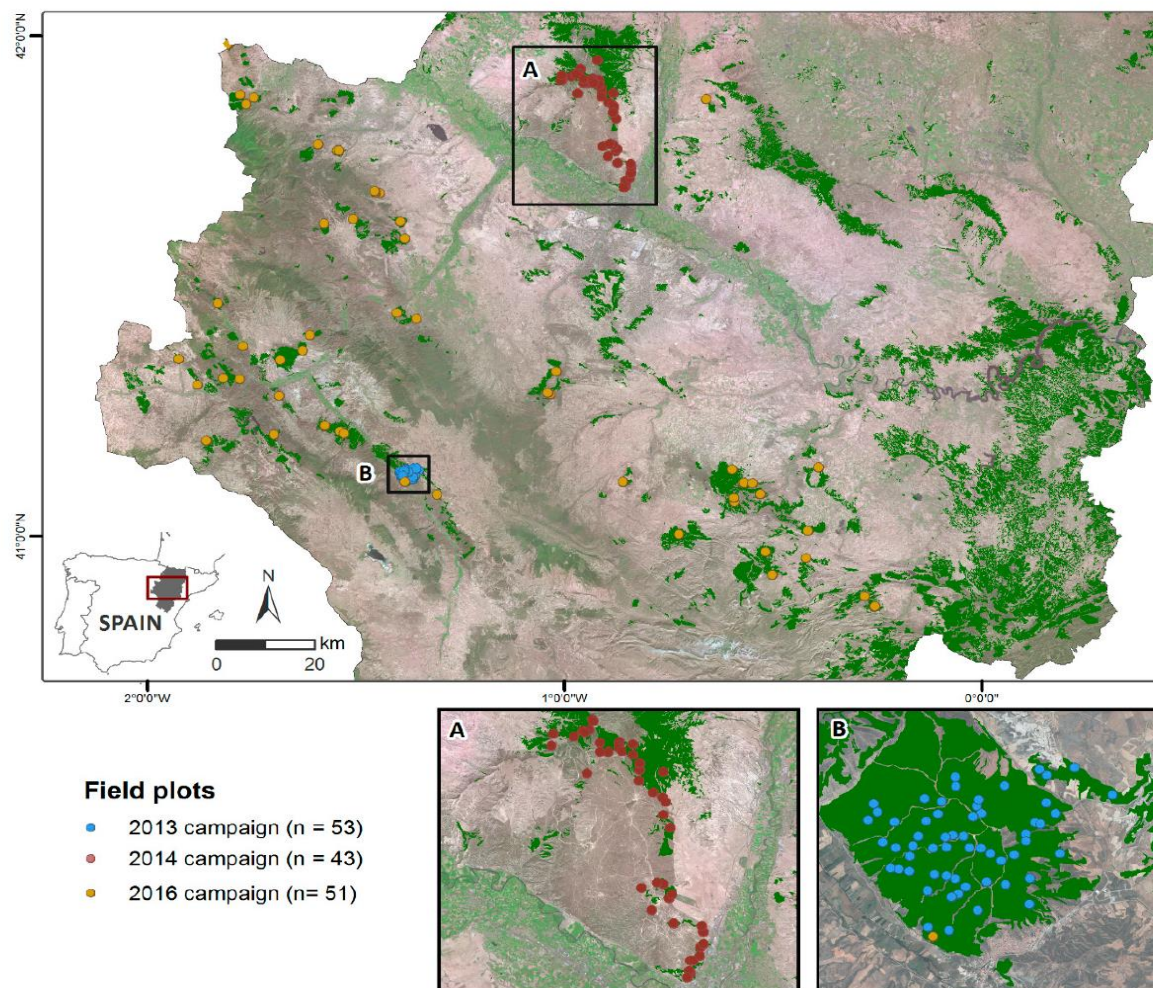
Thus, the aim of this study is to assess temporal transferability of several forest attributes models by comparing direct and indirect approaches using low-density ALS datasets collected in 2011 and 2016. Seven forestry attributes (i.e. stand density, basal area, squared mean diameter, dominant diameter, tree dominant height, timber volume, and total tree biomass) are estimated at regional scale in the Mediterranean Aleppo pine forest. First, an indirect approach fits two different models for 2011 and 2016 and estimates stand attributes for each point in time using different ALS-metrics and model parameters. Secondly, a direct approach extrapolates the models fitted previously, using the same variables and model parameters, to the other points in time.

Furthermore, the following secondary objectives were addressed: updating field inventory data collected in three different dates to the point clouds acquisition dates using single-tree growth models; comparing five selection methods and five regression methods in forest attributes modeling.

## 2. Materials and Methods

### 2.1. Study Area

The Aleppo pine (*Pinus halepensis* Mill.) forests under study are located in the Aragón region (Figure 1), Northeast Spain. This species represents 18.7% of the forested area, including semi-natural and reforested stands [48] and is well adapted to Mediterranean environmental conditions.



**Figure 1.** Study area with the location of forest inventory plots. High spatial resolution orthophotography from Spanish National Plan for Aerial Orthophotography spatial data infrastructure (SDI) service is used as a backdrop.

In this area, the annual precipitation ranges from 350 mm to 1000 mm [49]. The average annual temperature is 14 °C, with cold winters and hot and dry summers. Aleppo pine forests are characterized by a hilly topography, with elevations ranging from 300 to 1150 m above sea level and slopes from 0° to 39°. The lithology of the study area varies from Miocene carbonate and marl sediments to limestone platforms and Mesozoic and Eocene limestone.



Some pine stands are natural, but most stands were planted approximately forty to sixty years ago. The evergreen understory includes species such as *Quercus ilex* subsp. *rotundifolia*, *Quercus coccifera*, *Juniperus oxycedrus*, *Buxus sempervirens*, *Juniperus phoenicea*, *Rosmarinus officinalis*, and *Thymus vulgaris*, among many others. Reforested areas usually present a low presence of hardwood species and poor diversity [50,51], while natural stands are structurally complex with a developed and diverse understory [52].

## 2.2. Forest Inventory Data

Forest inventory data was acquired through 147 plots from three field campaigns performed during June to July 2013, July to September 2014, and April 2016 (Figure 1), from now on cited as first, second, and third campaign, respectively. The sampling data fulfilled the statistical requirements [53], considered the size of the study area, and the variability of the pine forest in terms of terrain slope, canopy height, and canopy cover (CC) [54].

Field data from the first campaign were acquired in 53 circular plots with a 15 m radius. A Leica VIVA®GS15 CS10 GNSS real-time kinematic global positioning system (GPS) was used to collect the center of each plot. Tree dbh, for those trees with a dbh larger than 7.5 cm, was measured at 1.3 m height using a diameter tape. Green crown height and height of up to 4 randomly selected sample trees within each plot were measured using a Suunto®hypsometer. Thus, diametric class was considered when selecting the sample trees.

Field data from the second campaign were collected in 43 circular plots with a 15 m radius. The center of the plots was referenced using the same Global Navigation Satellite System (GNSS) receiver as in the first campaign. Tree dbh was measured with the same criteria as the first campaign using a Haglöf Sweden®Mantax Precision Blue caliper. The green crown height and the total height of all trees in the plot were measured using a Haglöf Sweden®Vertex instrument.

Field data from the third campaign were acquired in 51 circular plots. The center of the designated circular plots was measured using a Trimble®GNSS receiver. Field plots with a 5.6 m (3 plots), 8.5 m (23 plots), 11.3 m (17 plots), and 14.10 m (7 plots) radius were collected. Tree dbh was measured at 1.3 m using the same caliper as in the second campaign. The green crown height and the height of up to the 6 nearest trees to the plot center were measured using a Haglöf Sweden®Vertex instrument. The sample was completed to achieve 100 dominant stems ha<sup>-1</sup> considering those with larger dbh.

The height for those trees not measured in the field plots was predicted using a height–diameter model developed from the sampled trees from all the field plots of the third campaign. Normality of the residuals, homoscedasticity, and independence or no auto-correlation in the residuals were verified for the linear regression fitted model.

## 2.3. Inventories Updating and Stand Variable Computation

Field data measurements were updated to year 2011 and 2016, which correspond to each ALS flight, to avoid any temporal lag between ALS-metrics and stand-level variables. The PHRAGON-2017 individual tree model was applied [55] through the forest simulator platform Simanfor [56]. This model enables tree-level distance-independent simulation of the development of Aleppo pine afforestations in Aragón. It consists of a set of equations for diameter over bark growth, diameter under bark growth, diameter under bark–diameter over bark ratio, generalized height–diameter relationship, volume over bark (taper equation) and crown ratio. In addition, it presents a survival model for the coming 10-year period and a classification tree for the regeneration of species of the genus *Pinus*, *Quercus*, and *Juniperus*, also in the coming decade. Explanatory variables included those related to tree size (diameter at breast height, total height), stand density (basal area, Hart–Becking index), dominant trees (dominant height, dominant diameter), competition (BALMOD) [57] and site quality (site index). Site index and dominant height evolution were estimated using the site index curves developed for natural Aleppo pine forests in the central Ebro valley [58].



Thus, when projecting future stand variables, diameter growth and survival equations were applied to every single tree in each plot, while the site index curve was used to forecast the future stand dominant height (and hence estimate the total height of each surviving tree). To reconstruct stand structure in the past, we need to deploy the diameter under bark growth equation, as it permits the use of the current stand features to predict the past growth of a tree (backdating procedure). Therefore, past tree diameters over bark are estimated through the diameter under bark –diameter over bark ratio–, while past dominant height can be calculated with the same site index curves, as they are dynamic, age-independent functions. Once every diameter and dominant height are known, the rest of the stand variables can be directly computed.

Seven forest inventory attributes were calculated from field data for each plot: stand density (N); basal area (G); squared mean diameter (Dg); dominant diameter (Do); dominant height (Ho); timber volume over bark of stem (V); and total tree biomass (W) [37] (Table 1). Thus, Ho is the mean height of the 100 trees per ha with largest dbh; Do is the mean dbh of the 100 trees per ha with largest dbh. V is estimated through the taper equation included in the PHRAGON-2017 individual-tree model [55]. W is computed as the sum of aboveground and belowground tree biomass using the Aleppo pine allometric equations developed by Ruiz-Peinado et al. [59].

**Table 1.** Forest inventory attributes.

| Date of the Campaign           | Field Data         | Variables                  | Units                           |
|--------------------------------|--------------------|----------------------------|---------------------------------|
| First: June to July 2013       | Green crown height | Stand density (N)          | stems ha <sup>-1</sup>          |
| Second: July to September 2014 | Total height       | Basal area (G)             | m <sup>2</sup> ha <sup>-1</sup> |
|                                | Dbh                | Squared mean diameter (Dg) | cm                              |
|                                |                    | Dominant diameter (Do)     | cm                              |
|                                |                    | Dominant height (Ho)       | m                               |
| Third: April 2016              |                    | Volume over bark (V)       | m <sup>3</sup> ha <sup>-1</sup> |
|                                |                    | Total tree biomass (W)     | tons ha <sup>-1</sup>           |

#### 2.4. ALS Data and Processing

The ALS data were acquired in 2011 and 2016 by the Spanish National Plan for Aerial Orthophotography (PNOA) [60]. The respective acquisition specifications are shown in Table 2. The point clouds, delivered in 2 × 2 km tiles in LAS binary file format, were captured with up to four returns measured per pulse. The x, y, and z coordinates were provided in European Terrestrial Reference System (ETRS) 1989 Universal Transverse Mercator (UTM) Zone 30 N.

**Table 2.** Technical specifications of airborne laser scanning (ALS) data.

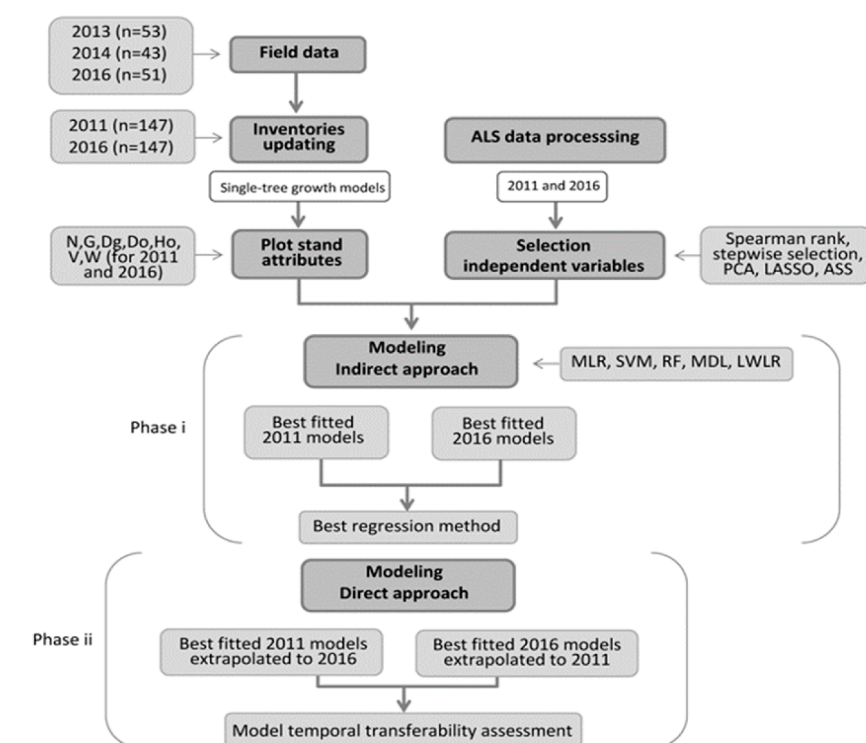
| Characteristics                        | Year 2011                   | Year 2016                   |
|--|-----------------------------|-----------------------------|
| Time period                            | January to February         | September to November       |
| Laser scanning system                  | Leica ALS60                 | Leica ALS80                 |
| Wavelength                             | 1,064 nm                    | 1064 nm                     |
| Average flying altitude over sea level | 3,000 m                     | 3150 m                      |
| Pulse repetition frequency             | ~70 kHz                     | 176–286 kHz                 |
| Scanning frequency                     | ~45 kHz                     | 28–59 Hz                    |
| Maximum scan angle                     | 29°                         | 25°                         |
| Nominal point density                  | 0.5 points m <sup>-2</sup>  | 1 points m <sup>-2</sup>    |
| Average point density                  | 0.64 points m <sup>-2</sup> | 1.25 points m <sup>-2</sup> |
| Accuracy of the point cloud (RMSEz)    | ≤0.2 m                      | 0.09 m                      |

After the noise removal from the point clouds, a verification of the overlapping returns was performed considering vertical and horizontal displacements. Thus, overlapping returns were removed from 105 tiles for the 2011 ALS flight. The subsequent steps were evenly applied for both ALS campaigns. The multiscale curvature classification algorithm [61], implemented in MCC 2.1

command line tool, was used to classify ground points according to Montealegre et al. [62]. A digital elevation model (DEM) with a 1-m size grid was generated using the Point-Triangulated Irregular Network-Raster interpolation method [61], implemented in ArcGIS 10.5 software. This DEM was used for point height normalization. The point clouds were clipped to the spatial extent of each field plot. Then, a full suite of statistical metrics related to height distribution and canopy cover was calculated [63] using FUSION LDV 3.60 [64] software. A threshold value of 2 m height was applied to remove ground and understory laser hits before generating the ALS-derived variables according to Nilsson [65] and Næsset and Økland [66]. For better understanding of the results, the ALS metrics were classified into three macro classes and seven classes (see Table A1 in Appendix A). Canopy height metrics (CHMs) were subdivided into lower, mean, and higher height variables; canopy height variability metrics (CHVMs) were subdivided into variability and variability metrics derived from the L moments [47]; and canopy density metrics (CDMs) were subdivided into percentage of first or all returns, canopy relief ratio (CRR), and the ratio between all returns and total returns.

## 2.5. Modeling of Forest Stand Attributes and Temporal Transferability Assessment

Figure 2 depicts in a graphical way the steps followed in the methodology in which two main phases can be distinguished: (i) the selection of variables and the forest attributes modeling in 2011 and 2016 using the indirect approach and (ii) the temporal transferability assessment applying a direct approach.



**Figure 2.** Methodology for forest stand attributes estimation using indirect and direct approaches.

### 2.5.1. Variable Selection and Attributes Modeling Using the Indirect Approach

Forest stand attributes modeling using the indirect approach was performed by two steps: (i) selection of the suitable ALS metrics using five selection approaches, and (ii) estimation of each stand attribute using five types of regression methods for 2011 and 2016 years (Figure 2). Thus, each of

the computed models have associated a selection approach, which determined the ALS metrics to be included in the models.

As described by Domingo et al. [36], different selection methods were applied to choose the ALS-metrics that present the best relationship with the forest inventory attribute at plot-level: (i) Spearman's rank correlation coefficient considering a minimum positive and negative rho value of 0.5; (ii) stepwise selection using backward, forward, and bidirectional approaches; (iii) principal component analysis (PCA) using varimax rotation to better interpret the results [67,68]; (iv) last absolute shrinkage and selection operator (LASSO) [69]; and (v) all subset selection (ASS) implementing exhaustive, backward, forward, and sequential replacement (Seqrep) approaches [70].

The selection of ALS-metrics was restricted to a combination of up to four independent variables using the mentioned selection methods to avoid variable multicollinearity, overfitting [71], and within the purpose of generating parsimonious models for forest management.

The estimation of forest stand attributes using an area-based approach and ALS data is usually done using either parametric (i.e. multiple linear regression) or non-parametric approaches such as regression trees, random forest, support vector machine, k-nearest neighbor, artificial neural network among others [72]. Five different regression methods, as described by Domingo et al. [36], were compared to estimate the seven forest inventory attributes (Table 1): multiple linear regression model (MLR), support vector machine (SVM), random forest (RF), locally weighted linear regression (LWLR), and linear model with a minimum length principle (MDL).

In the case of multiple linear regression model (MLR) normality, homoscedasticity, independence, and no auto-correlation of the residuals were verified. Logarithmic transformation of the dependent variables and the independent ones was also performed in those cases where statistical assumptions of linear regression were not fulfilled [73–75] or to improve the goodness-of-fit of the models. Support vector machine was computed using two kernel variants, linear (SVMl), and radial (SVMr) ones. Cost and gamma SVM parameters were tuned applying an interval of 1–1000 and 0.01–1, respectively. Random forest was implemented in R using “randomForest” [76] and “caret” packages [77], including “corr.bias” parameter to minimize bias effects. The model was tuned by applying a number of trees to growth (ntrees) within the interval 1–3000 and a number of variables to divide the nodes between 1 and 3. Locally weighted linear regression and MDL tree structures were computed using the R package “CORElearn” considering up to four ALS metrics. Model computation required the splitting of the original sample into a training set of 75% of the cases (110 plots) and a test set of 25% of the cases (37 plots). Validation was performed for all the models, being executed 100 times for those methods with associated randomness as SVM, RF, LWLR, and MDL to increase robustness in the results [78]. Furthermore, data were normalized in values ranging from 0 to 1 in order to avoid weights saturation according to Görgens [79].

Statistic performance of each computed model was reported including root mean square error (RMSE), relative RMSE respect to the mean (%RMSE) and bias. Differences between models were determined by using the Friedman nonparametric test according to the RMSE of each plot [80]. Furthermore, the Nemenyi post-hoc test was applied to determine whether differences were statistically significant, with a significance level of 0.05 [81]. This test was applied only when the null-hypothesis of the Friedman test was rejected, thus implying non-equivalence between models.

## 2.5.2. Assessment of Temporal Transferability by Applying a Direct Approach

The temporal transferability of models were assessed by three steps: (i) selection of the best regression model previously generated by the indirect approach for each forest stand attribute and year (2011 and 2016); (ii) extrapolation of the selected models from 2011 to 2016 ALS data, using the same variables and model parameters, and inversely from 2016 to 2011 by using the direct approach; (iii) performance comparison between extrapolated models for both years (Figure 2). Thus, Friedman and Nemenyi tests were applied for selecting the best regression model for each year (step i) and for selecting the best transferable models for both years (step iii).

### 3. Results

#### 3.1. Field Plot Computation

Equation (1) was used for estimating tree height (ht) for those trees not measured in the field plots for the first and third campaign. Model performance for the ht model was as follow: RMSE of 0.80 m and  $R^2$  of 0.93.

$$ht = \left( 1.3^{2.5511} + \left( H_0^{2.5511} - 1.3^{2.5511} \right) \cdot \frac{1 - \exp(-0.025687 \cdot dbh)}{1 - \exp(-0.025687 \cdot D_0)} \right)^{\frac{1}{2.5511}} \quad (1)$$

where  $ht$  is tree height (m),  $dbh$  is the diameter at breast height (cm),  $H_0$  and  $D_0$  are the dominant height and dominant diameter, as defined in Section 2.3.

Table 3 shows a summary of the forest inventory attributes obtained from the field plot data. The N values of the inventoried plots ranges from 99.03 to 3200 stems  $ha^{-1}$  and G ranges from 0.82 to 58.89  $m^2 ha^{-1}$ , presenting also a variety of diameters, from 9.21 to 47.96 cm. V and W data show also a high range of values with a standard deviation in both cases higher than 60 tons  $ha^{-1}$ . The high range and standard deviation values of the forest inventory attributes show the variability that characterizes Aleppo pine forest in Aragón region.

**Table 3.** Summary of the field plot characteristics ( $n = 147$ ).

| Forest Inventory Attribute | Min.  | Max.    | Range   | Mean   | Standard Deviation |
|----------------------------|-------|---------|---------|--------|--------------------|
| N (stems $ha^{-1}$ )       | 99.03 | 3200.00 | 3100.97 | 715.61 | 486.54             |
| G ( $m^2 ha^{-1}$ )        | 0.82  | 58.89   | 58.07   | 21.47  | 10.04              |
| Dg (cm)                    | 9.04  | 43.52   | 34.48   | 21.67  | 8.01               |
| Do (cm)                    | 9.21  | 47.96   | 38.76   | 27.79  | 8.73               |
| Ho (m)                     | 4.69  | 18.90   | 14.21   | 11.32  | 3.54               |
| V ( $m^3 ha^{-1}$ )        | 2.21  | 467.62  | 465.41  | 118.71 | 77.79              |
| W (tons $ha^{-1}$ )        | 2.89  | 373.02  | 370.14  | 101.91 | 60.69              |

Table 4 shows a summary of the estimated field plot attributes using single-tree growth models for each measured stand variable and both years. N shows a general decrease in the number of stems  $ha^{-1}$ , which may be caused by tree growth, resulting in an average of N change of 10.56 stems  $ha^{-1}$ . G shows average values of change of 2.55  $m^2 ha^{-1}$  and Dg and Do changes range from around 1.73 to 2.13 cm of growth, respectively. Ho values show an average increment of 0.62 m, ranging from 0.35 to 1.89 m. V and W changes show similar values ranging from around 2.00 to 50.00 tons  $ha^{-1}$  with average values around 17.00 tons  $ha^{-1}$ .

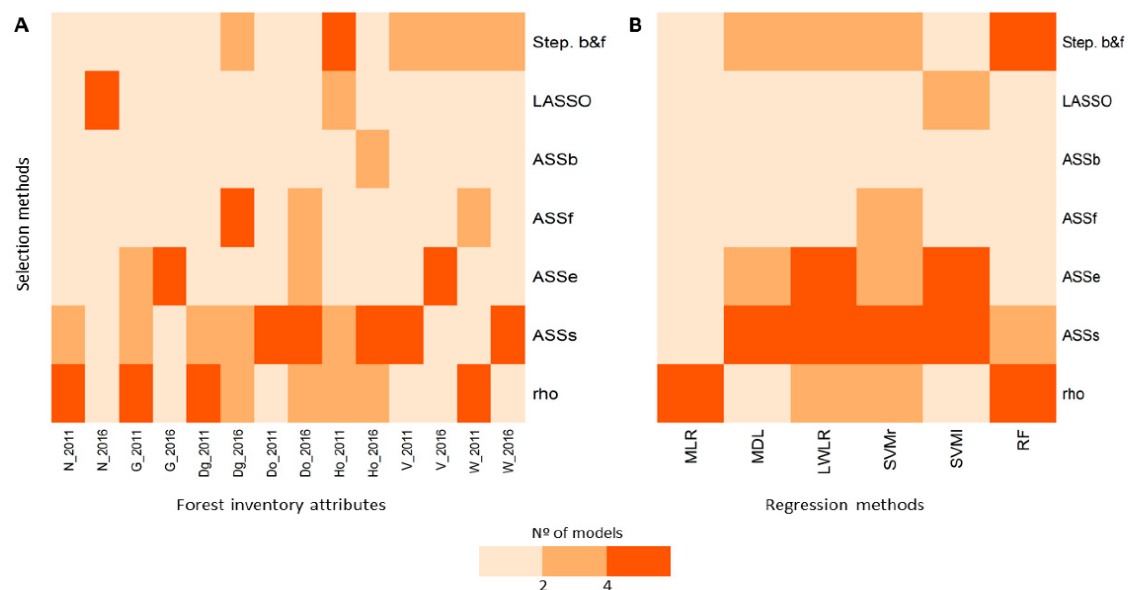
**Table 4.** Summary of the estimated field plot attributes using single-tree growth models for each year.

| Inventory Attribute  | Min. 2011 | Min. 2016 | Max. 2011 | Max. 2016 | Range 2011 | Range 2016 | Mean 2011 | Mean 2016 | SD 2011 | SD 2016 |
|----------------------|-----------|-----------|-----------|-----------|------------|------------|-----------|-----------|---------|---------|
| N (stems $ha^{-1}$ ) | 99.03     | 99.03     | 3405.67   | 3161.81   | 3306.64    | 3062.79    | 709.64    | 699.20    | 500.86  | 481.00  |
| G ( $m^2 ha^{-1}$ )  | 0.11      | 0.91      | 57.56     | 58.69     | 57.45      | 57.77      | 19.71     | 22.26     | 9.97    | 10.14   |
| Dg (cm)              | 3.29      | 9.55      | 41.41     | 45.05     | 38.12      | 35.50      | 20.72     | 22.45     | 7.99    | 8.40    |
| Do (cm)              | 3.35      | 9.72      | 45.85     | 49.19     | 42.50      | 39.47      | 26.59     | 28.72     | 8.84    | 9.09    |
| Ho (m)               | 4.24      | 4.90      | 18.46     | 19.08     | 14.22      | 14.17      | 10.97     | 11.58     | 3.70    | 3.60    |
| V ( $m^3 ha^{-1}$ )  | 0.35      | 2.51      | 454.77    | 476.02    | 454.42     | 473.51     | 107.31    | 126.45    | 74.83   | 81.48   |
| W (tons $ha^{-1}$ )  | 1.34      | 3.14      | 359.22    | 377.82    | 357.88     | 374.68     | 92.46     | 108.26    | 58.10   | 63.63   |

#### 3.2. Variable Selection

This section includes the results of the selection of ALS variables for the seven estimated forestry attributes modeled with the different regression methods.

Figure 3A synthesizes the performance of the analyzed selection methods for each forest stand attributes by year. All subsets regression Seqrep (ASSs) was the most powerful selection method. Spearman's rank correlation ( $\rho$ ) coefficient also showed good results, especially for selecting N, G, Dg, and W in 2011. All subsets regression Exhaustive (ASSe) and Stepwise (Step b&f) were both good selection methods for estimating G, Ho, and V. However, lasso selection (LASSO), all subsets regression Backward (ASSb) and all subsets regression Forward (ASSf) have been less utilized. The stepwise forward and PCA selection methods have not been included in Figure 3 as they did not determine the best variables for modeling in any of the cases. For detailed information of the selection methods performance, see Tables A2–A15 in the Appendix A.



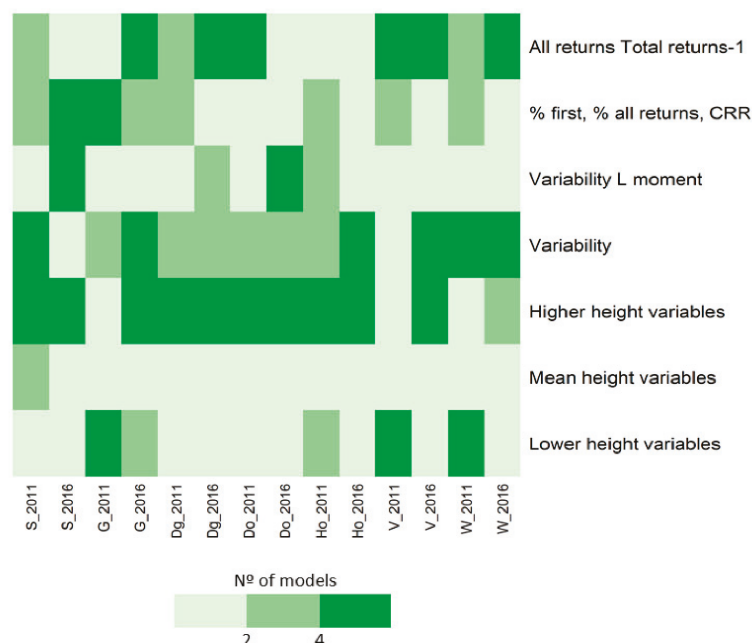
**Figure 3.** Performance of selection methods for each forest inventory attribute by year (3A) and performance of selection techniques for each regression method without considering the year of the model (3B). Maximum number of computed models in Figure 3A was six for all the stand attributes except for volume (V) and total tree biomass (W), which have a maximum number of five models. Maximum number of models in Figure 3B is 14, seven for each year and stand attribute, for all regression methods. Rho stands for Spearman Rank; ASSs stands for All Subject Selection Seqrep; ASSe stands for All Subject Selection Exhaustive; ASSf stands for All Subject Selection Forward; ASSb stands for All Subject Selection Backward; and Step. b&f stands for Stepwise Selection Both Backward and Forward.

Figure 3B depicts the performance of selection methods associated to each regression method without considering the year of the model. The ASSs was the most powerful method to select the best ALS metrics when using the MDL, LWLR, SVMr, and SVMl regression methods. Furthermore,  $\rho$  coefficient was the most powerful one when using the MLR and RF regression methods. The ASSe and Step b&f both were also good selection techniques for almost all the regression methods, excluding MLR. However, LASSO, ASSb, and ASSf were less effective.

Figure 4 shows the ALS selected metrics for estimating the forest inventory attributes for both 2011 and 2016 years. As mentioned in Section 2.4, for better understanding of results, the ALS metrics are classified into groups (see Table A1 in Appendix A). In general, higher height variables, variability, and variables related to the ratio between all returns and total returns were included in most of the models, while height variables and variability L moments were less demanded. Comparing the different estimated forest attributes, some differences can be observed. N, Dg, and Do estimations



included higher height variables, variability metrics, and CDM metrics. Ho estimation usually required higher height variables and variability metrics, while CDM metrics were not included. G, V, and W estimations included either lower or higher height metrics, variability metrics, and CDM ones.

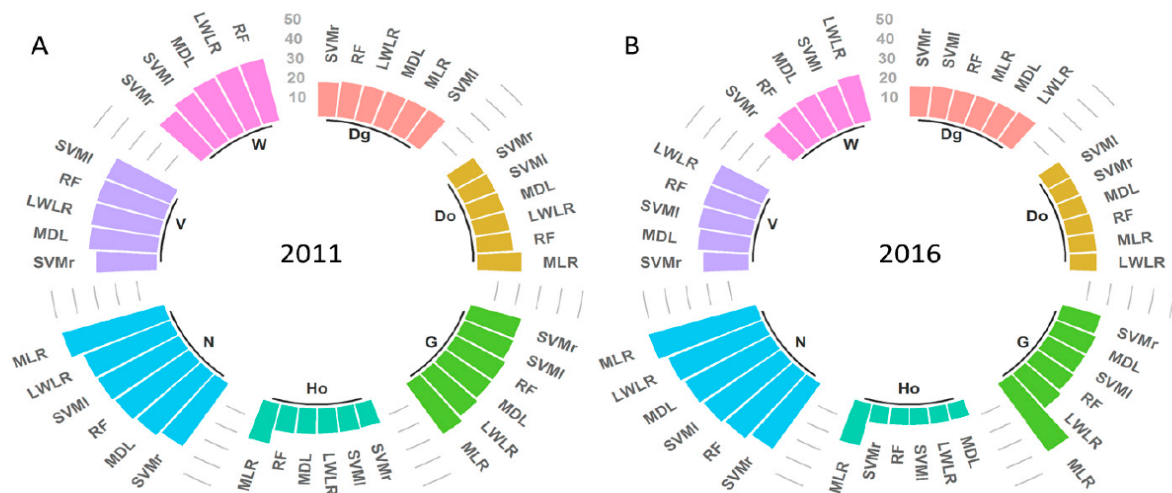


**Figure 4.** ALS selected metrics for estimating forest stand attributes for both 2011 and 2016 years. Maximum number of models is six for all the stand attributes except for V and W, which have a maximum number of five models.

### 3.3. Indirect Approach

The regression models to estimate Dg, Do, G, Ho, N, V, and W for 2011 and 2016 years using the indirect approach are summarized in Tables A2–A15 of the Appendix A. Figure 5 summarizes the %RMSE respect to the mean of the different regression methods for estimating the forest inventory attributes. Models developed for 2016 (Figure 5B) present generally higher accuracy than the ones generated for 2011 (Figure 5A). The point density of ALS datasets may determine these differences in accuracy. The SVMr shows the lower RMSE when modeling all the analyzed stand attributes in 2011 (Figure 5A). In this year, SVMl was the second best model when estimating Do, G, Ho, and W; MDL when estimating N and V; and RF when estimating Dg. The MLR regression method was not computed for V and W, as statistical assumptions of linear regression were not fulfilled, even considering logarithmic transformation. The MLR shows the lowest accuracy when estimating Do, G, Ho, and N.





**Figure 5.** % root mean square error (RMSE) respect to the mean of the different regression methods for estimating the forest inventory attributes for 2011 (A) and 2016 (B).

The SVMr is the best model for estimating N, V, W, Dg, and G in 2016 (Figure 5B). However, in this year, SVMl and MDL outperformed SVMr when estimating Do and Ho, respectively. In 2016, MLR shows the lowest accuracy when estimating G, Ho, and N.

Friedman tests shows that the models are not equivalent, with a  $p$ -value lower than 0.05 when testing whether there were statistically significant differences between regression methods for 2011 and 2016 years. Thereby, the post-hoc Nemenyi test indicates that no statistically significant differences exist between the methods, with 95% of probability, except for MLR. In this sense, statistically significant differences were found when comparing models in the following cases: between MLR and SVMr models when estimating Do and G for 2011; between MLR and MDL models when estimating Ho for 2011; between MLR and MDL models when estimating Dg for 2016; and between MLR and all the generated models when estimating G, Ho, and N for 2016.

### 3.4. Direct Approach

The SVMr was established as the regression method for analyzing how models fitted at 2011 perform at 2016, and inversely, following a direct approach to analyze temporal transferability. This method resulted the best estimator for all the models generated in 2011 and for the majority of forest attributes modeled in 2016.

Table 5 summarizes the best-selected SVMr models fitted in 2011 and the ones extrapolated to 2016 by using the same ALS metrics and model parameters. Table 6 summarizes the best-selected SVMr models fitted in 2016 and extrapolated to 2011 by using the same ALS metrics and model parameters. Furthermore, scatter plots of observed and predicted values for the analyzed forest stand attributes for both years are included in Figures A1 and A2 of Appendix A.

**Table 5.** Summary of the best-selected SVMr 2011 models and 2016 extrapolated ones. *ret.* refers to returns; *e* is extrapolated; N is stand density; G is basal area; Dg is squared mean diameter; Do is dominant diameter; Ho is dominant height; V is timber volume over bark of stem; W is total tree biomass.

| Attribute | ALS Metrics  | Fitting Phase |        |       | Validation |        |       |                |
|-----------|--|---------------|--------|-------|------------|--------|-------|----------------|
|           |  | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | R <sup>2</sup> |
| N 2011    | P99 + ElevIQ + % first ret. Above 2.00                                     | 257.09        | 36.34  | 28.81 | 272.76     | 38.55  | 26.99 | 0.72           |
| N 2016e   |  | 265.62        | 38.10  | 17.99 | 295.83     | 42.43  | 20.49 | 0.64           |
| G 2011    | Elev. minimum + Elev. kurtosis + % first ret. above mean                   | 4.43          | 22.77  | −0.10 | 4.77       | 24.51  | −0.10 | 0.77           |
| G 2016e   |  | 4.18          | 19.01  | 0.20  | 5.51       | 25.05  | 0.57  | 0.71           |
| Dg 2011   | P90 + Elev. SD + % first ret. above mean                                   | 3.38          | 16.38  | 0.19  | 3.56       | 17.25  | 0.06  | 0.81           |
| Dg 2016e  |  | 3.02          | 13.48  | 0.19  | 3.43       | 15.35  | 0.06  | 0.85           |
| Do 2011   | P90 + (All ret. Above 2)/(total first ret) × 100                           | 4.11          | 15.53  | 0.19  | 4.07       | 15.36  | 0.11  | 0.79           |
| Do 2016e  |  | 3.43          | 11.99  | 0.41  | 3.53       | 12.33  | 0.31  | 0.86           |
| Ho 2011   | P90 + Elev. variance + % all ret. above mean                               | 1.32          | 12.11  | 0.11  | 1.34       | 12.30  | 0.09  | 0.87           |
| Ho 2016e  |  | 0.86          | 7.47   | 0.10  | 0.98       | 8.54   | 0.10  | 0.93           |
| V 2011    | Elev. L2 + Elev. cubic mean + % first ret. above mean                      | 28.87         | 27.42  | 2.59  | 29.71      | 28.22  | 1.79  | 0.84           |
| V 2016e   |  | 25.03         | 20.15  | 3.14  | 26.00      | 20.92  | 2.64  | 0.90           |
| W 2011    | P10 + Elev. Quadratic mean + (All ret. Above mean)/(total first ret) × 100 | 23.00         | 25.29  | 0.75  | 24.29      | 26.71  | −0.03 | 0.82           |
| W 2016e   |  | 19.63         | 18.41  | 1.80  | 21.39      | 20.06  | 1.08  | 0.89           |

**Table 6.** Summary of the best-selected SVMr 2016 models and 2011 extrapolated ones. *ret.* refers to returns; *e* is extrapolated; N is stand density; G is basal area; Dg is squared mean diameter; Do is dominant diameter; Ho is dominant height; V is timber volume over bark of stem; W is total tree biomass.

| Attribute | ALS Metrics  | Fitting Phase |        |       | Validation |        |       |                |
|-----------|--|---------------|--------|-------|------------|--------|-------|----------------|
|           |  | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | R <sup>2</sup> |
| N 2011e   | Elev. maximum + Elev. L kurtosis + % first ret. Above 2.00             | 256.69        | 36.28  | 33.73 | 340.20     | 48.09  | 49.31 | 0.55           |
| N 2016    |  | 250.87        | 35.98  | 13.95 | 278.58     | 39.96  | 11.83 | 0.67           |
| G 2011e   | P75 + Elev. CUR mean CUBE + (All ret. Above 2)/(total first ret) × 100 | 4.97          | 25.54  | 0.26  | 5.04       | 25.88  | 0.13  | 0.74           |
| G 2016    |  | 3.88          | 17.61  | 0.41  | 4.14       | 18.80  | 0.30  | 0.84           |
| Dg 2011e  | Elev. maximum + Elev. IQ + (All ret. Above 2)/(total first ret) × 100  | 3.54          | 17.14  | 0.14  | 3.77       | 18.25  | 0.00  | 0.79           |
| Dg 2016   |  | 3.03          | 13.53  | 0.21  | 3.42       | 15.28  | 0.11  | 0.85           |
| Do 2011e  | P99 + Elev. CV   | 4.20          | 15.85  | 0.25  | 4.18       | 15.79  | 0.16  | 0.78           |
| Do 2016   |  | 3.25          | 11.35  | 0.40  | 3.40       | 11.89  | 0.33  | 0.87           |
| Ho 2011e  | P95 + Elev. SD   | 1.32          | 12.12  | 0.03  | 1.38       | 12.64  | 0.03  | 0.86           |
| Ho 2016   |  | 0.86          | 7.48   | 0.03  | 1.02       | 8.83   | 0.03  | 0.92           |
| V 2011e   | P75 + Elev. CUR mean CUBE + (All ret. Above 2)/(total first ret) × 100 | 29.97         | 28.46  | 1.51  | 30.96      | 29.40  | 0.84  | 0.83           |
| V 2016    |  | 24.69         | 19.87  | 2.65  | 26.35      | 21.20  | 1.92  | 0.90           |
| W 2011e   | Elev. L2 + Elev. CUR mean CUBE + % first ret. Above 2.00               | 23.11         | 25.42  | 0.96  | 23.36      | 25.69  | 0.27  | 0.83           |
| W 2016    |  | 18.82         | 17.65  | 1.23  | 20.06      | 18.81  | 0.56  | 0.90           |

In the case of the models fitted in 2011 and extrapolated to 2016 (Table 5), the %RMSE after validation ranges from 8.54 to 42.43% and R<sup>2</sup> ranges from 0.64 to 0.93 within the different stand attributes. As it is shown in Table 5, models are transferable. In fact, the average %RMSE differences between the fitted and the extrapolated models is 3.87%. Dg, Do, Ho, V, and W estimations for 2011 models have higher %RMSE than the one for models extrapolated to 2016. However, N and G models show higher %RMSE for the 2016 extrapolated ones.

In the case of the models fitted in 2016 and extrapolated to 2011 (Table 6), the %RMSE in the validation sample ranges from 8.83 to 48.09% and R<sup>2</sup> ranges from 0.55 to 0.92 within the different stand attributes. These models also show good temporal transferability, being the average %RMSE differences between the fitted and the extrapolated model 5.85%, even lower than the models fitted in

2011 and extrapolated to 2016. All the models fitted in 2016 for the analyzed stand attributes present lower RMSE than the ones extrapolated to 2011.

The comparison between fitted models generated for either 2011 or 2016 and the extrapolated ones were assessed using Friedman and post-hoc Nemenyi tests. Friedman test shows that the models for the analyzed stand attributes are not equivalent with a  $p$ -value lower than 0.05 when testing whether there were differences: (i) between models fitted in 2011 and the ones extrapolated to 2016; (ii) between models fitted in 2016 and the ones extrapolated to 2011; (iii) between models fitted in 2011 and models fitted in 2016; (iv) between models fitted in 2011 and models extrapolated to 2011; (v) between models fitted in 2016 and models extrapolated to 2016. Thereby, the required application of the post-hoc Nemenyi test indicates that no statistically significant differences exist between the methods for the proposed hypothesis, with 95% of probability.

#### 4. Discussion

Airborne laser scanning is considered the best technology for mapping 3D vegetation structures [3] allowing the measurement of fine-scale forestry metrics [82]. Multi-temporal ALS data has been less explored as the availability of two or more LiDAR surveys in the same area is still limited. Nevertheless, several studies have used multi-temporal small-footprint ALS to estimate total tree biomass or carbon dynamics [3,9,21–27], volume changes [17,21], height growth [14–19], and basal area [17]. This study estimates seven forest attributes (N, G, Dg, Do, Ho, V, W) using bi-temporal low-point density ALS data in Mediterranean Aleppo pine heterogeneous forests. The high number of field plots has allowed estimating the seven mentioned forest attributes for 18.7% of the forested area in Aragón, providing results at a regional scale. Moreover, model temporal transferability was demonstrated which could improve forest management in a cost-effective way in Mediterranean Aleppo pine forests.

Multi-temporal LiDAR estimations of forest attributes requires the support of accompanying field surveys [3] being desirable to have them corresponding to LiDAR surveys [9]. Field surveys are cost and time demanding specially when acquiring a high number of plots. The use of specific individual-tree growth equations, derived from dbh growth by extracting tree cores or from interval or permanent plots, is a good way to get value from field plot inventories acquired between different ALS surveys. Diameter at breast height and height growth values from general yield tables from the Spanish National Forest Inventory have been applied in other studies for estimating total tree biomass in Aleppo pine forests [36]. Nevertheless, in this work specific single-tree growth models, generalized height-diameter curves and taper equations were used to estimate all stand attributes in the measured field plots at two different points in time, which produces more accurate results, particularly when predicting at short term [83].

The use of selection methods reduces variable collinearity, modeling time and increases model parsimony. All subset selection Seqrep was the most powerful selection method in agreement with Hansen et al. [84] who also used similar best subset regression procedures to estimate biomass with ALS data. Spearman rho coefficients, proposed as a good tool for determining the relationships between ALS and field metrics by Kristensen et al. [85], also showed a good result, agreeing with our previous studies [37]. Furthermore, our results agree with García-Gutiérrez et al. [78], who found that stepwise was a powerless technique. Accordingly, the use of automatic selection methods such as ASSs is proposed when using MDL, LWLR, SVMr, and SVMl regression methods in Mediterranean Aleppo pine forest. Nevertheless, comparison between selection methods should be considered when working with other forest types or species. In this sense, Rho coefficients should be considered specially when using MLR and RF regression methods and PCA should be taken into account for a first attempt to reduce collinearity as proposed by Naesset [54] and Cao et al. [9], but afterwards another selection approach should be considered before modeling.

The most selected types of ALS metrics for estimating the seven analyzed stand attributes were higher height variables, variability ones and the ratio between all returns and total returns, while dominant height and variability L moment variables were less demanded. Ho estimation usually

required the inclusion of high height percentiles as concluded by Næsset and Gobakken [17]. V and W estimations normally included either lower or higher height variables, variability metrics, and/or CDM ones as proposed by Silva et al. [86] and Hopkinson et al. [87].

The comparison between regression methods shows that SVMr had the lowest RMSE when estimating the majority of the analyzed stand attributes for both dates, except for Do and Ho when using 2016 data. These results match with García-Gutiérrez et al. [78] and Guerra-Hernández et al. [38,88], which obtained higher accuracies when using non-parametric regression methods. Different results were found in our previous studies [36,89] as MLR slightly outperformed other nonparametric methods when estimating total tree biomass in Aleppo pine forests, but no statistically significant differences were found. Thus, in agreement with Gagliasso et al. [90], a high number of field plots may have boosted machine-learning performance. Furthermore, the broad range and standard deviation values of the field plot data that characterizes Aleppo pine forest at a regional scale is notoriously higher than in our previous studies. Thus, although logarithmic transformation of the dependent and independent variables was carried out, most of the data was not normally distributed. The limitation on using the best-suited ALS metrics, as most of them were not normally distributed, generates a considerable decrease of accuracy in MLR model performance. In this sense, comparison between regression methods is desirable, especially when working with big datasets in heterogeneous forest stands as the Mediterranean ones.

The comparison of direct and indirect approaches allowed us to assess model temporal transferability between 2011 and 2016. The direct approach was computed when extrapolating 2011 models to 2016 and inversely. The indirect approach has shown slightly better results when estimating N, G, Dg, Do, and V. Direct approach showed slightly better results in the estimation of W when extrapolating 2016 model to 2011, but not inversely. Similar results were found for the estimation of Ho when extrapolating 2011 model to 2016 data, but not inversely. Comparisons with previous studies cannot be done for N, Dg, and Do, as these attributes have not been previously estimated using multi-temporal data. Comparisons between Ho, G, and V results are neither possible as Næsset and Gobakken [17] performed only the indirect approach. Regarding W estimations, several results have been obtained using different regression methods, and even in our work both direct and indirect approaches performed in a different way when extrapolating the first-year model (2011) or the second one (2016). The indirect approach obtained better results when estimating W for 2011 data agreeing with Zhao et al. [3] and Meyer et al. [25]. Our results also agree with Cao et al. [9], Skowronski et al. [24], and Bollandsås et al. [26], which detected slightly better performance of direct approach. Similar results have been found in our study when extrapolating 2016 model to 2011. In general, the SVMr regression method shows good temporal transferability between both ALS acquisitions with average %RMSE differences for all the modeled stand attributes of 2.13% for 2011 and 1.58% for 2016.

Models generated using 2016 data (1.25 points m<sup>-2</sup>) showed generally higher accuracy than 2011 ones (0.64 points m<sup>-2</sup>). However, no statistically significant differences were found between the best-fitted models for each year. In agreement with Cao et al. [9], point density may influence model performance but did not significantly affect the estimations of forest attributes as point clouds has a consistent vertical pattern. According to Hudak et al. [27], the relatively large size of the sample plots is considered sufficient for generating canopy height metrics. Thus, the results confirm, as other previous studies based on low-density ALS from the Spanish National Plan for Aerial Orthophotography data (i.e., [33,36–38]), that this information is an accurate and economic alternative to perform forest inventories when higher point density data are not available.

Overall, the use of low-point ALS data for two dates and single-tree growth models for generating temporally-concomitant field data provides accurate estimations of forest stand attributes in Mediterranean Aleppo pine forests. The indirect approach produced higher precision, but the direct approach, within those conditions, may reduce fieldwork and time of model parametrization. When using a direct approach it would not be necessary to create one model for two different points in



time, as it will be possible to extrapolate a model generated for one date (validated with field data) to another date. Furthermore, the number of revisited field plots can be dismissed or even not required, when the time between ALS surveys is not large [28]. This will benefit not only forest managers but also enterprises devoted to forest inventories. The use of direct method and the possibility of model temporal transferability generates new alternatives to calibrate future ALS captures with a lower number of field plots and helps in designing the temporal gap between flights. Single-tree growth models constitute a useful and robust alternative to update field data to a point in time, allowing to accurately estimate forest inventory parameters with the use of ALS data. Future research using multi-temporal ALS data should focus on the inclusion of inference models to better understand uncertainties as well as on the analysis of field plot size and saturation effects in model accuracy. Furthermore, the analysis of forests structural biodiversity changes caused by wildfires or the fusion of ALS data with multi-temporal passive remote sensing series or unmanned aerial vehicle (UAV) point clouds may help to monitor forest dynamics over time.

## 5. Conclusions

Multi-temporal ALS data may improve forest management and planning, providing accurate forest inventory attribute estimations for different points in time. The results illustrate the usefulness of bitemporal low-point density ALS data and single-tree growth models, when lacking temporally-concomitant field campaigns, to accurately estimate seven forestry attributes, using an area-based approach. All subsets regression Seqrep was the most powerful selection method, followed by rho coefficient, to generate parsimonious models. Higher height metrics, canopy height variability, and canopy density variables were the most selected ALS-metrics, while mean height variables and variability L moments were less demanded. The SVM with radial kernel outperformed the analyzed non-parametric and multivariate linear regression methods for estimating all forest inventory attributes except from Do and Ho when using 2016 data. Thus, machine-learning performance may have been boosted by forest heterogeneity and an elevated number of field plots.

This study has assessed model temporal transferability by comparing direct and indirect approaches for the estimation of seven forestry attributes. Indirect approach have produced slightly more accurate results than the direct approach, but average %RMSE differences between both approaches for all modeled stand attributes ranged from 2.13% in 2011 to 1.58% in 2016. Thus, mixing the direct approach with single-tree growth methods provides a suitable alternative to reduce fieldwork and enhance ALS technology as a good tool for estimating forest attributes in two different dates. The utility of multi-temporal ALS data and the combination with multi-temporal series from passive remote sensing and UAV point clouds derived by using photogrammetric techniques would have great value for forest management and planning.

**Author Contributions:** J.d.I.R. and E.R. had the original idea. All co-authors conducted the fieldwork campaigns and D.D., R.A., M.T.L., and A.L.M. developed the methodology. All co-authors performed the analysis and D.D. and R.A. wrote this paper, incorporating suggestions from all co-authors, who approved the final manuscript.

**Funding:** This research is funded by the project SERGISAT [CGL2014-57013-C2-2-R] of the Department of Economy and Competitiveness and partially funded by the FEADER under the provisions of the Rural Development Program of Aragón 2014–2020.

**Acknowledgments:** This work was supported by the Government of Spain, Department of Education Culture and Sports under Grant FPU Grant BOE, 14/06250 and by the Department of Economy, Industry and Competitiveness under the Torres Quevedo Contract PTQ-15-08067. We thank the Spanish National Geographic Information Centre (CNIG) and the Geographic Institute of Aragón (IGEAR) for providing the ALS data. We are also grateful to the material resources provided by the Centro Universitario de la Defensa de Zaragoza (CUD). We express our appreciation to Alberto García-Martín, Jesús Cabrera, and Francisco Escribano for their help during the fieldwork.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Summary of the airborne laser scanning (ALS) computed metrics including the abbreviations, classes, and macro-classes defined.

| Macro-Classes                            | Classes                                     | ALS Computed Metrics   | Abbreviations   |
|--|---|--|---|
| Canopy height metrics (CHM)              | Lower height variables                      | Minimum elevation  | Elev. minimum   |
|  |   | 01th percentile of the return heights                                  | P01   |
|  |   | 05th percentile of the return heights                                  | P05   |
|  |   | 10th percentile of the return heights                                  | P10   |
|  |   | 20th percentile of the return heights                                  | P20   |
|  |   | 25th percentile of the return heights                                  | P25   |
|  |   | L moment 1 elevation   | Elev. L1  |
|  |   | L moment 2 elevation   | Elev. L2  |
|  | Mean height variables                       | Mean elevation   | Elev. Mean  |
|  |   | Mode elevation   | Elev. Mode  |
|  |   | 30th percentile of the return heights                                  | P30   |
|  |   | 40th percentile of the return heights                                  | P40   |
|  |   | 50th percentile of the return heights                                  | P50   |
|  |   | 60th percentile of the return heights                                  | P60   |
|  |   | 70th percentile of the return heights                                  | P70   |
|  |   | L moment 3 elevation   | Elev. L3  |
|  |   | Elevation quadratic mean   | Elev. SQRT mean SQ                                    |
|  |   | Elevation cubic mean   | Elev. CUR mean CUBE                                   |
|  | Higher height variables                     | 75th percentile of the return heights                                  | P75   |
|  |   | 80th percentile of the return heights                                  | P80   |
|  |   | 90th percentile of the return heights                                  | P90   |
|  |   | 95th percentile of the return heights                                  | P95   |
|  |   | 99th percentile of the return heights                                  | P99   |
|  |   | Maximum elevation  | Elev. maximum   |
|  |   | L moment 4 elevation   | Elev. L4  |
| Canopy height variability metrics (CHVM) | Variability                                 | Standard deviation of point heights distribution                       | Elev. SD  |
|  |   | Variance of point heights distribution                                 | Elev. Variance  |
|  |   | Coefficient of variation of point heights distribution                 | Elev. CV  |
|  |   | Skewness of point heights distribution                                 | Elev. Skewness  |
|  |   | kurtosis of point heights distribution                                 | Elev. Kurtosis  |
|  |   | Interquartile distance of point heights distribution                   | Elev. IQ  |
|  |   | Average Absolute Deviation of point heights distribution               | Elev. AAD   |
|  | Variability L moment                        | L moment coefficient of variation of point heights distribution        | Elev. LCV   |
|  |   | L moment skewness of point heights distribution                        | Elev. Lskewness                                       |
|  |   | L moment kurtosis of point heights distribution                        | Elev. Lkurtosis                                       |
| Canopy density metrics (CDM)             | % first, % all returns, canopy relief ratio | percentage of first returns above the 2.00                             | % first ret. above 2.00                               |
|  |   | percentage of all returns above the 2.00                               | % all ret. above 2.00                                 |
|  |   | percentage of first returns above the mean                             | % first ret. above mean                               |
|  |   | percentage of first returns above the mode                             | % first ret. above mode                               |
|  |   | percentage of all returns above the mean                               | % all ret. above mean                                 |
|  |   | percentage of all returns above the mode                               | % all ret. above mode                                 |
|  |   | Canopy relief ratio  | CRR   |
|  | All returns Total returns-1                 | All returns above 2.00 divided by the total first returns $\times 100$ | (All ret. above 2.00)/(total first ret.) $\times 100$ |
|  |   | All returns above mean divided by the total first returns $\times 100$ | (All ret. above mean)/(total first ret.) $\times 100$ |
|  |   | All returns above mode divided by the total first returns $\times 100$ | (All ret. above mode)/(total first ret.) $\times 100$ |



**Table A2.** Summary of the N models using 2011 ALS data. Validation results in terms of RMSE (stems  $\text{ha}^{-1}$ ), %RMSE, and bias (stems  $\text{ha}^{-1}$ ) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to support vector machine with radial kernel; SVM l. refers to support vector machine with linear kernel; ret. refers to returns.

| ALS Metrics   | Model | SM    | Fitting Phase |        |       | Validation |        |        |       |
|---|-------|-------|---------------|--------|-------|------------|--------|--------|-------|
|   |       |       | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias   | $R^2$ |
| P90 + (All ret. above mean)/(total first ret.) $\times 100$                       | MLR   | Step. | 347.22        | 49.08  | 0.00  | 350.67     | 49.57  | 8.76   | 0.53  |
| Elev. L2 + Elev. Variance + (All ret. above 2.00)/(total first ret.) $\times 100$ | MDL   | ASSs  | 235.89        | 33.34  | −0.83 | 292.37     | 41.33  | −1.48  | 0.68  |
| P99 + Elev. IQ + % first ret. above 2.00  | LWLR  | Rho   | 205.80        | 29.09  | −9.79 | 310.97     | 43.96  | −11.39 | 0.65  |
| P99 + Elev. IQ + % first ret. above 2.00  | SVMr  | rho   | 257.09        | 36.34  | 28.81 | 272.76     | 38.55  | 26.99  | 0.72  |
| Elev. L2 + Elev. Variance + (All ret. above 2.00)/(total first ret.) $\times 100$ | SVMl  | ASSs  | 319.34        | 45.14  | 60.68 | 309.56     | 43.76  | 64.83  | 0.65  |
| P99 + Elev. SD + % first ret. above 2.00  | RF    | rho   | 151.86        | 21.46  | 1.91  | 303.56     | 42.91  | 6.91   | 0.66  |

**Table A3.** Summary of the N models using 2016 ALS data. Validation results in terms of RMSE (stems  $\text{ha}^{-1}$ ), %RMSE, and bias (stems  $\text{ha}^{-1}$ ) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to support vector machine with radial kernel; SVM l. refers to support vector machine with linear kernel; ret. refers to returns.

| ALS Metrics  | Model | SM    | Fitting Phase |        |       | Validation |        |        |       |
|--|-------|-------|---------------|--------|-------|------------|--------|--------|-------|
|  |       |       | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias   | $R^2$ |
| Elev. mean + Elev. L kurtosis + Canopy relief ratio        | MLR   | Step. | 358.84        | 51.47  | 0.00  | 363.62     | 52.15  | 11.57  | 0.45  |
| Elev. maximum + Elev. L kurtosis + % first ret. above 2.00 | MDL   | LASSO | 243.13        | 34.87  | 1.14  | 322.89     | 46.31  | 8.15   | 0.61  |
| Elev. maximum + Elev. L kurtosis + % first ret. above 2.00 | LWLR  | LASSO | 204.63        | 29.35  | 4.55  | 333.20     | 47.79  | 11.06  | 0.57  |
| Elev. maximum + Elev. L kurtosis + % first ret. above 2.00 | SVMr  | LASSO | 250.87        | 35.98  | 13.95 | 278.58     | 39.96  | 11.83  | 0.67  |
| Elev. maximum + Elev. L kurtosis + % first ret. above 2.00 | SVMl  | LASSO | 322.11        | 46.20  | 29.31 | 313.41     | 44.95  | 36.04  | 0.59  |
| Elev. maximum + Elev. L kurtosis + % first ret. above 2.00 | RF    | LASSO | 159.15        | 22.83  | −1.71 | 302.57     | 43.40  | −10.81 | 0.60  |

**Table A4.** Summary of the G models using 2011 ALS data. Validation results in terms of RMSE ( $\text{m}^2 \text{ha}^{-1}$ ), %RMSE, and bias ( $\text{m}^2 \text{ha}^{-1}$ ) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to support vector machine with radial kernel; SVM l. refers to support vector machine with linear kernel; ret. refers to returns.

| ALS Metrics  | Model | SM   | Fitting Phase |        |       | Validation |        |       |       |
|--|-------|------|---------------|--------|-------|------------|--------|-------|-------|
|  |       |      | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | $R^2$ |
| Elev. minimum + Elev. Kurtosis + (All ret. above mode)/(total first ret.) $\times 100$ | MLR   | rho  | 5.80          | 29.80  | 0.00  | 6.01       | 30.89  | 0.19  | 0.64  |
| P10 + % first ret. above 2.00  | MDL   | rho  | 4.61          | 23.69  | 0.21  | 5.23       | 26.85  | 0.38  | 0.74  |
| P05 + % first ret. above mean  | LWLR  | ASSe | 4.07          | 20.92  | 0.01  | 5.53       | 28.42  | 0.12  | 0.70  |
| Elev. minimum + Elev. Kurtosis + (All ret. above mode)/(total first ret.) $\times 100$ | SVMr  | ASSs | 4.43          | 22.77  | −0.10 | 4.77       | 24.51  | −0.10 | 0.77  |
| Elev. minimum + Elev. Kurtosis + (All ret. above mode)/(total first ret.) $\times 100$ | SVMl  | ASSs | 4.85          | 24.92  | 0.10  | 4.87       | 25.05  | 0.05  | 0.75  |
| P10 + % first ret. above 2.00  | RF    | rho  | 2.61          | 13.41  | 0.02  | 5.19       | 26.69  | 0.06  | 0.73  |

**Table A5.** Summary of the G models using 2016 ALS data. Validation results in terms of RMSE ( $\text{m}^2 \text{ha}^{-1}$ ), %RMSE, and bias ( $\text{m}^2 \text{ha}^{-1}$ ) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to support vector machine with radial kernel; SVM l. refers to support vector machine with linear kernel; ret. refers to returns.

| ALS Metrics   | Model | SM   | Fitting Phase |        |      | Validation |        |       |       |
|---|-------|------|---------------|--------|------|------------|--------|-------|-------|
|   |       |      | RMSE          | % RMSE | Bias | RMSE       | % RMSE | Bias  | $R^2$ |
| Elev. minimum + % all ret. above mode   | MLR   | rho  | 9.27          | 42.11  | 0.00 | 9.19       | 41.76  | 0.21  | 0.15  |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) $\times 100$ | MDL   | ASSe | 3.65          | 16.57  | 0.19 | 4.43       | 20.11  | 0.12  | 0.82  |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) $\times 100$ | LWLR  | ASSe | 2.84          | 12.93  | 0.01 | 5.05       | 22.94  | −0.10 | 0.77  |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) $\times 100$ | SVMr  | ASSe | 3.88          | 17.61  | 0.41 | 4.14       | 18.80  | 0.30  | 0.84  |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) $\times 100$ | SVMl  | ASSe | 4.38          | 19.89  | 0.44 | 4.43       | 20.12  | 0.35  | 0.81  |
| P10 + Elev. minimum + % first ret. above mean                                     | RF    | ASSf | 2.32          | 10.56  | 0.08 | 4.64       | 21.06  | 0.37  | 0.81  |

**Table A6.** Summary of the Dg models using 2011 ALS data. Validation results in terms of RMSE (cm), %RMSE, and bias (cm) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to Support Vector Machine with radial kernel; SVM l. refers to Support Vector Machine with linear kernel; ret. refers to returns.

| ALS Metrics   | Model | SM   | Fitting Phase |        |       | Validation |        |       |       |
|---|-------|------|---------------|--------|-------|------------|--------|-------|-------|
|   |       |      | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | $R^2$ |
| P90 + % first ret. above mean   | MLR   | rho  | 3.80          | 18.44  | 0.00  | 3.84       | 18.60  | −0.03 | 0.77  |
| P90 + (All ret. above 2.00)/(total first ret.) $\times 100$                 | MDL   | ASSs | 3.30          | 15.97  | 0.16  | 3.78       | 18.34  | 0.00  | 0.78  |
| P90 + (All ret. above 2.00)/(total first ret.) $\times 100$                 | LWLR  | ASSs | 2.98          | 14.46  | −0.02 | 3.75       | 18.19  | −0.22 | 0.78  |
| P90 + Elev. Std.dev + % first ret. above mean                               | SVMr  | rho  | 3.38          | 16.38  | 0.19  | 3.56       | 17.25  | 0.06  | 0.81  |
| P90 + Elev. Std.dev + % first ret. above mean                               | SVMl  | rho  | 3.76          | 18.24  | 0.08  | 3.88       | 18.80  | −0.06 | 0.78  |
| P90 + Elev. Std.dev + (All ret. above mean)/(total first ret.) $\times 100$ | RF    | rho  | 1.89          | 9.17   | −0.02 | 3.75       | 18.18  | −0.04 | 0.79  |

**Table A7.** Summary of the Dg models using 2016 ALS data. Validation results in terms of RMSE (cm), %RMSE, and bias (cm) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to Support Vector Machine with radial kernel; SVM l. refers to Support Vector Machine with linear kernel; ret. refers to returns.

| ALS Metrics  | Model | SM    | Fitting Phase |        |       | Validation |        |       |       |
|--|-------|-------|---------------|--------|-------|------------|--------|-------|-------|
|  |       |       | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | $R^2$ |
| P90 + Elev. LCV + (All ret. above mean)/(total first ret.) $\times 100$          | MLR   | Step. | 3.48          | 15.53  | 0.00  | 3.63       | 16.22  | −0.07 | 0.82  |
| Elev. maximum + Elev. IQ + (All ret. above 2.00)/(total first ret.) $\times 100$ | MDL   | ASSf  | 3.12          | 13.93  | 0.23  | 3.71       | 16.58  | 0.21  | 0.82  |
| P90 + Elev. LCV + (All ret. above mean)/(total first ret.) $\times 100$          | LWLR  | Step. | 2.59          | 11.58  | −0.03 | 3.91       | 17.48  | −0.04 | 0.80  |
| Elev. maximum + Elev. IQ + (All ret. above 2.00)/(total first ret.) $\times 100$ | SVMr  | ASSf  | 3.03          | 13.53  | 0.21  | 3.42       | 15.28  | 0.11  | 0.85  |
| P90 + Elev. mode + % first ret. above mode                                       | SVMl  | ASSs  | 3.45          | 15.40  | 0.20  | 3.57       | 15.95  | 0.11  | 0.83  |
| P90 + Elev. LCV + (All ret. above mean)/(total first ret.) $\times 100$          | RF    | rho   | 1.65          | 7.39   | 0.01  | 3.59       | 16.05  | 0.05  | 0.82  |

**Table A8.** Summary of the Do models using 2011 ALS data. Validation results in terms of RMSE (cm), %RMSE, and bias (cm) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to Support Vector Machine with radial kernel; SVM l. refers to Support Vector Machine with linear kernel; ret. refers to returns.

| ALS Metrics  | Model | SM   | Fitting Phase |        |       | Validation |        |       |       |
|--|-------|------|---------------|--------|-------|------------|--------|-------|-------|
|  |       |      | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | $R^2$ |
| Elev. skewness + Elev. Lkurtosis + P25                     | MLR   | rho  | 5.34          | 20.17  | 0.00  | 5.42       | 20.47  | 0.07  | 0.62  |
| P90+ (All ret. above 2.00)/(total first ret.) $\times$ 100 | MDL   | ASSf | 3.99          | 15.07  | 0.10  | 4.27       | 16.12  | 0.11  | 0.77  |
| P90+ (All ret. above 2.00)/(total first ret.) $\times$ 100 | LWLR  | ASSf | 3.49          | 13.17  | −0.03 | 4.29       | 16.19  | −0.05 | 0.77  |
| P90+ (All ret. above 2.00)/(total first ret.) $\times$ 100 | SVMr  | ASSf | 4.11          | 15.53  | 0.19  | 4.07       | 15.36  | 0.11  | 0.79  |
| P90+ (All ret. above 2.00)/(total first ret.) $\times$ 100 | SVMl  | ASSf | 4.24          | 16.01  | 0.22  | 4.20       | 15.85  | 0.11  | 0.78  |
| P90 + % first ret. above mean                              | RF    | rho  | 2.13          | 8.04   | −0.11 | 4.38       | 16.55  | −0.36 | 0.76  |

**Table A9.** Summary of the Do models using 2016 ALS data. Validation results in terms of RMSE (cm), %RMSE, and bias (cm) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to support vector machine with radial kernel; SVM l. refers to support vector machine with linear kernel; ret. refers to returns.

| ALS Metrics  | Model | SM    | Fitting Phase |        |       | Validation |        |       |       |
|--|-------|-------|---------------|--------|-------|------------|--------|-------|-------|
|  |       |       | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | $R^2$ |
| P90 + Elev. CV + (All ret. above mean)/(total first ret.) $\times$ 100 | MLR   | Step. | 3.53          | 12.33  | 0.00  | 3.63       | 12.68  | −0.08 | 0.85  |
| P90 + Elev. variance + Elev. L2  | MDL   | ASSs  | 3.26          | 11.40  | 0.18  | 3.47       | 12.13  | 0.23  | 0.86  |
| P95 + Elev. CV   | LWLR  | ASSs  | 2.88          | 10.07  | −0.03 | 3.63       | 12.70  | −0.09 | 0.84  |
| P95 + Elev. CV   | SVMr  | ASSs  | 3.25          | 11.35  | 0.40  | 3.40       | 11.89  | 0.33  | 0.87  |
| Elev. Std.dev + Elev. Variance + P05                                   | SVMl  | ASSe  | 3.26          | 11.40  | 0.18  | 3.36       | 11.75  | 0.11  | 0.87  |
| P95 + Elev. CV   | RF    | ASSs  | 1.61          | 5.64   | −0.02 | 3.62       | 12.64  | 0.00  | 0.85  |

**Table A10.** Summary of the Ho models using 2011 ALS data. Validation results in terms of RMSE (m), %RMSE, and bias (m) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to Support Vector Machine with radial kernel; SVM l. refers to Support Vector Machine with linear kernel; ret. refers to returns.

| ALS Metrics                                  | Model | SM    | Fitting Phase |        |       | Validation |        |       |       |
|--|-------|-------|---------------|--------|-------|------------|--------|-------|-------|
|  |       |       | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | $R^2$ |
| Elev. LCV + Elev. Lkurtosis + P01            | MLR   | rho   | 2.21          | 20.24  | 0.00  | 2.26       | 20.71  | 0.05  | 0.63  |
| P90 + Elev. kurtosis                         | MDL   | Step. | 1.24          | 11.36  | 0.10  | 1.44       | 13.18  | 0.08  | 0.85  |
| P90 + Elev. skewness                         | LWLR  | ASSs  | 1.16          | 10.69  | −0.02 | 1.40       | 12.83  | 0.01  | 0.86  |
| P90 + Elev. variance + % All ret. above mean | SVMr  | ASSf  | 1.32          | 12.11  | 0.11  | 1.34       | 12.30  | 0.09  | 0.87  |
| Elev. L1 + Elev. maximum                     | SVMl  | LASSO | 1.42          | 12.99  | 0.09  | 1.40       | 12.82  | 0.05  | 0.86  |
| P90 + Canopy relief ratio                    | RF    | Step. | 0.72          | 6.65   | −0.01 | 1.46       | 13.41  | −0.06 | 0.84  |

**Table A11.** Summary of the Ho models using 2016 ALS data. Validation results in terms of RMSE (m), %RMSE, and bias (m) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to Support Vector Machine with radial kernel; SVM l. refers to Support Vector Machine with linear kernel; ret. refers to returns.

| ALS Metrics  | Model | SM   | Fitting Phase |        |       | Validation |        |       |       |
|--|-------|------|---------------|--------|-------|------------|--------|-------|-------|
|  |       |      | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | $R^2$ |
| Elev. minimum + Elev. CV + Canopy relief ratio                               | MLR   | rho  | 2.41          | 20.90  | 0.00  | 2.52       | 21.91  | 0.01  | 0.51  |
| P95 + Elev. Std.dev  | MDL   | ASSs | 0.92          | 7.99   | 0.07  | 0.95       | 8.27   | 0.07  | 0.93  |
| P95 + Elev. variance   | LWLR  | ASSs | 0.79          | 6.87   | 0.02  | 0.98       | 8.48   | 0.00  | 0.93  |
| P95 + Elev. Std.dev  | SVMr  | ASSs | 0.86          | 7.48   | 0.03  | 1.02       | 8.83   | 0.03  | 0.92  |
| P90 + Elev. variance + Elev. SQRT mean SQ                                    | SVMl  | ASSb | 0.96          | 8.30   | 0.12  | 1.00       | 8.69   | 0.08  | 0.93  |
| P95 + Elev. variance + (All ret. above mean)/(total first ret.) $\times 100$ | RF    | rho  | 0.43          | 3.76   | −0.01 | 1.00       | 8.72   | −0.03 | 0.92  |

**Table A12.** Summary of the V models using 2011 ALS data. Validation results in terms of RMSE ( $m^3 ha^{-1}$ ), %RMSE, and bias ( $m^3 ha^{-1}$ ) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to support vector machine with radial kernel; SVM l. refers to support vector machine with linear kernel; ret. refers to returns.

| ALS Metrics  | Model | SM    | Fitting Phase |        |      | Validation |        |       |       |
|--|-------|-------|---------------|--------|------|------------|--------|-------|-------|
|  |       |       | RMSE          | % RMSE | Bias | RMSE       | % RMSE | Bias  | $R^2$ |
| P20 + (All ret. above 2.00)/(total first ret.) $\times 100$                    | MDL   | ASSs  | 30.15         | 28.63  | 1.69 | 33.39      | 31.71  | 2.13  | 0.81  |
| P20 + (All ret. above 2.00)/(total first ret.) $\times 100$                    | LWLR  | ASSs  | 25.89         | 24.58  | 0.05 | 34.09      | 32.37  | 0.24  | 0.80  |
| Elev. L2 + Elev. CUR mean CUBE + % first ret. above mean                       | SVMr  | Step. | 28.87         | 27.42  | 2.59 | 29.71      | 28.22  | 1.79  | 0.84  |
| P20 + Elev. L skewness + (All ret. above mean)/(total first ret.) $\times 100$ | SVMl  | ASSs  | 34.25         | 32.52  | 0.88 | 34.30      | 32.58  | 0.09  | 0.79  |
| P20 + Elev. L skewness + % first ret. above 2.00                               | RF    | ASSs  | 16.80         | 15.96  | 0.17 | 34.28      | 32.55  | −0.56 | 0.78  |

**Table A13.** Summary of the V models using 2016 ALS data. Validation results in terms of RMSE ( $m^3 ha^{-1}$ ), %RMSE, and bias ( $m^3 ha^{-1}$ ) and  $R^2$ . SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to support vector machine with radial kernel; SVM l. refers to support vector machine with linear kernel; ret. refers to returns.

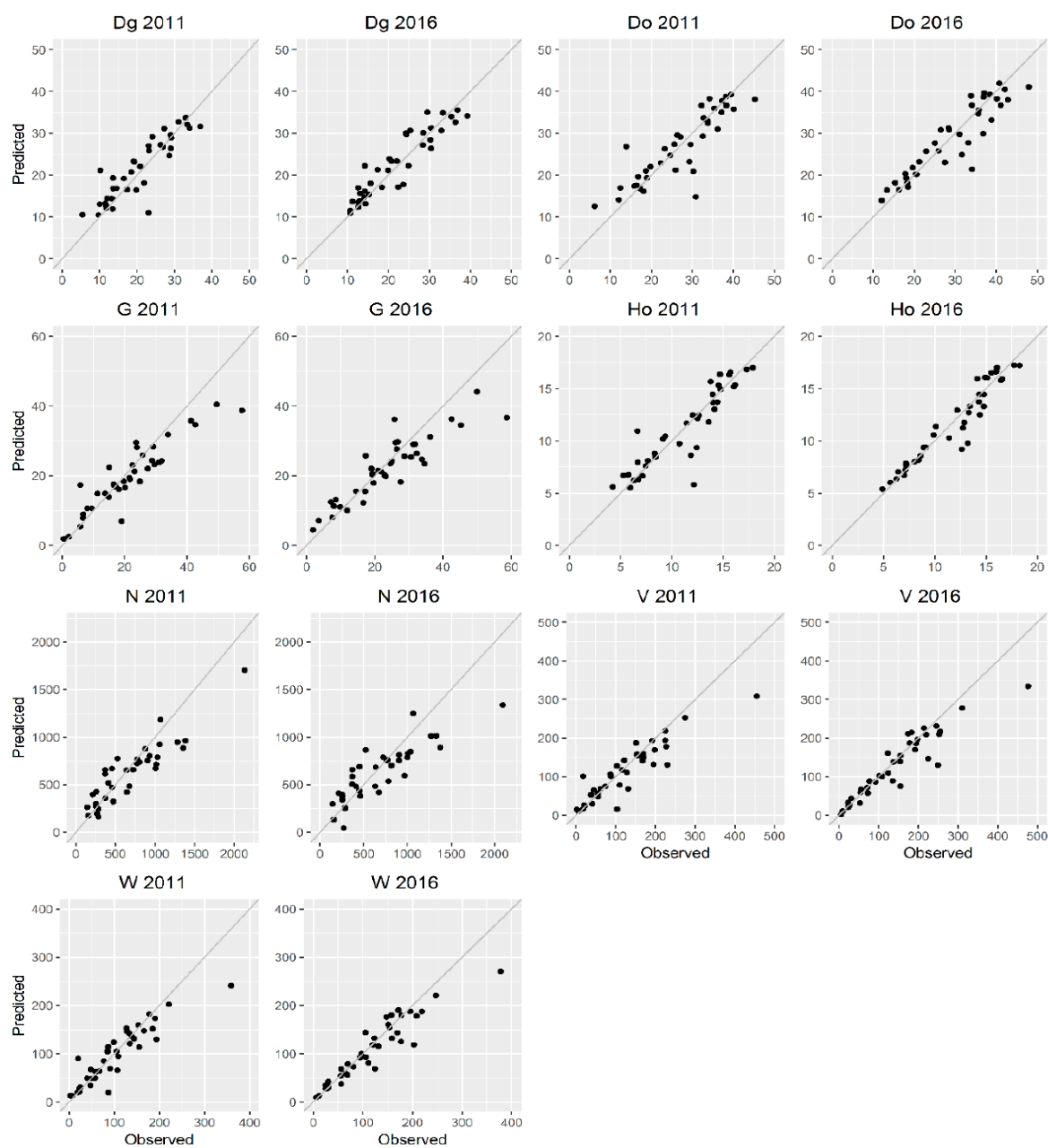
| ALS Metrics   | Model | SM    | Fitting Phase |        |       | Validation |        |       |       |
|---|-------|-------|---------------|--------|-------|------------|--------|-------|-------|
|   |       |       | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | $R^2$ |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) $\times 100$ | MDL   | ASSe  | 24.87         | 20.02  | −0.34 | 29.63      | 23.85  | −0.19 | 0.88  |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) $\times 100$ | LWLR  | ASSe  | 20.26         | 16.30  | −0.08 | 31.80      | 25.59  | −0.06 | 0.85  |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) $\times 100$ | SVMr  | ASSe  | 24.69         | 19.87  | 2.65  | 26.35      | 21.20  | 1.92  | 0.90  |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) $\times 100$ | SVMl  | ASSe  | 30.49         | 24.54  | 2.60  | 31.14      | 25.06  | 1.48  | 0.86  |
| Elev. L2 + Elev. CUR mean CUBE + % first ret. above 2.00                          | RF    | Step. | 15.25         | 12.27  | −0.38 | 31.73      | 25.53  | 0.32  | 0.86  |

**Table A14.** Summary of the W models using 2011 ALS data. Validation results in terms of RMSE (tons ha<sup>-1</sup>), %RMSE, and bias (tons ha<sup>-1</sup>) and R<sup>2</sup>. SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to support vector machine with radial kernel; SVM l. refers to support vector machine with linear kernel; ret. refers to returns.

| ALS Metrics  | Model | SM    | Fitting Phase |        |       | Validation |        |       |                |
|--|-------|-------|---------------|--------|-------|------------|--------|-------|----------------|
|  |       |       | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | R <sup>2</sup> |
| Elev. L2 + Elev. CUR mean CUBE + % first ret. above 2.00                   | MDL   | Step. | 22.02         | 24.21  | 0.68  | 28.30      | 31.12  | 1.82  | 0.76           |
| P10 + Elev. CUR mean CUBE + % first ret. above mean                        | LWLR  | rho   | 18.34         | 20.17  | 0.23  | 29.12      | 32.02  | 0.04  | 0.74           |
| P10 + Elev. SQRT mean SQ + (All ret. above mean)/(total first ret.) × 100  | SVMr  | rho   | 23.00         | 25.29  | 0.75  | 24.29      | 26.71  | −0.03 | 0.82           |
| P10 + Canopy relief ratio + (All ret. above mean)/(total first ret.) × 100 | SVMl  | ASSf  | 26.60         | 29.25  | 0.50  | 26.82      | 29.49  | 0.11  | 0.79           |
| P10 + Elev. CUR mean CUBE + (All ret. above mean)/(total first ret.) × 100 | RF    | rho   | 14.39         | 15.83  | −0.05 | 29.43      | 32.36  | 0.24  | 0.75           |

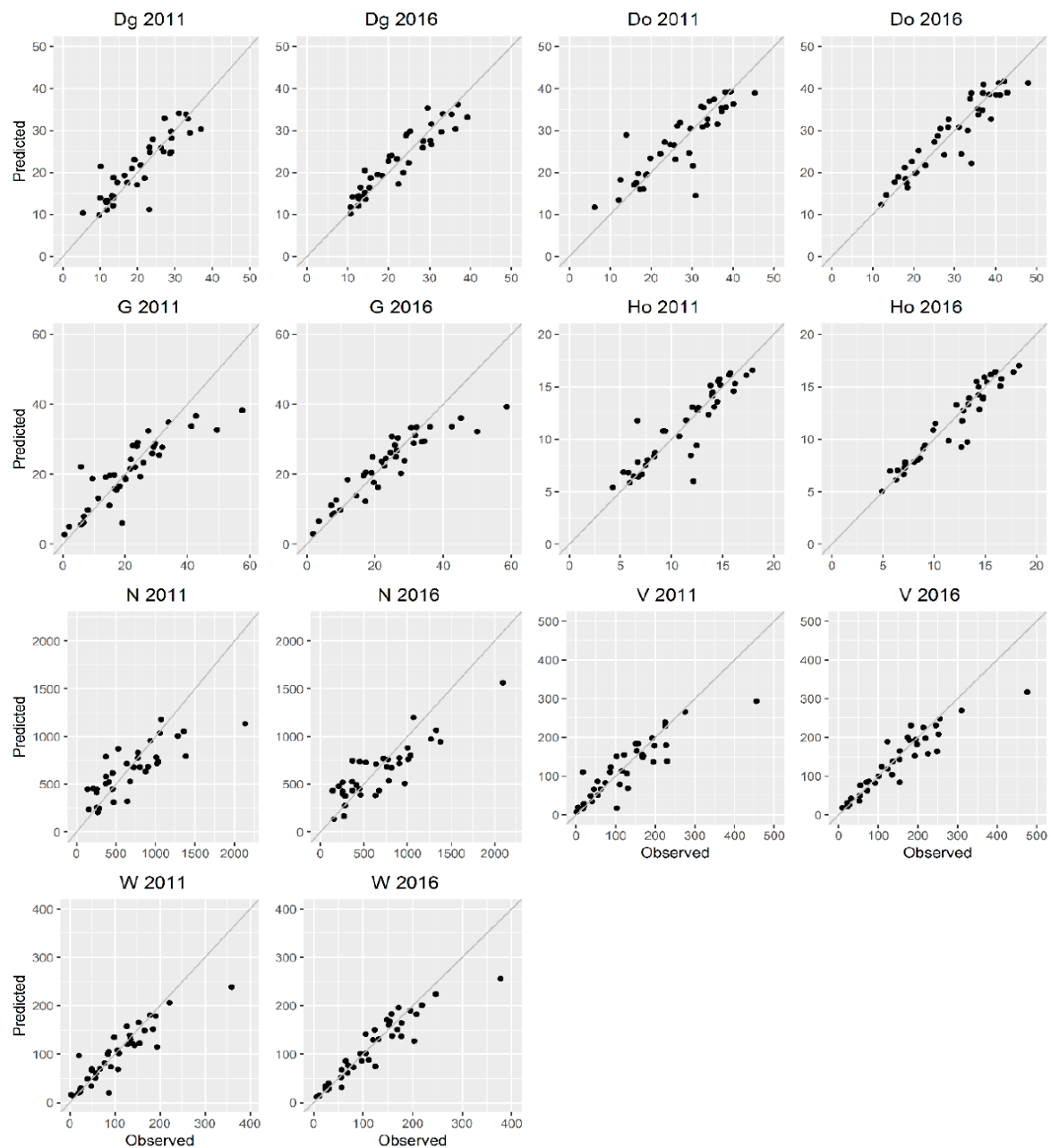
**Table A15.** Summary of the W models using 2016 ALS data. Validation results in terms of RMSE (tons ha<sup>-1</sup>), %RMSE, and bias (tons ha<sup>-1</sup>) and R<sup>2</sup>. SM refers to selection method; Step. stands for Stepwise both and forward; SVMr. refers to support vector machine with radial kernel; SVM l. refers to support vector machine with linear kernel; ret. refers to returns.

| ALS Metrics  | Model | SM    | Fitting Phase |        |       | Validation |        |       |                |
|--|-------|-------|---------------|--------|-------|------------|--------|-------|----------------|
|  |       |       | RMSE          | % RMSE | Bias  | RMSE       | % RMSE | Bias  | R <sup>2</sup> |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) × 100 | MDL   | ASSe  | 19.66         | 18.44  | −0.64 | 23.44      | 21.98  | −0.43 | 0.88           |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) × 100 | LWLR  | ASSe  | 16.11         | 15.11  | −0.11 | 25.75      | 24.15  | 0.18  | 0.85           |
| Elev. L2 + Elev. CUR mean CUBE + % first ret. above 2.00                   | SVMr  | Step. | 18.82         | 17.65  | 1.23  | 20.06      | 18.81  | 0.56  | 0.90           |
| P75 + Elev. CUR mean CUBE + (All ret. above 2.00)/(total first ret.) × 100 | SVMl  | ASSe  | 22.78         | 21.37  | 1.65  | 23.43      | 21.98  | 0.80  | 0.87           |
| P20 + Elev. CUR mean CUBE + All ret. above 2.00)/(total first ret.) × 100  | RF    | Step. | 12.58         | 11.80  | 0.13  | 22.38      | 20.99  | 0.01  | 0.87           |



**Figure A1.** Scatterplot of predicted vs. observed values of the forest stand variables using the best-selected SVM with radial kernel 2011 models and 2016 extrapolated models. Dg and Do are expressed in cm; G is expressed in  $\text{m}^2 \text{ha}^{-1}$ ; H is expressed in m; N is expressed in stems  $\text{ha}^{-1}$ ; V is expressed in  $\text{m}^3 \text{ha}^{-1}$ ; W is expressed in tons  $\text{ha}^{-1}$ .





**Figure A2.** Scatterplot of predicted vs. observed values of the forest stand variables using the best-selected SVM with radial kernel 2016 models and 2011 extrapolated models. Dg and Do are expressed in cm; G is expressed in  $\text{m}^2 \text{ha}^{-1}$ ; H is expressed in m; N is expressed in stems  $\text{ha}^{-1}$ ; V is expressed in  $\text{m}^3 \text{ha}^{-1}$ ; W is expressed in tons  $\text{ha}^{-1}$ .

# References

1. Lal, R. Sequestration of atmospheric CO<sub>2</sub> in global carbon pools. *Energy Environ. Sci.* **2008**, *1*, 86–100. [[CrossRef](#)]
2. Pan, Y.; Birdsey, R.A.; Fang, J.; Houghton, R.; Kauppi, P.E.; Kurz, W.A.; Phillips, O.L.; Shvidenko, A.; Lewis, S.L.; Canadell, J.G.; et al. A large and persistent carbon sink in the world's forests. *Science* **2011**, *333*, 988–993. [[CrossRef](#)] [[PubMed](#)]
3. Zhao, K.; Suarez, J.C.; Garcia, M.; Hu, T.; Wang, C.; Londo, A. Utility of multitemporal lidar for forest and carbon monitoring: Tree growth, biomass dynamics, and carbon flux. *Remote Sens. Environ.* **2018**, *204*, 883–897. [[CrossRef](#)]
4. Zald, H.S.J.; Wulder, M.A.; White, J.C.; Hilker, T.; Hermosilla, T.; Hobart, G.W.; Coops, N.C. Integrating Landsat pixel composites and change metrics with lidar plots to predictively map forest structure and aboveground biomass in Saskatchewan, Canada. *Remote Sens. Environ.* **2016**, *176*, 188–201. [[CrossRef](#)]
5. Pflugmacher, D.; Cohen, W.B.; Kennedy, R.E.; Yang, Z. Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics. *Remote Sens. Environ.* **2014**, *151*, 124–137. [[CrossRef](#)]
6. Ferraz, A.; Saatchi, S.; Bormann, K.J.; Painter, T.H. Fusion of NASA Airborne Snow Observatory (ASO) Lidar Time Series over Mountain Forest Landscapes. *Remote Sens.* **2018**, *10*, 164. [[CrossRef](#)]
7. Hummel, S.; Hudak, A.T.; Uebler, E.H.; Falkowski, M.J.; Megown, K.A. A Comparison of Accuracy and Cost of LiDAR versus Stand Exam Data for Landscape Management on the Malheur National Forest. *J. For.* **2011**, *109*, 267–273.
8. Hudak, A.T.; Evans, J.S.; Stuart Smith, A.M. LiDAR Utility for Natural Resource Managers. *Remote Sens.* **2009**, *1*, 934–951. [[CrossRef](#)]
9. Cao, L.; Coops, N.C.; Innes, J.L.; Sheppard, S.R.J.; Fu, L.; Ruan, H.; She, G. Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data. *Remote Sens. Environ.* **2016**, *178*, 158–171. [[CrossRef](#)]
10. Dubayah, R.O.; Sheldon, S.L.; Clark, D.B.; Hofton, M.A.; Blair, J.B.; Hurtt, G.C.; Chazdon, R.L. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *J. Geophys. Res. Biogeosci.* **2010**, *115*, 1–17. [[CrossRef](#)]
11. Stoker, J.M.; Harding, D.; Parrish, J. The need for a national LIDAR dataset. *Photogramm. Eng. Remote Sensing* **2008**, *74*, 1066–1068.
12. Marinelli, D.; Paris, C.; Bruzzone, L. A Novel Approach to 3-D Change Detection in Multitemporal LiDAR Data Acquired in Forest Areas. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3030–3046. [[CrossRef](#)]
13. Fekety, P.A.; Falkowski, M.J.; Hudak, A.T. Temporal transferability of LiDAR-based imputation of forest inventory attributes. *Can. J. Forest Res.* **2015**, *45*, 422–435. [[CrossRef](#)]
14. Socha, J.; Pierzchalski, M.; Bałazy, R.; Ciesielski, M. Modelling top height growth and site index using repeated laser scanning data. *For. Ecol. Manag.* **2017**, *406*, 307–317. [[CrossRef](#)]
15. Gatzliolis, D.; Fried, J.S.; Monleon, V.S. Challenges to Estimating Tree Height via LiDAR in Closed-Canopy Forests: A Parable from Western Oregon. *For. Sci.* **2010**, *56*, 139–155.
16. Hopkinson, C.; Chasmer, L.; Hall, R.J. The uncertainty in conifer plantation growth prediction from multi-temporal lidar datasets. *Remote Sens. Environ.* **2008**, *112*, 1168–1180. [[CrossRef](#)]
17. Næsset, E.; Gobakken, T. Estimating forest growth using canopy metrics derived from airborne laser scanner data. *Remote Sens. Environ.* **2005**, *96*, 453–465. [[CrossRef](#)]
18. Yu, X.; Hyypä, J.; Kaartinen, H.; Maltamo, M.; Hyypä, H. Obtaining plotwise mean height and volume growth in boreal forests using multi-temporal laser surveys and various change detection techniques. *Int. J. Remote Sens.* **2008**, *29*, 1367–1386. [[CrossRef](#)]
19. Yu, X.; Hyypä, J.; Hyypä, H.; Maltamo, M. Effects of flight altitude on tree height estimation using airborne laser scanning. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2004**, XXXVI, 96–101.
20. St-Onge, B.; Vepakomma, U. Assessing forest gap dynamics and growth using multi-temporal laser-scanner data. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2004**, XXXVI, Part 1, 173–178.
21. Poudel, K.P.; Flewelling, J.W.; Temesgen, H. Predicting Volume and Biomass Change from Multi-Temporal Lidar Sampling and Remeasured Field Inventory Data in Panther Creek Watershed, Oregon, USA. *Forests* **2018**, *9*, 28. [[CrossRef](#)]

22. Temesgen, H.; Strunk, J.; Andersen, H.E.; Flewelling, J. Evaluating different models to predict biomass increment from multi-temporal lidar sampling and remeasured field inventory data in south-central Alaska. *Math. Comput. For. Nat. Res. Sci.* **2015**, *7*, 66–80.
23. Réjou-Méchain, M.; Tymen, B.; Blanc, L.; Fauset, S.; Feldpausch, T.R.; Monteagudo, A.; Phillips, O.L.; Richard, H.; Chave, J. Using repeated small-footprint LiDAR acquisitions to infer spatial and temporal variations of a high-biomass Neotropical forest. *Remote Sens. Environ.* **2015**, *169*, 93–101. [[CrossRef](#)]
24. Skowronski, N.S.; Clark, K.L.; Gallagher, M.; Birdsey, R.A.; Hom, J.L. Airborne laser scanner-assisted estimation of aboveground biomass change in a temperate oak–pine forest. *Remote Sens. Environ.* **2014**, *151*, 166–174. [[CrossRef](#)]
25. Meyer, V.; Saatchi, S.S.; Chave, J.; Dalling, J.W.; Bohlman, S.; Fricker, G.A.; Robinson, C.; Neumann, M.; Hubbell, S. Detecting tropical forest biomass dynamics from repeated airborne lidar measurements. *Biogeosciences* **2013**, *10*, 5421–5438. [[CrossRef](#)]
26. Bollandsås, O.M.; Gregoire, T.G.; Næsset, E.; Øyen, B.H. Detection of biomass change in a Norwegian mountain forest area using small footprint airborne laser scanner data. *Stat. Methods Appl.* **2013**, *22*, 113–129. [[CrossRef](#)]
27. Hudak, A.T.; Strand, E.K.; Vierling, L.A.; Byrne, J.C.; Eitel, J.U.H.; Martinuzzi, S.; Falkowski, M.J. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sens. Environ.* **2012**, *123*, 25–40. [[CrossRef](#)]
28. Noordermeer, L.; Bollandsås, O.M.; Gobakken, T.; Næsset, E. Direct and indirect site index determination for Norway spruce and Scots pine using bitemporal airborne laser scanner data. *For. Ecol. Manag.* **2018**, *428*, 104–114. [[CrossRef](#)]
29. Mccarley, T.R.; Kolden, C.A.; Vaillant, N.M.; Hudak, A.T.; Smith, A.M.S.; Wing, B.M.; Kellogg, B.S.; Kreidler, J. Multi-temporal LiDAR and Landsat quantification of fire-induced changes to forest structure. *Remote Sens. Environ.* **2017**, *419*–432. [[CrossRef](#)]
30. Vepakomma, U.; Kneeshaw, D.; St-Onge, B. Interactions of multiple disturbances in shaping boreal forest dynamics: A spatially explicit analysis using multi-temporal lidar data and high-resolution imagery. *J. Ecol.* **2010**, *98*, 526–539. [[CrossRef](#)]
31. Vepakomma, U.; St-Onge, B.; Kneeshaw, D. Spatially explicit characterization of boreal forest gap dynamics using multi-temporal lidar data. *Remote Sens. Environ.* **2008**, *112*, 2326–2340. [[CrossRef](#)]
32. Solberg, S.; Næsset, E.; Hanssen, K.H.; Christiansen, E. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sens. Environ.* **2006**, *102*, 364–376. [[CrossRef](#)]
33. Guerra-Hernández, J.; Görgens, E.B.; García-Gutiérrez, J.; Rodríguez, L.C.E.; Tomé, M.; González-Ferreiro, E. Comparison of ALS based models for estimating aboveground biomass in three types of Mediterranean forest. *Eur. J. Remote Sens.* **2016**, *49*, 185. [[CrossRef](#)]
34. Mehtatalo, L.; Virolainen, A.; Tuomela, J.; Packalen, P. Estimating Tree Height Distribution Using Low-Density ALS Data With and Without Training Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 1432–1441. [[CrossRef](#)]
35. Shendryk, I.; Hellström, M.; Klemetsson, L.; Kljun, N. Low-Density LiDAR and Optical Imagery for Biomass Estimation over Boreal Forest in Sweden. *Forests* **2014**, *5*, 992–1010. [[CrossRef](#)]
36. Domingo, D.; Lamelas, M.; Montealegre, A.; García-Martín, A.; de la Riva, J.; Domingo, D.; Lamelas, M.T.; Montealegre, A.L.; García-Martín, A.; de la Riva, J. Estimation of Total Biomass in Aleppo Pine Forest Stands Applying Parametric and Nonparametric Methods to Low-Density Airborne Laser Scanning Data. *Forests* **2018**, *9*, 158. [[CrossRef](#)]
37. Montealegre, A.L.; Lamelas, M.T.; de la Riva, J.; García-Martín, A.; Escribano, F. Use of low point density ALS data to estimate stand-level structural variables in Mediterranean Aleppo pine forest. *Forestry* **2016**, *89*, 373–382. [[CrossRef](#)]
38. Garcia-Gutierrez, J.; Gonzalez-Ferreiro, E.; Riquelme-Santos, J.C.; Miranda, D.; Dieguez-Aranda, U.; Navarro-Cerrillo, R.M. Evolutionary feature selection to estimate forest stand variables using LiDAR. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *26*, 119–131. [[CrossRef](#)]
39. Jakubowski, M.K.; Guo, Q.; Kelly, M. Tradeoffs between lidar pulse density and forest measurement accuracy. *Remote Sens. Environ.* **2013**, *130*, 245–253. [[CrossRef](#)]

40. Gobakken, T.; Næsset, E. Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Can. J. For. Res.* **2008**, *38*, 1095–1109. [\[CrossRef\]](#)
41. Lovell, J.L.; Jupp, D.L.B.; Newnham, G.J.; Coops, N.C.; Culvenor, D.S. Simulation study for finding optimal lidar acquisition parameters for forest height retrieval. *For. Ecol. Manag.* **2005**, *214*, 398–412. [\[CrossRef\]](#)
42. Roussel, J.-R.; Caspersen, J.; Béland, M.; Thomas, S.; Achim, A. Removing bias from LiDAR-based estimates of canopy height: Accounting for the effects of pulse density and footprint size. *Remote Sens. Environ.* **2017**, *198*, 1–16. [\[CrossRef\]](#)
43. Garcia, M.; Saatchi, S.; Casas, A.; Koltunov, A.; Ustin, S.; Ramirez, C.; Garcia-Gutierrez, J.; Balzter, H. Quantifying biomass consumption and carbon release from the California Rim fire by integrating airborne LiDAR and Landsat OLI data. *J. Geophys. Res. Biogeosci.* **2017**. [\[CrossRef\]](#)
44. Singh, K.K.; Chen, G.; McCarter, J.B.; Meentemeyer, R.K. Effects of LiDAR point density and landscape context on estimates of urban forest biomass. *ISPRS J. Photogramm. Remote Sens.* **2015**, *101*, 310–322. [\[CrossRef\]](#)
45. Ruiz, L.; Hermosilla, T.; Mauro, F.; Godino, M.; Ruiz, L.A.; Hermosilla, T.; Mauro, F.; Godino, M. Analysis of the Influence of Plot Size and LiDAR Density on Forest Structure Attribute Estimates. *Forests* **2014**, *5*, 936–951. [\[CrossRef\]](#)
46. García, O. Dimensionality reduction in growth models: An example. *FBMIS* **2003**, *1*, 1–15.
47. Thornley, J. Modelling forest ecosystems: The Edinburgh Forest Model. In *Forest Sustainability: Theory and Practice*; CAB International: Wallingford, UK, 2006.
48. Montero, G.; Serrada, R. *La situación de los bosques y el sector forestal en España - ISFE 2013*; Sociedad Española de Ciencias Forestales: Lourizán (Pontevedra), Spain, 2013.
49. Cuadrat, J.M.; Saz, M.A.; Vicente-Serrano, S.M. *Atlas Climático de Aragón*; Gobierno de Aragón: Aragón, Zaragoza, Spain, 2007.
50. Granados, M.E.; Vilagrosa, A.; Chirino, E.; Vallejo, V.R. Reforestation with resprouter species to increase diversity and resilience in Mediterranean pine forests. *For. Ecol. Manag.* **2016**, *362*, 231–240. [\[CrossRef\]](#)
51. Gasque, M.; García-Fayos, P. Interaction between *Stipa tenacissima* and *Pinus halepensis*: Consequences for reforestation and the dynamics of grass steppes in semi-arid Mediterranean areas. *For. Ecol. Manag.* **2004**, *189*, 251–261. [\[CrossRef\]](#)
52. Hernandez-Tecles, E.; Osem, Y.; Alfaro-Sanchez, R.; de las Heras, J. Vegetation structure of planted versus natural Aleppo pine stands along a climatic gradient in Spain. *Ann. For. Sci.* **2015**, *72*, 641–650. [\[CrossRef\]](#)
53. Hair, J.F.; Prentice, E.; Cano, D. *Análisis Multivariante*; Prentice-Hall: Madrid, Spain, 1999; pp. 1–795. ISBN 9788483220351.
54. Næsset, E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* **2002**, *80*, 88–99. [\[CrossRef\]](#)
55. Alonso Ponce SIMANFOR - Sistema de Simulación de Manejo Forestal Sostenible. Gestión Forestal Sostenible 2018. Available online: <http://sostenible.palencia.uva.es/content/simanfor-sistema-de-simulacion-de-manejo-forestal-sostenible> (accessed on 28 September 2018).
56. Bravo, F.; Rodríguez, F.; Ordóñez, C. A web-based application to simulate alternatives for sustainable forest management: SIMANFOR. *For. Syst.* **2012**, *21*, 4. [\[CrossRef\]](#)
57. Schröder, J.; Gadow, K. von Testing a new competition index for Maritime pine in northwestern Spain. *Can. J. For. Res.* **1999**, *29*, 280–283. [\[CrossRef\]](#)
58. Rojo-Alboreca, A.; Cabanillas-Saldaña, A.M.; Barrio-Anta, M.; Notivol-Paño, E.; Gorgoso-Varela, J.J. Site index curves for natural Aleppo pine forests in the central Ebro valley (Spain). *Madera y Bosques* **2017**, *23*, 143. [\[CrossRef\]](#)
59. Ruiz-Peinado, R.; Del Rio, M.; Montero, G. New models for estimating the carbon sink capacity of Spanish softwood species. *For. Syst.* **2011**, *20*, 176. [\[CrossRef\]](#)
60. PNOA Plan Nacional de Ortofotografía Aérea. Available online: <http://pnoa.ign.es/presentacion> (accessed on 12 April 2016).
61. Evans, J.S.; Hudak, A.T. A Multiscale Curvature Algorithm for Classifying Discrete Return LiDAR in Forested Environments. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1029–1038. [\[CrossRef\]](#)
62. Montealegre, A.; Lamelas, M.; Riva, J. Interpolation Routines Assessment in ALS-Derived Digital Elevation Models for Forestry Applications. *Remote Sens.* **2015**, *7*, 8631–8654. [\[CrossRef\]](#)

63. Evans, J.S.; Hudak, A.T.; Faux, R.; Smith, A.M.S. Discrete Return Lidar in Natural Resources: Recommendations for Project Planning, Data Processing, and Deliverables. *Remote Sens.* **2009**, *1*, 776–794. [CrossRef]
64. McGaughey, R. FUSION/LDV: Software for LIDAR data analysis and visualization - V3.10. *USDA For Serv.* **2014**, 1–212.
65. Nilsson, M. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sens. Environ.* **1996**, *56*, 1–7. [CrossRef]
66. Næsset, E.; Økland, T. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sens. Environ.* **2002**, *79*, 105–115. [CrossRef]
67. Darlington, R.B.; Horst, P. Factor Analysis of Data Matrices. *Am. J. Psychol.* **1966**, *79*, 344–346. [CrossRef]
68. Kaiser, H.F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **1958**, *23*, 187–200. [CrossRef]
69. Tibshirani, R. Regression shrinkage and selection via the lasso: A retrospective. *J. R. Stat. Soc. Stat. Methodol. Ser. B* **2011**, *73*, 273–282. [CrossRef]
70. Miller, A.J. Subset Selection in Regression. In *Monographs on Statistics and Applied Probability* 95; Isham, V., Keiding, T., Louis, N., Tibshirani, R.R., Tong, H., Eds.; Chapman & Hall/CRC: New York, NY, USA, 2002; p. 256. Available online: [https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Subset\\_Selection\\_in\\_Multiple\\_Regression.pdf](https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Subset_Selection_in_Multiple_Regression.pdf) (accessed on 23 January 2019).
71. Hermosilla, T.; Ruiz, L.A.; Kazakova, A.N.; Coops, N.C.; Moskal, L.M. Estimation of forest structure and canopy fuel parameters from small-footprint full-waveform LiDAR data. *Int. J. Wildland Fire* **2014**, *23*, 224. [CrossRef]
72. Kumar, L.; Sinha, P.; Taylor, S.; Alqurashi, A.F. Review of the use of remote sensing for biomass estimation to support renewable energy generation. *J. Appl. Remote Sens.* **2015**, *9*, 097696. [CrossRef]
73. Means, J.; Acker, S.; Harding, D.; Blair, J.; Lefsky, M.; Cohen, W.; Harmon, M.; McKee, W. Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western Cascades of Oregon. *Remote Sens. Environ.* **1999**, *67*, 298–308. [CrossRef]
74. García, D.; Godino, M.; Mauro, F. *Lidar: Aplicación Práctica Al Inventario Forestal*; Editorial Academica Española: Madrid, Spain, 2012; p. 196. ISBN 3659009555.
75. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
76. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
77. Van Essen, D.C.; Drury, H.A.; Dickson, J.; Harwell, J.; Hanlon, D.; Anderson, C.H. An Integrated Software Suite for Surface-based Analyses of Cerebral Cortex. *J. Am. Med. Inf. Assoc.* **2001**, *8*, 443–459. [CrossRef]
78. García-Gutiérrez, J.; Martínez-Álvarez, F.; Troncoso, A.; Riquelme, J.C. A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing* **2015**, *167*, 24–31. [CrossRef]
79. Görgens, E.B.; Montagni, A.; Rodríguez, L.C.E. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Comput. Electron. Agric.* **2015**, *116*, 221–227. [CrossRef]
80. Stojanova, D.; Panov, P.; Gjorgjioski, V.; Kobler, A.; Džeroski, S. Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecol. Inf.* **2010**, *5*, 256–266. [CrossRef]
81. Nemenyi, P. Distribution-Free Multiple Comparisons. Ph.D. Thesis, Princeton University, Princeton, NY, USA, 1963.
82. Watt, M.S.; Meredith, A.; Watt, P.; Gunn, A.; Andersen, H.; Reutebuch, S.; Nelson, R. Use of LiDAR to estimate stand characteristics for thinning operations in young Douglas-fir plantations. *N. Z. J. For. Sci.* **2013**, *43*, 1–10. [CrossRef]
83. Amaro, A.; Reed, D.; Soares, P. *Modelling Forest Systems*; CABI publishing: Lisbon, Portugal, 2003; Available online: <https://bit.ly/2FNpTVj> (accessed on 23 January 2019).
84. Hansen, E.; Gobakken, T.; Bollandsås, O.; Zahabu, E.; Næsset, E.; Hansen, E.H.; Gobakken, T.; Bollandsås, O.M.; Zahabu, E.; Næsset, E. Modeling Aboveground Biomass in Dense Tropical Submontane Rainforest Using Airborne Laser Scanner Data. *Remote Sens.* **2015**, *7*, 788–807. [CrossRef]
85. Kristensen, T.; Næsset, E.; Ohlson, M.; Bolstad, P.V.; Kolka, R. Mapping Above- and Below-Ground Carbon Pools in Boreal Forests: The Case for Airborne Lidar. *PLoS ONE* **2015**, *10*. [CrossRef] [PubMed]



86. Silva, C.A.; Klauber, C.; Hudak, A.T.; Vierling, L.A.; Liesenberg, V.; Carvalho, S.P.C.E.; Rodriguez, L.C.E. A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. *Forestry* **2016**, *89*. [[CrossRef](#)]
87. Hopkinson, C.; Chasmer, L.; Barr, A.G.; Kljun, N.; Black, T.A.; McCaughey, J.H. Monitoring boreal forest biomass and carbon storage change by integrating airborne laser scanning, biometry and eddy covariance data. *Remote Sens. Environ.* **2016**, *181*, 82–95. [[CrossRef](#)]
88. Guerra-Hernandez, J.; Gonzalez-Ferreiro, E.; Sarmiento, A.; Silva, J.; Nunes, A.; Correia, A.C.; Fontes, L.; Tomé, M.; Diaz-Varela, R. Short Communication. Using high resolution UAV imagery to estimate tree variables in Pinus pinca plantation in Portugal. *For. Syst.* **2016**, *25*. [[CrossRef](#)]
89. Domingo, D.; Lamelas-Gracia, M.T.; Montealegre-Gracia, A.L.; de la Riva-Fernández, J. Comparison of regression models to estimate biomass losses and CO<sub>2</sub> emissions using low-density airborne laser scanning data in a burnt Aleppo pine forest. *Eur. J. Remote Sens.* **2017**, *50*, 384–396. [[CrossRef](#)]
90. Gagliasso, D.; Hummel, S.; Temesgen, H. A comparison of selected parametric and non-parametric imputation methods for estimating forest biomass and basal area. *Open J. For.* **2014**, *4*, 42–48. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



## 4. Conclusions and future working lines

*The last chapter of this PhD Thesis summarizes the main conclusions of the use of ALS-PNOA data for forest stand variables estimation in Mediterranean environments. The results, presented in the different papers, allow us to synthesize the general and specific conclusions as well as introduce some of the possible future research proposals.*



## 4.1. Main conclusions

Airborne laser scanning is considered one of the best technologies for 3D characterization of vegetation structure and fine-scale forestry metrics estimation. The use of low-density ALS-PNOA data in coniferous Mediterranean forest, characterized by a rugged topography and structural heterogeneity, gives value to these public data and provides useful information for forest management at local and regional scales.

This PhD Thesis addressed the estimation of different forest variables and the temporal transferability of models, focusing on the comparison of selection and regression methods, and the effects of ALS and environmental characteristics in error modelling.

The fusion of ALS and passive optical data evidenced the suitability of these data for wildfire emissions quantification. The use of ALS data capabilities to better describe forest structure and its fusion with optical passive data might be considered as suitable tools to enhance greenhouse gases quantification caused by fires, while solving the lack of post-fire structural information. This methodological approach could be also applied when having pre and post-fire ALS data by determining severity indexes.

We have proven the usefulness of low-density ALS-PNOA data for forest variable estimation at local and regional scales in Aleppo pine Mediterranean forests. The model accuracies reached similar values to those proposed in the literature in applications using low-density ALS data. Apart from the estimation of traditional inventory variables, the estimation of total biomass, including shrubs, and different biomass fractions, such as forest residual biomass, improved Mediterranean Aleppo pine forest characterization. The performance in the prediction of dominant height, volume and different biomass fractions achieved higher precision than number of stem predictions. The assessment of total biomass better quantified Aleppo pine forests as carbon stocks. The quantification of forest residual biomass might act as a first step to enhance the use of this renewable energy, whose exploitation has socio-economic benefits and may increase rural development. Furthermore, this forest parameter is related with canopy bulk density and canopy fuel weight, whose control can reduce fire risk in forested Mediterranean ecosystems.

After comparison between direct and indirect approach to assess temporal transferability, it was concluded that there is no perfect approach. The results demonstrate that both approaches were suitable for estimating forest stand variables using multi-temporal ALS data. Direct approach reduces modelling time and fieldwork costs in cases where the time between the ALS surveys is not large. Accordingly, the last mentioned approach benefits not only forest managers but also enterprises devoted to forest inventories and might serve as an alternative to design the temporal gap between ALS flights to predict accurately forest variables over time. This research provides guidance for potential users to select the most favourable approach according to the aim and time between ALS surveys.

The determination of metric suitability for estimating forest variables potentially reduces modelling time and improves model accuracy. In this sense, the proposed methodology compares

five selection processes to determine the most favourable one. In general, the results showed that any of the selection methods is preferred and their performance depends on the estimated forest variable. Nevertheless, this research provides guidance for potential users of the most favourable selection method, according to the analysed variable and conditions.

The generation of parsimonious and understandable models requires the selection of a reduced number of metrics that show a comprehensive relationship with the modelled variable. These requirements were considered for all the generated models in this research, selecting a maximum number of three ALS metrics. The models to estimate forest variables generally included one independent variable related to canopy height, one associated to height variability and one linked to canopy density.

After comparison between regression methods, we concluded that no perfect approach exists, being influenced by sample size and sample heterogeneity. Multivariate linear regression outperformed non-parametric methods in estimating aboveground tree biomass and total biomass with small or medium sample sizes (46 and 83 plots). However, with a bigger sample size (147 and 192 plots) multivariate linear model showed a limited applicability, being SVM the most accurate method. The main restriction for the multivariate linear model was the non-normal distribution of the best-suited ALS metrics.

From the analysis of the effects of ALS characteristics and environmental variables on forest residual biomass prediction accuracy, we concluded that the presence of higher point cloud densities, lower sensor scan angles and higher canopy pulse penetration values increased model performance. On the contrary, slope steepness of the terrain decreased model accuracy, which may be partially explained by the errors in DTMs generation. Finally, the effect of shrub presence was not clear. In this sense, further research could be done to analyse the effect on the prediction of other forest variables and in different forest environments.

This research enriched the knowledge of Mediterranean Aleppo pine forests using ALS-PNOA data and enhanced the utility of these public data, which is of great utility in forest management at local and regional scales.

## 4.2. Specific conclusions

The main conclusions for each of the specific aims detailed in chapter 1 are presented below:

*Explore the usefulness of low-density ALS data to new applications: estimation of biomass losses and CO<sub>2</sub> emissions to atmosphere, quantification and mapping of forest residual biomass and estimation of different tree fractions and shrub fraction of biomass at stand level*

- The estimation of pre-fire above ground tree biomass using ALS data showed a good model performance, allowing an accurate quantification of this forest variable. The models included two ALS-derived metrics: the 40<sup>th</sup> percentile of the return heights, related with canopy height, and the percentage of first returns above 2 m, associated to canopy density.

The determination of fire severity using Landsat 8 dNBR index allowed applying different burning efficiency factors, solving one of the traditional inaccuracies in fire emission estimations: the assumption of total loss of biomass after burning. The determination of three burning efficiency factors associated to three general ranges of dNBR permitted extrapolating the methodology to other environments. Biomass losses accounted for 262,659.7 ton and a total of 426,754.8 ton of CO<sub>2</sub> were emitted by the Aleppo pine forested burned areas. The application of regional conversion factor to estimate carbon content and, subsequently, CO<sub>2</sub> emissions reduces the uncertainties related to the use of general emission equations, while providing an accurate estimation. The three-phase methodological approach provides a suitable option for estimating wildfire biomass losses, minimizing fieldwork, while solving the lack of post-fire structural information.

- Spearman rank's rank determined the more suitable ALS variables to estimate forest residual biomass: the 70<sup>th</sup> percentile of return heights, the elevation variance and the percentage of first returns above mean, representing canopy height, variability and canopy density metrics. The 197,951.24 ha mapped represent 87.6% of the Aleppo pine forested area of Aragón, summing up 3,627,021.25 tons. Low-density ALS-PNOA data provide accurate information for forest managers about green renewable energy resources and canopy fuel distribution at regional scales.
- Three ALS metrics were selected to estimate total biomass, including a metric that characterizes the vertical distribution of ALS returns (the 25<sup>th</sup> percentile of the return heights), a metric related with the variability of the return heights (the elevation variance) and a metric associated with canopy density (the percentage of first returns above mean). Higher canopy density and variability of the point cloud are associated with higher total biomass content. Lower height values of the 25<sup>th</sup> percentile is related with a higher presence of shrubs in the understory and, subsequently, and increase of the total biomass. The presence of shrub in Aleppo pine Mediterranean forest, especially for those semi-natural stands, constitutes a relevant fraction of total biomass. The estimation and cartography of these forest resources using low-density ALS data provided useful information for forest management purposes. Total biomass mapping showed values ranging from less than 10 tons/ha up to 150 tons/ha.

*Explore the temporal transferability of models for estimating forest stands attributes at regional scale using multi-temporal ALS-PNOA data*

- The acquisition of a high number of field plots to support ALS inventories at regional scales is cost and time demanding. In this research, the estimation of stand attributes to generate temporal-concomitant field data for two different points in time (2011 and 2016) were performed using specific single-tree growth models, generalized height-diameter curves and taper equations. This methodology produced more accurate results than using yield tables in tree growth at short term predictions.
- Temporal transferability was performed by using indirect and direct approaches. The indirect approach, consisting on fitting two different models (one for each analysed year),

produced generally higher precision. However, no strong differences in precision with the direct approach were found. Besides, the last mentioned approach reduces fieldwork and time of model parametrization, being possible to extrapolate a model generated for one date (validated with field data) to another date, especially if the time between ALS surveys is not large. These potentialities benefits forest managers and enterprises devoted to forest inventories reducing time and cost when modelling changes in forest stand attributes.

*Compare different parametric and non-parametric methods in forest variable modelling*

- The comparisons between eight regression models to estimate pre-fire aboveground biomass determined that multivariate linear model outperform the two machine learning algorithms and five regression trees structures. The models were developed with 46 plots and two ALS metrics were included. Model accuracy ranged from 11.1 %RMSE, using the multivariate regression model, up to 23.2 %RMSE, using the KNN algorithm. Models were not equivalent according to Friedman test, but only WKNN and KNN presented statistically significant differences respect to the linear, the two machine learning and the three regression trees structures analysed.
- The best regression model to estimate total biomass was the multivariate linear model, outperforming the four non-parametric analysed regression methods. The models were developed with 83 plots and three ALS metrics were included. Model accuracy ranged from 19.21 up to 24.93 %RMSE, using the multivariate linear model and random forest, respectively. In this approach, statistically significant differences were found neither between the analysed methods. Random forest and locally weighted linear regression showed overestimation before performing cross-validation.
- SVM with radial kernel was the most accurate regression method to estimate forest residual biomass at regional scale (26.38 % RMSE). The higher number of plots (192) and higher sample variability might have boosted the performance of non-parametric regression methods over linear ones. While generalization of regional models might generate less accurate results, the use of SVM with radial kernel could be considered when working with large field datasets and ALS data.
- The SVM with radial kernel was also the best method to analyse temporal transferability between 2011 and 2016. Multivariate linear model showed a limited applicability with a high number of plots (147) as the best-suited ALS metrics were not normally distributed. This fact may have caused a decrease in accuracy of model performance. Stem number, basal area, squared mean diameter, dominant diameter, dominant height, volume and aboveground tree biomass showed a %RMSE of 38.55, 24.51, 17.25, 15.36, 28.22 and 26.71, respectively, using ALS data captured in 2011. Generally lower %RMSE were obtained using the ALS data captured in 2016, obtaining values of 39.96, 18.80, 15.28, 11.89, 8.83, 21.20 and 18.81, respectively (statistics are presented in the same order as in 2011).
- Generally, lower accuracy was obtained using SVM with linear kernel, random forest, LWLR and MDL. The use of WKNN and KNN was tested for estimating pre-fire above ground tree biomass, presenting the lowest accuracy and denoting statistically significant



differences respect to the rest of the analysed methods. RF showed an overestimating tendency during the fitting phase, even applying bias correction methods. However, this method provided the best results to estimate dominant height with a size sample of 147. Further tests are required to determine whether large sample sizes improve accuracy in more complex stand variables prediction.

*Assessing the suitability of different variable selection methods in order to improve accuracy following the principle of parsimony*

- The comparison between selection methods determined that all subset regression and selection based on Spearman's rank were the most powerful techniques. LASSO selection, forward Stepwise selection and PCA showed a lower power in determining the best subsets. All Subset Selection, which is an automatic selection method, allows selecting the maximum number of metrics and provides faster results than Spearman's rank, which was applied using "manual expert criteria".
- All Subset Selection was the best method for estimating basal area, dominant diameter, dominant height and volume, while Spearman rank determined the best metrics in estimating forest residual biomass, tree aboveground biomass and stem number. Equal usefulness of both methods was detected in squared mean diameter and total biomass predictions. Stepwise selection was also used for estimating aboveground tree biomass and volume, and LASSO for estimating stem number.
- High percentiles from the return height distribution were usually selected to estimate dominant height. Higher canopy height metrics, variability metrics and canopy density metrics were chosen for estimating stand density, dominant diameter, squared mean diameter and residual biomass. The last three mentioned metric types and lower canopy height metrics were used for estimating basal area, volume, aboveground tree biomass and total biomass.

*Assessing the effect of some ALS parameters and environmental conditions in model performance*

- Densities lower than 1 point  $\text{m}^{-2}$  increases MPE values in 0.56 tons  $\text{ha}^{-1}$ . However, the accuracy remained relatively high using low point densities, confirming that ALS-PNOA information is an accurate alternative to estimate residual biomass when higher point density are not available.
- Scan angles higher than  $15^\circ$  increases MPU values in 1.94 tons  $\text{ha}^{-1}$  compared to angles close to nadir ( $<5^\circ$ ), but no statistically significant differences were found between the analysed ranges. Agreeing with previous studies, data collection should be minimized at scan angles greater than  $\sim 15^\circ$ .
- The decrease of CPP, especially for those areas densely covered by high strata, increases MPE model values up to 5.47 tons  $\text{ha}^{-1}$ , especially with CPP values lower than 25% respect to plots with CPP values higher than 75%. The decrease of CPP implies a reduction of returns number in lower strata and, accordingly, a lower accuracy of DTM.

- The presence of steep slopes decreases model accuracy up to 1.44 tons ha<sup>-1</sup> in comparison with smooth slopes lower than 15%, but do not represent a significant change. The decrease of accuracy may be partially explained by the lower accuracy of DTMs, considering that filtering algorithms have more errors on steep slopes.
- Shrub presence slightly increases MPE values in 0.11 tons ha<sup>-1</sup>. The effect of shrubs in error modelling is not clear and further research is needed to analyse the effect in the prediction of different forest variables.

## 4.1. Conclusiones principales

El escáner láser aeroportado (ALS) es una de las tecnologías más útiles para caracterizar tridimensionalmente la estructura de la vegetación y estimar variables forestales a escala de detalle. El uso de los datos ALS de baja densidad del Plan Nacional de Ortofotografía Aérea (PNOA) en masas forestales de coníferas, en entornos mediterráneos, caracterizados por una topografía compleja y una heterogeneidad estructural, pone en valor estos datos públicos y proporciona información útil para la gestión forestal a escala local y regional.

Esta Tesis analiza la estimación de diversas variables forestales y la transferibilidad temporal de modelos, enfocándose en la comparación de métodos de selección y regresión, así como en el análisis de los efectos que diversas variables ambientales y características ALS tienen en los errores de modelado.

La fusión de datos ALS e imágenes capturadas por sensores óptico-pasivos evidencia la idoneidad de esta información para la cuantificación de las emisiones generadas por un incendio. El uso de los datos ALS para una mejor caracterización de la estructura forestal y su fusión con sensores óptico-pasivos pueden ser de utilidad para la cuantificación de gases de efecto invernadero generados por incendios, a la par que solventar la falta de información estructural post-incendio. Este enfoque metodológico puede aplicarse también en los casos en los que exista disponibilidad de datos ALS pre y post-incendio, mediante la determinación de índices de severidad.

Los datos ALS-PNOA de baja densidad se han mostrado útiles para estimar variables forestales a escalas locales y regionales en masas forestales mediterráneas de Pino Carrasco. Los ajustes de los modelos muestran valores similares a los propuestos en la literatura cuando se utilizan datos ALS de baja densidad. Además de la estimación de variables de inventario forestal tradicional, la estimación de biomasa total, incluyendo el matorral, y diferentes fracciones de biomasa, como la biomasa residual, han permitido mejorar la caracterización de las masas forestales mediterráneas de Pino carrasco. En este sentido, la estimación de la altura dominante, el volumen y diversas fracciones de biomasa presenta mayor precisión que la estimación del número de pies. La estimación de la biomasa total proporciona una mejora en la cuantificación de las reservas de carbono en masas forestales de Pino carrasco. Del mismo modo, la cuantificación de la biomasa residual forestal pone en valor el uso de este recurso de energía renovable, cuya explotación tiene beneficios socio-económicos, que pueden redundar en un mejor desarrollo de las áreas rurales. Además, este parámetro forestal está asociado a la densidad del dosel y el peso de combustible, cuyo control puede minimizar el riesgo de incendios en ecosistemas forestales mediterráneos.

De la evaluación de la transferibilidad de los modelos realizada mediante la comparación del enfoque directo e indirecto se concluye que no existe una metodología perfecta. Los resultados demuestran que ambos enfoques son óptimos para estimar variables forestales utilizando datos ALS multi-temporales. El enfoque directo reduce el tiempo de modelado y los costes asociados al trabajo de campo en aquellos casos en los que el tiempo entre adquisiciones ALS no es amplio. Este enfoque beneficia no sólo a los gestores de los espacios forestales sino también a las empresas

dedicadas a la realización de inventarios forestales, pudiendo también servir como alternativa para diseñar las fechas entre dos vuelos ALS y mantener una precisión adecuada en la estimación de variables forestales. Este estudio proporciona una guía para que los usuarios potenciales seleccionen el enfoque más favorable, acorde a sus objetivos y tiempo transcurrido entre adquisiciones ALS.

La determinación de la idoneidad de las métricas ALS para la estimación de variables forestales, reduce potencialmente el tiempo de modelado, a la par que mejora la precisión de los modelos. La metodología propuesta se basa en la comparación de cinco métodos de selección para determinar las métricas más favorables en la estimación de diferentes variables forestales. En general, los resultados muestran que el método de selección varía con respecto a la variable a estimar. Sin embargo, el análisis realizado proporciona una guía para usuarios potenciales sobre la utilización de métodos de selección más favorables, considerando la variable analizada y las condiciones ambientales.

La generación de modelos parsimoniosos requiere la selección de un número reducido de métricas ALS que muestren una relación comprensible con la variable modelada. Este supuesto fue considerado para el conjunto de modelos generados a lo largo de la Tesis, seleccionando un número máximo de tres métricas ALS. Los modelos de estimación de variables forestales generalmente incluyen una variable asociada a la altura del dosel, una relacionada con la variabilidad del dosel y una descriptora de la densidad del dosel.

La comparación entre métodos de regresión determinó que no existe un tipo de modelo que funcione mejor en todas las situaciones, estando influenciado por el tamaño de la muestra y su heterogeneidad. En este sentido, los modelos de regresión linear multivariante generaron mejores resultados que los no-paramétricos cuando se estimó la biomasa arbórea y la biomasa total, utilizando un número de parcelas de campo pequeño o medio (46 a 83 parcelas de campo). Por el contrario, cuando la muestra de parcelas de campo fue superior (147 o 192 parcelas de campo) el método de regresión linear multivariante mostró una aplicabilidad menor, siendo el método SVM el más preciso. La principal restricción de los modelos de regresión linear multivariante fue la distribución no-normal de las métricas ALS con mayor capacidad explicativa.

El análisis del efecto de las características ALS y las variables ambientales en la precisión de los modelos de estimación de biomasa residual forestal a escala regional determinó que una mayor densidad de puntos, menores ángulos de escaneo y una mayor penetración de los pulsos ALS en el dosel incrementa la precisión de los modelos. Por el contrario, la presencia de pendientes fuertes genera menores precisiones, pudiendo ser parcialmente explicadas por los errores generados en la derivación del modelo digital de elevaciones (MDE). La presencia de arbustos no generó un efecto claro en la ganancia o pérdida de precisión de los modelos, por lo que se debe continuar investigando sobre su efecto en otras masas forestales, así como en la estimación de otras variables.

La investigación ha enriquecido el conocimiento de las masas forestales mediterráneas de Pino carrasco utilizando datos ALS-PNOA poniendo en valor esta información pública y de gran utilidad para la gestión forestal a escala local y regional.

## 4.2. Conclusiones específicas

Las principales conclusiones para cada uno de los objetivos detallados en el capítulo 1 se presentan a continuación:

*Explorar la utilidad de los datos ALS de baja densidad en nuevas aplicaciones: estimación de la pérdida de biomasa y emisiones de CO<sub>2</sub> a la atmósfera, cuantificar y cartografiar la biomasa residual forestal y estimar diferentes fracciones de biomasa incluyendo la fracción arbustiva*

- La estimación de la biomasa arbórea pre-fuego utilizando datos ALS mostró un buen ajuste de los modelos, posibilitando la cuantificación precisa de esta variable forestal. Los modelos incluyen dos métricas ALS: el percentil 40 de la altura de los retornos, relacionado con la altura del dosel, y el porcentaje de primeros retornos sobre 2 m, asociado con la densidad del dosel. El índice dNBR obtenido mediante imágenes del sensor Landsat 8 permitió estimar la severidad del incendio y aplicar diferentes factores de combustión, solventando una de las fuentes de error tradicionales en la estimación de emisiones asociadas a incendios forestales, que considera la pérdida total de la biomasa tras el incendio. El establecimiento de tres factores de combustión asociados a tres rangos generales del dNBR permitió la extrapolación de la metodología a otros ambientes. Las pérdidas de biomasa se cuantificaron en 262.659,7 toneladas, lo que supone un total de 426.754,8 toneladas de CO<sub>2</sub> emitidas a la atmósfera por la combustión de las masas de Pino carrasco. La aplicación de factores de conversión a escala regional para estimar el contenido de carbono y posteriormente las emisiones de CO<sub>2</sub> reduce la incertidumbre asociada al uso de ecuaciones de carácter global, incrementando la precisión de las estimaciones. El enfoque metodológico en tres fases constituye una opción buena para estimar las pérdidas de biomasa forestal, minimizando el trabajo de campo y solventando la falta de información estructural post-incendio.
- Los coeficientes de correlación de Spearman determinaron las métricas ALS más adecuadas para la estimación de la biomasa residual forestal: el percentil 70 de la altura de los retornos, la varianza de la elevación y el porcentaje de primeros retornos sobre la media, representando la altura del dosel, la variabilidad y la densidad del mismo. Las 197.951,24 ha cartografiadas representan el 87,6% de la superficie forestal de Pino carrasco de Aragón, la cual alberga 3.627.021,25 toneladas. Los datos ALS-PNOA de baja densidad proporcionan información precisa para la gestión forestal relacionada con la estimación de este recurso de energía renovable, así como de la distribución de los combustibles del dosel a una escala regional.
- La estimación de biomasa total se realizó utilizando tres métricas ALS que caracterizan la distribución vertical de los retornos (el percentil 25 de la altura de los retornos), una métrica asociada con la variabilidad de la altura de los retornos (la varianza) y una métrica

que expresa la densidad del dosel (el porcentaje de primeros retornos sobre la media). La presencia de una densidad del dosel mayor y variabilidad de la nube de puntos están asociadas con un mayor contenido total de biomasa. Valores de altura más bajos en el percentil 25 determinan una mayor presencia de arbustos en el sotobosque y, consecuentemente, un incremento en la biomasa total. La presencia de sotobosque en los pinares de Pino carrasco mediterráneos, especialmente en aquellas masas semi-naturales, constituye una parte importante de la biomasa total. La estimación y cartografía de estos recursos forestales utilizando datos ALS de baja densidad proporciona información útil para la gestión forestal. La cartografía de biomasa total muestra valores que varían entre zonas con menos de 10 toneladas por hectárea a otras que superan las 150 toneladas por hectárea.

*Explorar la transferibilidad temporal de los modelos para la estimación de variables forestales a escala regional utilizando datos ALS-PNOA multitemporales*

- La adquisición de un número elevado de parcelas de campo para la realización de los inventarios con tecnología ALS a escala regional es costosa desde un punto de vista económico y demanda gran cantidad de tiempo. En este estudio, la estimación de las variables de campo para generar información concomitante al vuelo ALS en los años 2011 y 2016 se realizó utilizando modelos de crecimiento de árbol individual específicos, curvas generalizadas de altura-diámetro y ecuaciones cónicas. El uso de esta metodología produce resultados más precisos que la utilización de tablas de crecimiento para predecir el crecimiento de los árboles a corto plazo.
- La transferibilidad temporal de los modelos se realizó comparando los enfoques directo e indirecto. El enfoque indirecto, que consiste en el ajuste de un modelo diferente para cada uno de los años analizados, ha mostrado generalmente una mayor precisión. Sin embargo, utilizando el enfoque directo se obtuvieron resultados similares. Además, este enfoque reduce considerablemente el tiempo invertido en trabajo de campo, así como en la parametrización de los modelos, siendo posible extrapolar un modelo generado para una fecha (validado con datos de campo) a otra fecha en la que haya datos ALS, siempre que no hayan pasado muchos años entre ambas adquisiciones. Esta potencialidad beneficia a los gestores forestales y las empresas encargadas de realizar los inventarios forestales, reduce el tiempo y el coste de modelar parámetros forestales y sus cambios.

*Comparación de métodos paramétricos y no-paramétricos en el modelado de variables forestales*

- La comparación de ocho modelos de regresión para estimar la biomasa arbórea pre-fuego determinó que el modelo de regresión lineal multivariante era el mejor, superando a dos algoritmos de aprendizaje automático y cinco árboles de regresión. Los modelos se desarrollaron con 46 parcelas de campo y utilizando dos métricas ALS. La precisión de los modelos varía entre un 11,10 y un 23,20 %RMSE cuando se utiliza el modelo de regresión lineal multivariante y el algoritmo KNN, respectivamente. El test de Friedman determinó



que los modelos no eran equivalentes pero sólo se encontraron diferencias estadísticamente significativas entre el modelo lineal y el WKNN y KNN.

- El mejor modelo de regresión para estimar la biomasa total fue el modelo de regresión lineal multivariante, mejorando a los cuatro modelos no paramétricos analizados. Los modelos se generaron a partir de 83 parcelas de campo y constan de tres métricas ALS. El ajuste de los modelos varía de 19,21 hasta 24,93 %RMSE al utilizar el modelo de regresión lineal multivariante y random forest, respectivamente. No se encontraron diferencias estadísticamente significativas entre los métodos de regresión analizados. Todos los modelos se desarrollaron con el mismo número de parcelas y las mismas métricas ALS. Los modelos Random forest y regresión lineal ponderada localmente mostraron una sobreestimación previa a realizar la validación cruzada.
- El método SVM con kernel radial fue el más preciso para estimar la biomasa residual forestal a escala regional, con un 26,38 %RMSE. El mayor número de parcelas (192) y el incremento en la variabilidad de la muestra puede haber incrementado el rendimiento de los modelos de regresión no-paramétricos respecto a los modelos lineales. Aunque la generalización de modelos regionales puede generar resultados menos precisos para otros tipos de bosques utilizando el método SVM con kernel radial, consideramos que este modelo es útil para trabajar con un número de parcelas de campo elevadas y datos ALS.
- El método SVM con kernel radial fue el que obtuvo los mejores resultados en el análisis de la transferibilidad temporal de modelos entre 2011 y 2016. Los modelos de regresión lineal multivariante mostraron una aplicabilidad limitada con un número de parcelas de campo elevado (147) dado que las métricas ALS más representativas no se distribuían de forma normal, generando una pérdida de precisión y ajuste de los modelos. El ajuste de los modelos de estimación, con los datos ALS de 2011, de número de pies, área basal, diámetro cuadrático medio, diámetro dominante, altura dominante, volumen y biomasa arbórea fue de 38,55; 24,51; 17,25; 15,36; 28,22 y 26,71 %RMSE. Valores menores de error (%RMSE) se obtuvieron utilizando los datos ALS de 2016: 39,96; 18,80; 15,28; 11,89; 8,83; 21,20 y 18,81, respectivamente (las estadísticas se presentan en el mismo orden que las del año 2011).
- Los modelos SVM con kernel lineal, random forest, LWLR y MDL obtuvieron en su mayoría precisiones menores que el SVM con kernel radial. El uso de WKNN y KNN presentó una menor precisión al estimar la biomasa arbórea pre-fuego y mostró diferencias estadísticamente significativas con respecto al resto de métodos analizados. El método random forest mostró una tendencia a la sobreestimación durante la fase de entrenamiento, incluso aplicando métodos de corrección del sesgo. Sin embargo, este método proporcionó los mejores resultados para estimar la altura dominante con una muestra de 147 parcelas de campo. En este sentido, se debe continuar analizado si un mayor número de parcelas de campo en la muestra mejora la precisión de este método cuando se estiman variables más complejas.

*Evaluación de la idoneidad de diferentes métodos de selección de métricas para mejorar la precisión de los modelos siguiendo el principio de parsimonia*

- La comparación entre métodos de selección determinó que el método de selección de todos los subconjuntos y selección basada en los coeficientes de correlación de Spearman fueron las técnicas más adecuadas. La selección LASSO, selección paso a paso y análisis de componentes principales mostraron una menor capacidad para determinar los mejores subconjuntos de métricas. La selección de todos los subconjuntos es un método automático que permite determinar el número máximo de métricas a incluir en el modelo y proporciona resultados más rápidos que la selección basada en los coeficientes de correlación de Spearman, siendo esta última aplicada de forma manual con criterio experto.
- La selección de todos los subconjuntos fue el mejor método para la estimación de área basal, diámetro dominante, altura dominante y volumen, mientras que el coeficiente de correlación de Spearman determinó las mejores métricas para estimar la biomasa residual forestal, la biomasa arbórea y el número de pies. Ambos métodos mostraron la misma capacidad de selección al estimar el diámetro cuadrático medio y la biomasa total. Los métodos de selección paso a paso fueron utilizados para estimar la biomasa arbórea y el volumen, mientras que LASSO se utilizó para estimar el número de pies.
- Los percentiles elevados de altura de los retornos se seleccionaron recurrentemente para la estimación de la altura dominante. Métricas asociadas a la altura del dosel, a la variabilidad y a la densidad del dosel se seleccionaron para estimar la densidad de pies, el diámetro dominante, el diámetro cuadrático medio y la biomasa residual. Además de los tres tipos de métricas ALS mencionadas con anterioridad, también se seleccionaron métricas asociadas a las partes del dosel con menor altura para estimar el área basal, volumen, biomasa arbórea y biomasa total.

#### *Evaluación del efecto de diversos parámetros ALS y variables ambientales en el ajuste de los modelos*

- El error medio predicho se incrementó con densidades de puntos inferiores a 1 punto por metro cuadrado en 0,56 toneladas por hectárea. Sin embargo, la precisión se mantuvo en valores elevados, confirmando la utilidad de los datos ALS-PNOA de baja densidad. Esta información constituyen una alternativa óptima para estimar variables forestales cuando no existan nubes de puntos con una densidad superior.
- Los ángulos de escaneo superiores a 15° incrementan los valores de error medio predicho en 1,94 toneladas por hectárea, en comparación a ángulos próximos al nadir (<5°). Pese a ello, no se encontraron diferencias estadísticamente significativas entre los rangos analizados. Además, el análisis confirma los resultados de estudios previos, que proponían realizar la adquisición de datos evitando ángulos de escaneo superiores a 15°.
- La reducción de la penetración del pulso en el dosel, especialmente en aquellas áreas densamente cubiertas por estratos altos, incrementa el error medio predicho de los modelos hasta 5,47 toneladas por hectárea, cuando la penetración es inferior al 25%, respecto a parcelas cuya penetración del pulso es superior al 75%. La reducción de la penetración del pulso en el dosel genera también una reducción del número de retornos en los estratos inferiores y una menor precisión del MDE.

- Las pendientes fuertes generan un incremento del error de hasta 1,44 toneladas por hectárea, en comparación con pendientes suaves inferiores al 15%, si bien, no representan un cambio significativo. La reducción de la precisión puede ser explicada de forma parcial por la menor precisión del MDE, considerando que los algoritmos de filtrado tienen mayor error en las pendientes fuertes.
- La presencia de arbustos incrementa ligeramente el error medio predicho en 0,11 toneladas por hectárea. El efecto de la presencia de arbustos en el error de los modelos requiere ser analizado en la estimación de otras variables forestales, dado que los resultados obtenidos no muestran un patrón significativo.



### 4.3. Future working proposals

This research deepened into the use of low density ALS-PNOA data from two coverages to provide several forest applications in Aleppo pine forest of Aragón region. The comparison between parametric and non-parametric models under different conditions; the exploration of different variable selection methods; the analysis of ALS characteristics and environmental effects in error modelling; the cartography of different forestry variables at local and regional scales; and the analysis of model transferability between two ALS dates provided an insight of this public ALS data potentiality. This PhD Thesis also constitutes the “begin” for new research, which can be even more promising, either improving some of the presented applications or considering new research proposals.

The research field that might be more promising for future developments is the combination of ALS data with long temporal series of satellite images to characterize forest evolution during the last decades. These data may be also combined with 3D data derived from orthophotography to provide more accurate structural data from the past. Furthermore, the development of unmanned aerial vehicles provides new spatial and temporal scales of analysis for forest environments. The fusion with other remote sensing data would be one of the keys to develop regional products that may improve forest management.

Several research proposals are summarized below, some of which are currently under development:

- The estimation of forest variables for other Mediterranean species, including deciduous ones. The second ALS-PNOA coverage, mostly captured in the vegetative period, may provide suitable information for characterizing deciduous forests. In addition, the increase in point density of this second coverage may improve the generation of models in the shrubland areas that characterize the Mediterranean environments.
- The analysis of the effect of ALS filtering and interpolation errors in forest variable prediction. The comparison between several filtering and interpolation methods, including the performed by the PNOA project, could be tested to analyse modelling errors at regional scales.
- The refinement of existing DTMs and the use of triangulation or Scale Invariant Feature Transform (SIFT) for key point detection might enhance metrics calculations using low-density ALS data.
- The analysis of model uncertainty and cartography, using different methods as covariance values, Bayesian inference or Monte Carlo simulations, might improve forest decision making.
- The exploration of other variable heuristic-based selection methods such as genetic algorithms, simulated annealing or status quo model optimizations might be compared with the ones explored in this PhD Thesis.
- Fuel model classification can contribute to better managing Mediterranean forests, frequently affected by fires. The comparison of parametric and non-parametric methods

and the fusion with other remote sensing data, including images and point clouds derived from unmanned aerial vehicles, could be explored.

- The use of radiative transfer models like DART to simulate ALS data, and the subsequent comparison with ALS-PNOA data, might improve generalization and characterization of fuel models at regional scales.
- The use of pre- and post- fire ALS data derived metrics can be used to create severity indexes. The comparison between these indexes and the ones frequently generated with passive remote sensing data for determining fire severity requires further research. Furthermore, the fusion of structural and optical data may improve fire severity estimations.
- Structural diversity and textural metrics might improve traditional forest inventory variables prediction. These metrics also open new possibilities to characterize some fractions of forest diversity, analyse forest changes after disturbances or provide useful information to characterize pasture potential areas within the EU's Common Agricultural Policy.
- The positioning improvements of permanent plots from the National Forest Inventory might contribute to use these public datasets to estimate forestry variables at a regional scale. The availability of two or more ALS coverages may provide accurate structural data for characterizing forest changes.
- The determination of disturbances using algorithms like LandTrendr and long temporal series of satellite images could be explored to characterize forest evolution during the last decades. These data may be processed using new platforms as Google Earth Engine combined with ALS data to better characterize forest structure and improve forestry parameters estimation. The use of point clouds derived from orthophotography may also improve structural characterization.
- The effects of UAV flight configuration to characterize forest parameters at Mediterranean environments should be explored. The determination of the most suitable post-processing methods to generate point clouds from UAV images, filtering and interpolation algorithms should be tested. The fusion of UAV spectral and derived point cloud data with other remote sensing data may enhance forest management at local and regional scales.
- The need of generating a DEM to normalize a point cloud and subsequently estimate forestry variables should be explored when using ALS or UAV data.
- The use of ALS data and fusion with UAV data could be explored to estimate dominant trees structural characteristics to improve sampling for dendrochronology purposes.



## References



- Aguilar, F. J., Aguilar, M. A., Agüera, F., & Sánchez, J. (2006). The accuracy of grid digital elevation models linearly constructed from scattered sample data. *International Journal of Geographical Information Science*, 20(2), 169–192. <https://doi.org/10.1080/13658810500399670>
- Akagi, S. K., R. J. Yokelson, I. R. Burling, S. Meinardi, I. Simpson, D. R. Blake, G. R. McMeeking, A. Sullivan, T. Lee, S. Kreidenweis, S. Urbanski, J. Reardon, D. W. T. Griffith, T. J. Johnson, & D. R. Weise. (2013). Measurements of reactive trace gases and variable O<sub>3</sub> formation rates in some South Carolina biomass burning plumes. *Atmospheric Chemistry and Physics*, 13(3), 1141–1165. <https://doi.org/10.5194/acp-13-1141-2013>
- Aldred, A. H., & Bonnor, G. M. (1985). *Applications of airborne lasers to forest surveys* (Vol. 51). Petawawa National Forestry Institute, Canadian Forestry Service, Agriculture Canada.
- Alonso, R. (2018). SIMANFOR - Sistema de Simulación de Manejo Forestal Sostenible | Gestión Forestal Sostenible. Retrieved from <http://sostenible.palencia.uva.es/content/simanfor-sistema-de-simulacion-de-manejo-forestal-sostenible>
- Andersen, H.-E., McGaughey, R. J., & Reutebuch, S. E. (2005). Estimating forest canopy fuel parameters using LIDAR data. *Remote Sensing of Environment*, 94(4), 441–449. <https://doi.org/10.1016/j.rse.2004.10.013>
- Andersen, H.-E., Reutebuch, S. E., & McGaughey, R. J. (2006). A rigorous assessment of tree height measurements obtained using airborne lidar and conventional field methods. *Canadian Journal of Remote Sensing*, 32(5), 355–366. <https://doi.org/10.5589/m06-030>
- Bagaram, M., Giularelli, D., Chirici, G., Giannetti, F., & Barbati, A. 2018. “UAV Remote Sensing for Biodiversity Monitoring: Are Forest Canopy Gaps Good Covariates?” *Remote Sensing* 2018, Vol. 10, Page 1397 10 (9). Multidisciplinary Digital Publishing Institute: 1397. doi:10.3390/RS10091397.
- Baltsavias, E. (1999). Airborne laser scanning: basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2–3), 199–214. [https://doi.org/10.1016/S0924-2716\(99\)00015-5](https://doi.org/10.1016/S0924-2716(99)00015-5)
- Bollandsås, O. M., Gregoire, T. G., Næsset, E., & Øyen, B. H. (2013a). Detection of biomass change in a Norwegian mountain forest area using small footprint airborne laser scanner data. *Statistical Methods and Applications*, 22(1), 113–129. <https://doi.org/10.1007/s10260-012-0220-5>
- Bollandsås, O. M., Maltamo, M., Gobakken, T., & Naesset, E. (2013b). Comparing parametric and non-parametric modelling of diameter distributions on independent data using airborne laser scanning in a boreal conifer forest. *Forestry*, 86(4), 493–501. <https://doi.org/10.1093/forestry/cpt020>
- Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., & Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Québec. *Remote Sensing of Environment*, 112(10), 3876–3890. <https://doi.org/10.1016/J.RSE.2008.06.003>
- Boyd, D. S., & Danson, F. M. (2005). Satellite remote sensing of forest resources: three decades of research development. *Progress in Physical Geography: Earth and Environment*, 29(1), 1–26. <https://doi.org/10.1191/0309133305pp432ra>
- Bravo, F., Rodriguez, F., & Ordóñez, C. (2012). A web-based application to simulate alternatives for sustainable forest management: SIMANFOR. *Forest Systems*, 21(1), 4.

<https://doi.org/10.5424/fs/2112211-01953>

- Breidenbach, J., Koch, B., Kändler, G., & Kleusberg, A. (2008). Quantifying the influence of slope, aspect, crown shape and stem density on the estimation of tree height at plot level using lidar and InSAR data. *International Journal of Remote Sensing*, 29(5), 1511–1536. <https://doi.org/10.1080/01431160701736364>
- Breiman, L. (2001). Random Forest. *Machine Learning* 45, 5–32. <http://dx.doi.org/10.1023/A:1010933404324>
- Cabanillas, A. M. (2010). Bases para la gestión de masas naturales de *Pinus halepensis* Mill. en el Valle del Ebro. PhD Thesis, E.T.S.I. Montes Universidad Politécnica de Madrid. <http://oa.upm.es/4960/>
- Cabrera, J. (2013). Estimación de variables dasométricas a partir de datos LiDAR PNOA en masas regulares de *Pinus halepensis*, Daroca (Zaragoza). MSc Thesis, Dep. Geografía y Ordenación del Territorio, Universidad de Zaragoza. <https://zaguan.unizar.es/record/13229?ln=es>
- Cámara, A. (2001) Temperamento, aptitud y aplicaciones del pino carrasco (*Pinus halepensis* Mill.) en España. Análisis mediante un S.I.G. Tesis doctoral. Universidad Politécnica de Madrid. 357 p.
- Campbell, J. B. (2006). Introduction to remote sensing. Taylor & Francis, London.
- Cao, L., Coops, N. C., Innes, J. L., Sheppard, S. R. J., Fu, L., Ruan, H., & She, G. (2016). Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data. *Remote Sensing of Environment*, 178, 158–171. <https://doi.org/10.1016/j.rse.2016.03.012>
- Castro, K. L., Sanchez-Azofeifa, G. A., & Rivard, B. (2003). Monitoring secondary tropical forests using space-borne data: Implications for Central America. *International Journal of Remote Sensing*, 24(9), 1853–1894. <https://doi.org/10.1080/01431160210154056>
- Chasmer, L., Hopkinson, C., Smith, B., & Treitz, P. (2006a). Examining the Influence of Changing Laser Pulse Repetition Frequencies on Conifer Forest Canopy Returns. *Photogrammetric Engineering & Remote Sensing*, 72(12), 1359–1367. <https://doi.org/10.14358/PERS.72.12.1359>
- Chasmer, L., Hopkinson, C., & Treitz, P. (2006b). Investigating laser pulse penetration through a conifer canopy by integrating airborne and terrestrial lidar. *Canadian Journal of Remote Sensing*, 32(2), 116–125. <https://doi.org/10.5589/m06-011>
- Chirici, G., Barbati, A., Corona, P., Marchetti, M., Travaglini, D., Maselli, F., & Bertini, R. (2008). Non-parametric and parametric methods using satellite images for estimating growing stock volume in alpine and Mediterranean forest ecosystems. *Remote Sensing of Environment*, 112(5), 2686–2700. <https://doi.org/10.1016/J.RSE.2008.01.002>
- Chuvieco, E. (2010). Teledetección ambiental: La observación de la Tierra desde el espacio. Ariel Ciencia, Barcelona.
- Clark, M. L., Clark, D. B., & Roberts, D. A. (2004). Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sensing of Environment*, 91(1), 68–89. <https://doi.org/10.1016/J.RSE.2004.02.008>

- Cleveland, W. S., & Devlin, S. J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, 83(403), 596–610.
- Cohen, W. B., Harmon, M. E., Wallin, D. O., & Fiorella, M. (1996). Two Decades of Carbon Flux from Forests of the Pacific Northwest. *BioScience*, 46(11), 836–844. <https://doi.org/10.2307/1312969>
- Coops, N. ., & Waring, R. . (2001). The use of multiscale remote sensing imagery to derive regional estimates of forest growth capacity using 3-PGS. *Remote Sensing of Environment*, 75(3), 324–334. [https://doi.org/10.1016/S0034-4257\(00\)00176-0](https://doi.org/10.1016/S0034-4257(00)00176-0)
- Cowen, D. J., Jensen, J. R., Hendrix, C., Hodgson, M. E., & Schili, S. R. (2000). A GIS-Assisted Rail Construction Econometric Model that Incorporates LIDAR Data. *Photogrammetric Engineering & Remote Sensing*, 66(11), 1323–1328.
- Cuadrat, J. M. (2004). El clima de Aragón. In: J.L. Peña, L.A. Longares and M. Sánchez (Eds.), *Geografía Física de Aragón. Aspectos Generales y Temáticos*, Univresidad de Zaragoza e Institución Fernando el Católico, Zaragoza, 15–26.
- Darlington, R. B., & Horst, P. (1966). Factor Analysis of Data Matrices. *The American Journal of Psychology*, 79(2), 344. <https://doi.org/10.2307/1421153>
- De Santis, A., Asner, G. P., Vaughan, P. J., & Knapp, D. E. (2010). Mapping burn severity and burning efficiency in California using simulation models and Landsat imagery. *Remote Sensing of Environment*, 114(7), 1535–1545. <https://doi.org/10.1016/j.rse.2010.02.008>
- Disney, M. I., Kalogirou, V., Lewis, P., Prieto-Blanco, A., Hancock, S., & Pfeifer, M. (2010). Simulating the impact of discrete-return lidar system and survey characteristics over young conifer and broadleaf forests. *Remote Sensing of Environment*, 114(7), 1546–1560. <https://doi.org/10.1016/J.RSE.2010.02.009>
- Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., & Chazdon, R. L. (2010). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at la Selva, Costa Rica. *Journal of Geophysical Research: Biogeosciences*, 115(2), 1–17. <https://doi.org/10.1029/2009JG000933>
- Dube, T., & Mutanga, O. (2015). Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 36–46. <https://doi.org/10.1016/J.ISPRSJPRS.2014.11.001>
- Efroymson, M. A. (1960). Multiple regression analysis. In H. S. Ralston A. and Wilf (Ed.), *Mathematical Methods for Digital Computers*. Wiley, New York. Retrieved from <https://www.scopus.com/record/display.uri?eid=2-s2.0-0002879388&origin=inward>
- Estornell, J., Ruiz, L. A., Velázquez-Martí, B., & Hermosilla, T. (2012). Estimation of biomass and volume of shrub vegetation using LiDAR and spectral data in a Mediterranean environment. *Biomass and Bioenergy*, 46, 710–721. <https://doi.org/10.1016/j.biombioe.2012.06.023>
- Evans, J. S., & Hudak, A. T. (2007). A Multiscale Curvature Algorithm for Classifying Discrete Return LiDAR in Forested Environments. *IEEE Transactions on Geoscience and Remote*

- Sensing, 45(4), 1029–1038. <https://doi.org/10.1109/TGRS.2006.890412>
- Evans, J. S., Hudak, A. T., Faux, R., & Smith, A. M. S. (2009). Discrete Return Lidar in Natural Resources: Recommendations for Project Planning, Data Processing, and Deliverables. *Remote Sensing*, 1(4), 776–794. <https://doi.org/10.3390/rs1040776>
- Fekety, P. A., Falkowski, M. J., & Hudak, A. T. (2015). Temporal transferability of LiDAR-based imputation of forest inventory attributes. *Canadian Journal of Forest Research*, 45(4), 422–435. <https://doi.org/10.1139/cjfr-2014-0405>
- Fernández-Álvarez, M., Armesto, J., & Picos, J. (2019). LiDAR-Based Wildfire Prevention in WUI: The Automatic Detection, Measurement and Evaluation of Forest Fuels. *Forests*, 10(2), 148. <https://doi.org/10.3390/f10020148>
- Ferraz, A., Saatchi, S., Bormann, K., Painter, T., Ferraz, A., Saatchi, S., ... Painter, T. H. (2018). Fusion of NASA Airborne Snow Observatory (ASO) Lidar Time Series over Mountain Forest Landscapes. *Remote Sensing*, 10(2), 164. <https://doi.org/10.3390/rs10020164>
- Fix, E., Hodges, J.L. (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas.
- French, N. H. F., Goovaerts, P., & Kasischke, E. S. (2004). Uncertainty in estimating carbon emissions from boreal forest fires. *Journal of Geophysical Research*, 109(D14), D14S08. <https://doi.org/10.1029/2003JD003635>
- Gagliasso, D., Hummel, S., & Temesgen, H. (2014). A comparison of selected parametric and non-parametric imputation methods for estimating forest biomass and basal area. *Open J. For.*, 4, 42–48. <http://dx.doi.org/10.4236/ojf.2014.41008>
- García-Gutiérrez, J., Martínez-Álvarez, F., Troncoso, A., & Riquelme, J. C. (2015). A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. *Neurocomputing*, 167, 24–31. <https://doi.org/10.1016/j.neucom.2014.09.091>
- García-Llamas, P., Suárez-Seoane, S., Fernández-Guisuraga, J. M., Fernández-García, V., Fernández-Manso, A., Quintano, C., ... Calvo, L. (2019). Evaluation and comparison of Landsat 8, Sentinel-2 and Deimos-1 remote sensing indices for assessing burn severity in Mediterranean fire-prone ecosystems. *International Journal of Applied Earth Observation and Geoinformation*, 80, 137–144. <https://doi.org/10.1016/J.JAG.2019.04.006>
- García, D., Godino, M., & Mauro, F. (2012). Lidar: Aplicación Práctica Al Inventario Forestal. *Academica Española*.
- García, M., Riaño, D., Chuvieco, E., & Danson, F. M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data. *Remote Sensing of Environment*, 114(4), 816–830. <https://doi.org/10.1016/j.rse.2009.11.021>
- García, M., Saatchi, S., Ferraz, A., Silva, C. A., Ustin, S., Koltunov, A., & Balzter, H. (2017). Impact of data model and point density on aboveground forest biomass estimation from airborne LiDAR. *Carbon Balance and Management*, 12(1), 4. <https://doi.org/10.1186/s13021-017-0073-1>
- García, O. (2003). Dimensionality reduction in growth models: an example. *Forest biometry*,

- Modelling and Information Sciences 1, 1-15.
- Gatzliolis, D., Fried, J. S., & Monleon, V. S. (2010). Challenges to estimating tree height via LiDAR in closed-canopy forests: A parable from Western Oregon. *Forest Science*, 56(2), 139–155.
- Giannetti, F., Chirici, G., Gobakken, T., Næsset, E., Travaglini, D., & Puliti, S. (2018). A new approach with DTM-independent metrics for forest growing stock prediction using UAV photogrammetric data. *Remote Sensing of Environment*, 213(May), 195–205. <https://doi.org/10.1016/j.rse.2018.05.016>
- Gleason, C. J., & I. Jungho. (2012). A Review of Remote Sensing of Forest Biomass and Biofuel: Options for Small-Area Applications. *GIScience & Remote Sensing* 48 (2): 141–170. <https://doi.org/10.2747/1548-1603.48.2.141>
- Gobakken, T., & Næsset, E. (2008). Assessing effects of laser point density, ground sampling intensity, and field sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, 38(5), 1095–1109. <https://doi.org/10.1139/X07-219>
- González-de Vega, S., de las Heras, J., & Moya, D. (2016). Resilience of Mediterranean terrestrial ecosystems and fire severity in semiarid areas: Responses of Aleppo pine forests in the short, mid and long term. *Science of The Total Environment*, 573, 1171–1177. <https://doi.org/10.1016/J.SCITOTENV.2016.03.115>
- González-Ferreiro, E. M., Miranda, D., Barreiro-Fernandez, L., Bujan, S., Garcia-Gutierrez, J., & Dieguez-Aranda, U. (2013). Modelling stand biomass fractions in Galician Eucalyptus globulus plantations by use of different LiDAR pulse densities. *Forest Systems*, 22(3), 510–525.
- Görgens, E.B., Montaghi, A. & Rodriguez, L.C.E. (2015). A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Comput. Electron. Agric.*, 116, 221–227. <https://doi.org/10.1016/j.compag.2015.07.004>
- Guerra-Hernández, J., Tomé, M., & González-Ferreiro, E. (2016a). Using low density LiDAR data to map Mediterranean forest characteristics by means of an area-based approach and height threshold analysis. *Revista de Teledetección*, 0(46), 103–117.
- Guerra-Hernández, J., Görgens, E. B., García-Gutiérrez, J., Rodriguez, L. C. E., Tomé, M., & González-Ferreiro, E. (2016b). Comparison of ALS based models for estimating aboveground biomass in three types of Mediterranean forest. *European Journal of Remote Sensing*, 49, 185. <https://doi.org/10.5721/EuJRS20164911>
- Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3(Mar), 1157–1182.
- Hair, J. F., Prentice, E., & Cano, D. (1999). *Análisis multivariante*. Prentice-Hall.
- Hamelin, L., Borzęcka, M., Kozak, M., & Pudełko, R. (2019). A spatial approach to bioeconomy: Quantifying the residual biomass potential in the EU-27. *Renewable and Sustainable Energy Reviews*, 100, 127–142. <https://doi.org/10.1016/j.rser.2018.10.017>
- Hauglin, M., Gobakken, T., Astrup, R., Ene, L., & Næsset, E. (2014). Estimating single-tree crown biomass of norway spruce by airborne laser scanning: A comparison of methods with and



- without the use of terrestrial laser scanning to obtain the ground reference data. *Forests*, 5(3), 384–403. <https://doi.org/10.3390/f5030384>
- Hauglin, M., Gobakken, T., Lien, V., Bollandsås, O. M., & Næsset, E. (2012). Estimating potential logging residues in a boreal forest by airborne laser scanning. *Biomass and Bioenergy*, 36, 356–365. <https://doi.org/10.1016/j.biombioe.2011.11.004>
- Hazelton, M. L. (2015). Nonparametric Regression. *International Encyclopedia of the Social & Behavioral Sciences*, 867–877. <https://doi.org/10.1016/B978-0-08-097086-8.42124-0>
- Hechenbichler, K., & Schliep, K. (2014). Weighted k-Nearest-Neighbor Techniques and Ordinal Classification. <https://doi.org/https://doi.org/10.5282/ubm/epub.1769>
- Hill, R. A., & Broughton, R. K. (2009). Mapping the understorey of deciduous woodland from leaf-on and leaf-off airborne LiDAR data: A case study in lowland Britain. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(2), 223–233. <https://doi.org/10.1016/j.isprsjprs.2008.12.004>
- Hirata, Y. (2004). The effects of footprint size and sampling density in airborne laser scanning to extract individual trees in mountainous terrain. *International Archives for Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(8), 102–107.
- Hollaus, M., Wagner, W., Eberhöfer, C., & Karel, W. (2006). Accuracy of large-scale canopy heights derived from LiDAR data under operational constraints in a complex alpine environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(5), 323–338. <https://doi.org/10.1016/j.isprsjprs.2006.05.002>
- Holmgren, J. (2004). Prediction of tree height, basal area and stem volume in forest stands using airborne laser scanning. *Scandinavian Journal of Forest Research*, 19(6), 543–553. <https://doi.org/10.1080/02827580410019472>
- Hudak, A. T., Strand, E. K., Vierling, L. A., Byrne, J. C., Eitel, J. U. H., Martinuzzi, S., & Falkowski, M. J. (2012). Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sensing of Environment*, 123, 25–40. <https://doi.org/10.1016/j.rse.2012.02.023>
- Hyde, P., Dubayah, R., Walker, W., Blair, J. B., Hofton, M., & Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, 102(1–2), 63–73. <https://doi.org/10.1016/J.RSE.2006.01.021>
- Hyypä, J., Hyypä, H., Litkey, P., Yu, X., Haggren, H., Rönholm, P., Maltamo, M. (2000). Algorithms and Methods of Airborne Laser Scanning for Forest Measurements. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(8), 82–89. <https://doi.org/10.1.1.150.8427>
- Jakubowski, M. K., Guo, Q., & Kelly, M. (2013). Tradeoffs between lidar pulse density and forest measurement accuracy. *Remote Sensing of Environment*, 130, 245–253. <https://doi.org/10.1016/J.RSE.2012.11.024>
- Kachamba, D., Ørka, H., Næsset, E., Eid, T., & Gobakken, T. (2017). Influence of Plot Size on Efficiency of Biomass Estimates in Inventories of Dry Tropical Forests Assisted by

- Photogrammetric Data from an Unmanned Aircraft System. *Remote Sensing*, 9(6), 610. <https://doi.org/10.3390/rs9060610>
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187–200. <https://doi.org/10.1007/BF02289233>
- Kane, V. R., Gersonde, R. F., Lutz, J. A., McGaughey, R. J., Bakker, J. D., & Franklin, J. F. (2011). Patch dynamics and the development of structural and spatial heterogeneity in Pacific Northwest forests. *Canadian Journal of Forest Research*, 41(12), 2276–2291. <https://doi.org/10.1139/x11-128>
- Key, C. H., & Benson, N. (2005). Landscape Assessment: Ground Measure of Severity, the Composite Burn Index; and remote sensing of severity, the normalized burn ratio. In L. J. Lutes, D.C., Keane, R.E., Caratti, J.F., Key, C.H., Benson, N.C., Gangi (Ed.), FIREMON: Fire Effects Monitoring and Inventory System. Ogden, UT, USDA Forest Service, Rocky Mountain Research Station, Gen. Tech. Rep. RMRS-GTR-164.
- Khosravipour, A., Skidmore, A. K., Wang, T., Isenburg, M., & Khoshelham, K. (2015). Effect of slope on treetop detection using a LiDAR Canopy Height Model. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104, 44–52. <https://doi.org/10.1016/J.ISPRSJPRS.2015.02.013>
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2), 273–324. [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)
- Korpela, I., Ørka, H. O., Hyyppä, J., Heikkinen, V., & Tokola, T. (2010). Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(4), 369–379. <https://doi.org/10.1016/J.ISPRSJPRS.2010.04.003>
- Lal, R. (2008). Sequestration of atmospheric CO<sub>2</sub> in global carbon pools. *Energy and Environmental Science*, 1, 86–100. <https://doi.org/10.1039/b809492f>
- Latifi, H., Nothdurft, A., & Koch, B. (2010). Non-parametric prediction and mapping of standing timber volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry*, 83(4), 395–407. <https://doi.org/10.1093/forestry/cpq022>
- Latifi, H., Fassnacht, F. E., Müller, J., Tharani, A., Dech, S., & Heurich, M. (2015). Forest inventories by LiDAR data: A comparison of single tree segmentation and metric-based methods for inventories of a heterogeneous temperate forest. *International Journal of Applied Earth Observation and Geoinformation*, 42, 162–174. <https://doi.org/10.1016/J.JAG.2015.06.008>
- Latypov, D. (2002). Estimating relative lidar accuracy information from overlapping flight lines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 56(4), 236–245. [https://doi.org/10.1016/S0924-2716\(02\)00047-3](https://doi.org/10.1016/S0924-2716(02)00047-3)
- Lefsky, M.A., Cohen, W. B., Acker, S. A., Parker, G. G., Spies, T. A., & Harding, D. (1999). Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sensing of Environment*, 70(3), 339–361. [https://doi.org/10.1016/S0034-4257\(99\)00052-8](https://doi.org/10.1016/S0034-4257(99)00052-8)

- Lefsky, Michael A., Cohen, W. B., Parker, G. G., & Harding, D. J. (2002). Lidar Remote Sensing for Ecosystem Studies. *BioScience*, 52(1), 19. [https://doi.org/10.1641/0006-3568\(2002\)052\[0019:LRSFES\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052[0019:LRSFES]2.0.CO;2)
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News* 2/3, 18-22.
- Liu, J., Skidmore, A. K., Jones, S., Wang, T., Heurich, M., Zhu, X., & Shi, Y. (2018). Large off-nadir scan angle of airborne LiDAR can severely affect the estimates of forest structure metrics. *ISPRS Journal of Photogrammetry and Remote Sensing*, 136, 13–25. <https://doi.org/10.1016/J.ISPRSJPRS.2017.12.004>
- Longares, J.L. (2004). Variedad biogeográfica del territorio aragonés. In: J.L., Peña, L.A., Longares, & M. Sánchez (Eds.), *Geografía Física de Aragón: Aspectos generales y temáticos*. Universidad de Zaragoza e Institución Fernando el Católico, Zaragoza, pp. 27–40.
- Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, 27(7), 1297–1328. <https://doi.org/10.1080/01431160500486732>
- Maclean, G. A., & Krabill, W. B. (1986). Gross-Merchantable Timber Volume Estimation Using an Airborne Lidar System. *Canadian Journal of Remote Sensing*, 12(1), 7–18. <https://doi.org/10.1080/07038992.1986.10855092>
- Maltamo, M., Næsset, E., & Vauhkonen, J. (2014). *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*. (M. Maltamo, E. Næsset, & J. Vauhkonen, Eds.) (Vol. 27). Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-017-8663-8>
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. Academic Press. ISBN: 0124712525.
- Martens, H., & Naes, T. (1992). *Multivariate calibration*. John Wiley & Sons. ISBN: 978-0-471-93047-1
- Matasci, G., Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C., Hobart, G. W., ... Bater, C. W. (2018). Three decades of forest structural dynamics over Canada's forested ecosystems using Landsat time-series and lidar plots. *Remote Sensing of Environment*, 216, 697–714. <https://doi.org/10.1016/j.rse.2018.07.024>
- Mauya, E. W., Hansen, E. H., Gobakken, T., Bollandsås, O. M., Malimbwi, R. E., & Næsset, E. (2015). Effects of field plot size on prediction accuracy of aboveground biomass in airborne laser scanning-assisted inventories in tropical rain forests of Tanzania. *Carbon Balance and Management*, 10(1), 10. <https://doi.org/10.1186/s13021-015-0021-x>
- McCarley, T. R., Kolden, C. A., Vaillant, N. M., Hudak, A. T., Smith, A. M. S., Wing, B. M., & Kreidler, J. (2017). Multi-temporal LiDAR and Landsat quantification of fire-induced changes to forest structure. *Remote Sensing of Environment*, 191, 419–432. <https://doi.org/10.1016/j.rse.2016.12.022>
- McGaughey, R. (2009). *FUSION/LDV: Software for LIDAR Data Analysis and Visualization*. US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

- Means, J., Acker, S., Harding, D., Blair, J., Lefsky, M., Cohen, W., ... McKee, W. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western Cascades of Oregon. *Remote Sensing of Environment*, 67, 298–308.
- Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems*, 118, 62–69. <https://doi.org/10.1016/J.CHEMOLAB.2012.07.010>
- Meng, X., Currit, N., & Zhao, K. (2010). Ground Filtering Algorithms for Airborne LiDAR Data: A Review of Critical Issues. *Remote Sensing*, 2(3), 833–860. <https://doi.org/10.3390/rs2030833>
- Meyer, V., Saatchi, S. S., Chave, J., Dalling, J. W., Bohlman, S., Fricker, G. A., Robinson, C., Neumann, M., & Hubbell, S. (2013). Detecting Tropical Forest Biomass Dynamics from Repeated Airborne Lidar Measurements. *Biogeosciences* 10(8):5421–38.
- Mieville, A., Granier, C., Lioussé, C., Guillaume, B., Mouillot, F., Lamarque, J.-F., ... Pétron, G. (2010). Emissions of gases and particles from biomass burning during the 20th century using satellite data and an historical reconstruction. *Atmospheric Environment*, 44(11), 1469–1477. <https://doi.org/10.1016/j.atmosenv.2010.01.011>
- Miller, A. J. (2002). Subset selection in regression. Chapman & Hall/CRC.
- Miller, J. D., & Thode, A. E. (2007). Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). *Remote Sensing of Environment*, 109(1), 66–80. <https://doi.org/10.1016/J.RSE.2006.12.006>
- Montaghi, A. (2013). Effect of scanning angle on vegetation metrics derived from a nationwide Airborne Laser Scanning acquisition. *Canadian Journal of Remote Sensing*, 39(sup1), S152–S173. <https://doi.org/10.5589/m13-052>
- Montealegre, A. L., Lamelas, M. T., de la Riva, J., García-Martín, A., & Escribano, F. (2016). Use of low point density ALS data to estimate stand-level structural variables in Mediterranean Aleppo pine forest. *Forestry*, cpw008. <https://doi.org/10.1093/forestry/cpw008>
- Montealegre, A. L., Lamelas, M. T., & de la Riva, J. (2015a). A Comparison of Open-Source LiDAR Filtering Algorithms in a Mediterranean Forest Environment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8), 4072–4085. <https://doi.org/10.1109/JSTARS.2015.2436974>
- Montealegre, A., Lamelas, M., & de la Riva, J. (2015b). Interpolation Routines Assessment in ALS-Derived Digital Elevation Models for Forestry Applications. *Remote Sensing*, 7(7), 8631–8654. <https://doi.org/10.3390/rs70708631>
- Montero, G., Pasalodos-Tato, M., López-Senespleda, E., Onrubia, R., & Madrigal, G. (2013). Ecuaciones para la estimación de la biomasa en matorrales y arbustados mediterráneos. In *Proceedings of the 6th Congreso Forestal Español*, Pontevedra, Spain, 10 June–14 June 2013; Sociedad Española de Ciencias Forestales: Pontevedra, Spain.
- Montero, G., Ruiz-Peinado, R., & Muñoz, M. (2005). Producción de biomasa y fijación de CO<sub>2</sub> por los bosques españoles. *Producción De Biomasa y Fijación De CO<sub>2</sub> Por Los Bosques Españoles*. Instituto Nacional de Investigación y Técnica Agraria y Alimentaria.

- Mountrakis, G., Im, J., & Ogole, C. (2011): Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66 (3) Issue 3, 247-259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Næsset, E. (1997). Determination of mean tree height of forest stands using airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(2), 49–56. [https://doi.org/10.1016/S0924-2716\(97\)83000-6](https://doi.org/10.1016/S0924-2716(97)83000-6)
- Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, 80(1), 88–99. [https://doi.org/10.1016/S0034-4257\(01\)00290-5](https://doi.org/10.1016/S0034-4257(01)00290-5)
- Næsset, E. (2009). Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, 113(1), 148–159. <https://doi.org/10.1016/j.rse.2008.09.001>
- Næsset, E. (2011). Estimating above-ground biomass in young forests with airborne laser scanning. *International Journal of Remote Sensing*, 32(2), 473–501. <https://doi.org/10.1080/01431160903474970>
- Næsset, E., & Gobakken, T. (2005). Estimating forest growth using canopy metrics derived from airborne laser scanner data. *Remote Sensing of Environment*, 96(3–4), 453–465. <https://doi.org/10.1016/j.rse.2005.04.001>
- Næsset, E., & Økland, T. (2002). Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*, 79(1), 105–115. [https://doi.org/10.1016/S0034-4257\(01\)00243-7](https://doi.org/10.1016/S0034-4257(01)00243-7)
- Nemenyi, P. (1963). *Distribution-Free Multiple Comparisons*. Ph.D. Thesis, Princeton University, Princeton, NY, USA.
- Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*, 56(1), 1–7. [https://doi.org/10.1016/0034-4257\(95\)00224-3](https://doi.org/10.1016/0034-4257(95)00224-3)
- Noordermeer, L., Bollandsås, O. M., Gobakken, T., & Næsset, E. (2018). Direct and indirect site index determination for Norway spruce and Scots pine using bitemporal airborne laser scanner data. *Forest Ecology and Management*, 428(June), 104–114. <https://doi.org/10.1016/j.foreco.2018.06.041>
- Noordermeer, L., Bollandsås, O. M., Ørka, H. O., Næsset, E., & Gobakken, T. (2019). Comparing the accuracies of forest attributes predicted from airborne laser scanning and digital aerial photogrammetry in operational forest inventories. *Remote Sensing of Environment*, 226, 26–37. <https://doi.org/10.1016/J.RSE.2019.03.027>
- Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. C. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*, 275, 117–129. <https://doi.org/10.1016/j.foreco.2012.03.003>
- Ørka, H. O., Bollandsås, O. M., Hansen, E. H., Næsset, E., & Gobakken, T. (2018). Effects of terrain

- slope and aspect on the error of ALS-based predictions of forest attributes. *Forestry: An International Journal of Forest Research*, 91(2), 225–237. <https://doi.org/10.1093/forestry/cpx058>
- Pausas, J. G., Llovet, J., Rodrigo, A., & Vallejo, R. (2008). Are wildfires a disaster in the Mediterranean basin? – A review. *International Journal of Wildland Fire*, 17(6), 713–723. <https://doi.org/10.1071/WF07151>
- Peña, D. (2002). *Regresión y diseño de experimentos*. Alianza. ISBN 8420686956.
- Peña, J., & Lozano, M.V. (2004). “Las unidades del relieve aragonés.” In: J.L. Peña, L.A. Longares, & M. Sánchez (Eds.), *Geografía Física de Aragón: Aspectos generales y temáticos*. Universidad de Zaragoza e Institución Fernando el Católico, Zaragoza, pp. 3–14.
- Penner, M., Pitt, D. G., & Woods, M. E. (2013). Parametric vs. non parametric LiDar models for operational forest inventory in boreal Ontario.
- Periasamy, S. (2018). Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1. *Remote Sensing of Environment*, 217, 537–549. <https://doi.org/10.1016/J.RSE.2018.09.003>
- Pétron, G., Granier, C., Khatatov, B., Yudin, V., Lamarque, J.-F., Emmons, L., Edwards, D. P. (2004). Monthly CO surface sources inventory based on the 2000-2001 MOPITT satellite data. *Geophysical Research Letters*, 31(21). <https://doi.org/10.1029/2004GL020560>
- Pfeifer, M., Gonsamo, A., Disney, M., Pellikka, P., & Marchant, R. (2012). Leaf area index for biomes of the Eastern Arc Mountains: Landsat and SPOT observations along precipitation and altitude gradients. *Remote Sensing of Environment*, 118, 103–115. <https://doi.org/10.1016/J.RSE.2011.11.009>
- Pflugmacher, D., Cohen, W. B., Kennedy, R. E., & Yang, Z. (2014). Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics. *Remote Sensing of Environment*, 151, 124–137. <https://doi.org/10.1016/j.rse.2013.05.033>
- Poudel, K. P., Flewelling, J. W., & Temesgen, H. (2018). Predicting volume and biomass change from multi-temporal lidar sampling and remeasured field inventory data in Panther Creek Watershed, Oregon, USA. *Forests*, 9(1). <https://doi.org/10.3390/f9010028>
- Puliti, S., Ene, L. T., Gobakken, T., & Næsset, E. (2017). Use of partial-coverage UAV data in sampling for large scale forest inventories. *Remote Sensing of Environment*, 194, 115–126. <https://doi.org/10.1016/j.rse.2017.03.019>
- Puliti, S., Talbot, B., Astrup, R., Puliti, S., Talbot, B., & Astrup, R. (2018). Tree-Stump Detection, Segmentation, Classification, and Measurement Using Unmanned Aerial Vehicle (UAV) Imagery. *Forests*, 9(3), 102. <https://doi.org/10.3390/f9030102>
- Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, 27(3), 221–234. [https://doi.org/10.1016/S0020-7373\(87\)80053-6](https://doi.org/10.1016/S0020-7373(87)80053-6)
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5), 465–471. [https://doi.org/10.1016/0005-1098\(78\)90005-5](https://doi.org/10.1016/0005-1098(78)90005-5)

- Rodrigues, M., Ibarra, P., Echeverria, M., Pérez-Cabello, F., & de la Riva, J. (2014). A method for regional-scale assessment of vegetation recovery time after high-severity wildfires: Case study of Spain. *Progress in Physical Geography*, 38(5), 556–575. <https://doi.org/10.1177/0309133314542956>
- Royo-Alboreca, A., Cabanillas-Saldaña, A. M., Barrio-Anta, M., Notivol-Paíno, E., & Gorgoso-Varela, J. J. (2017). Site index curves for natural Aleppo pine forests in the central Ebro valley (Spain). *Madera y Bosques*, 23(1), 143. <https://doi.org/10.21829/myb.2017.231495>
- Roussel, J.-R., Caspersen, J., Béland, M., Thomas, S., & Achim, A. (2017). Removing bias from LiDAR-based estimates of canopy height: Accounting for the effects of pulse density and footprint size. *Remote Sensing of Environment*, 198, 1–16. <https://doi.org/10.1016/J.RSE.2017.05.032>
- Roy, D. P., Jin, Y., Lewis, P. E., & Justice, C. O. (2005). Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sensing of Environment*, 97(2), 137–162. <https://doi.org/10.1016/j.rse.2005.04.007>
- Ruiz, L., 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>
- Sá, A. C. L., Pereira, J. M. C., & Silva, J. M. N. (2005). Estimation of combustion completeness based on fire-induced spectral reflectance changes in a dambo grassland (Western Province, Zambia). *International Journal of Remote Sensing*, 26(19), 4185–4195. <https://doi.org/10.1080/01431160500113468>
- Sačkov, I., Kulla, L., Bucha, T., Sačkov, I., Kulla, L., & Bucha, T. (2019). A Comparison of Two Tree Detection Methods for Estimation of Forest Stand and Ecological Variables from Airborne LiDAR Data in Central European Forests. *Remote Sensing*, 11(12), 1431. <https://doi.org/10.3390/rs11121431>
- Schröder, J., & Gadow, K. von. (1999). Testing a new competition index for Maritime pine in northwestern Spain. *Canadian Journal of Forest Research*, 29(2), 280–283. <https://doi.org/10.1139/x98-199>
- Sebastián-López, A., Salvador-Civil, R., Gonzalo-Jiménez, J., & SanMiguel-Ayanz, J. (2008). Integration of socio-economic and environmental variables for modelling long-term fire danger in Southern Europe. *European Journal of Forest Research*, 127(2), 149–163. <https://doi.org/10.1007/s10342-007-0191-5>
- Seiler, W., & Crutzen, P. J. (1980). Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. *Climatic Change*, 2(3), 207–247. <https://doi.org/10.1007/BF00137988>
- Shan, J., & Toth, C. K. (2018). *Topographic laser ranging and scanning : principles and processing*. CRC Press. ISBN 9781498772273
- Shifley, S. R., He, H. S., Lischke, H., Wang, W. J., Jin, W., Gustafson, E. J., Yang, J. (2017). The past and future of modeling forest dynamics: from growth and yield curves to forest landscape models. *Landscape Ecology*, 32(7), 1307–1325. <https://doi.org/10.1007/s10980-017-0540-9>



- Shin, P., Sankey, T., Moore, M., & Thode, A. E. (2018). Evaluating Unmanned Aerial Vehicle Images for Estimating Forest Canopy Fuels in a Ponderosa Pine Stand. *Remote Sensing*, 10(8), 1266. <https://doi.org/10.3390/rs10081266>
- Silva, C. A., Klauberg, C., Hudak, A. T., Vierling, L. A., Liesenberg, V., Carvalho, S. P. C. E., & Rodriguez, L. C. E. (2016). A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data. *Forestry*, 89(4), 422-433. <https://doi.org/10.1093/forestry/cpw016>
- Singh, K. K., Chen, G., McCarter, J. B., & Meentemeyer, R. K. (2015). Effects of LiDAR point density and landscape context on estimates of urban forest biomass. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 310-322. <https://doi.org/10.1016/j.isprsjprs.2014.12.021>
- Sithole, G., & Vosselman, G. (2004). Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(1-2), 85-101. <https://doi.org/10.1016/J.ISPRSJPRS.2004.05.004>
- Skowronski, N. S., Clark, K. L., Gallagher, M., Birdsey, R. A., & Hom, J. L. (2014). Airborne laser scanner-assisted estimation of aboveground biomass change in a temperate oak-pine forest. *Remote Sensing of Environment*, 151, 166-174. <https://doi.org/10.1016/j.rse.2013.12.015>
- Socha, J., Pierzchalski, M., Bałazy, R., & Ciesielski, M. (2017). Modelling top height growth and site index using repeated laser scanning data. *Forest Ecology and Management*, 406, 307-317. <https://doi.org/10.1016/J.FORECO.2017.09.039>
- Solberg, S., Næsset, E., Hanssen, K. H., & Christiansen, E. (2006). Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. *Remote Sensing of Environment*, 102(3-4), 364-376. <https://doi.org/10.1016/J.RSE.2006.03.001>
- Soverel, N. O., Perrakis, D. D. B., & Coops, N. C. (2010). Estimating burn severity from Landsat dNBR and RdNBR indices across western Canada. *Remote Sensing of Environment*, 114(9), 1896-1909. <https://doi.org/10.1016/J.RSE.2010.03.013>
- Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1), 72. <https://doi.org/10.2307/1412159>
- Stojanova, D., Panov, P., Gjorgjioski, V., Kobler, A., & Džeroski, S. (2010). Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecological Informatics*, 5(4), 256-266. <https://doi.org/10.1016/j.ecoinf.2010.03.004>
- Strunk, J., Temesgen, H., Andersen, H.-E., Flewelling, J. P., & Madsen, L. (2012). Effects of lidar pulse density and sample size on a model-assisted approach to estimate forest inventory variables. *Canadian Journal of Remote Sensing*, 38(5), 644-654. <https://doi.org/10.5589/m12-052>
- Tanase, M. A., Panciera, R., Lowell, K., Tian, S., Hacker, J. M., & Walker, J. P. (2014). Airborne multi-temporal L-band polarimetric SAR data for biomass estimation in semi-arid forests. *Remote Sensing of Environment*, 145, 93-104. <https://doi.org/10.1016/j.rse.2014.01.024>
- Tesch, S. D. (1980). The evolution of forest yield determination and site classification. *Forest*

- Ecology and Management, 3, 169–182. [https://doi.org/10.1016/0378-1127\(80\)90014-6](https://doi.org/10.1016/0378-1127(80)90014-6)
- Thornley, J. (2006). Modelling forest ecosystems : the Edinburgh Forest Model.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*. WileyRoyal Statistical Society. <https://doi.org/10.2307/2346178>
- Trozzi, C., Vaccaro, R., & Piscitello, R. (2002). Emissions estimate from forest fires: methodology, software and European case studies. In *Proceedings of the 11th International Conference, on Emission Inventories – Partnering for the Future*, United States Environmental Protection Agency, Atlanta, GA, USA.
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., ... van Leeuwen, T. T. (2010). Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmospheric Chemistry and Physics*, 10(23), 11707–11735. <https://doi.org/10.5194/acp-10-11707-2010>
- Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of the American Medical Informatics Association : JAMIA*, 8(5), 443–459.
- Vanclay, J. K. (1994). Modelling forest growth and yield : applications to mixed tropical forests. CAB International, Wallingford, UK.
- Vepakomma, U., St-Onge, B., & Kneeshaw, D. (2008). Spatially explicit characterization of boreal forest gap dynamics using multi-temporal lidar data. *Remote Sensing of Environment*, 112(5), 2326–2340. <https://doi.org/10.1016/J.RSE.2007.10.001>
- Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2016.04.008>
- Vosselman, G. (2012). Automated planimetric quality control in high accuracy airborne laser scanning surveys. *ISPRS Journal of Photogrammetry and Remote Sensing*, 74, 90–100. <https://doi.org/10.1016/J.ISPRSJPRS.2012.09.002>
- Vosselman, G., & Maas, H.-G. (2010). *Airborne and Terrestrial Laser Scanning*. Whittles Publishing.
- Wang, Y., & Witten, I. H. (1996). Induction of model trees for predicting continuous classes. *Computer Science Working Papers*.
- Wasser, L., Day, R., Chasmer, L., & Taylor, A. (2013). Influence of vegetation structure on lidar-derived canopy height and fractional cover in forested riparian buffers during leaf-off and leaf-on conditions. *PloS One*, 8(1), e54776. <https://doi.org/10.1371/journal.pone.0054776>
- Watt, M. S., Adams, T., Gonzalez Aracil, S., Marshall, H., & Watt, P. (2013). The influence of LiDAR pulse density and plot size on the accuracy of New Zealand plantation stand volume equations. *New Zealand Journal of Forestry Science*, 43(1), 15. <https://doi.org/10.1186/1179-5395-43-15>

- 
- Watt, P. J., Cox, N. J., & Wilson, J. (2007). Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data. *Remote Sensing of Environment*, 110(4), 509–522. <https://doi.org/10.1016/j.rse.2007.02.032>
- Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., & Soja, A. J. (2011). The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development*, 4(3), 625–641. <https://doi.org/10.5194/gmd-4-625-2011>
- Yu, X., Hyypä, J., Kaartinen, H., Maltamo, M., & Hyypä, H. (2008). Obtaining plotwise mean height and volume growth in boreal forests using multi-temporal laser surveys and various change detection techniques. *International Journal of Remote Sensing*, 29(5), 1367–1386. <https://doi.org/10.1080/01431160701736356>
- Yu, Xiaowei, Hyypä, J., Hyypä, H., & Maltamo, M. (2004). Effects of flight altitude on tree height estimation using airborne laser scanning. *International Archives Of Photogrammetry Remote Sensing And Spatial Information Sciences*, XXXVI(8-W2), 96–101.
- Zhao, K., Suarez, J. C., Garcia, M., Hu, T., Wang, C., & Londo, A. (2018). Utility of multitemporal lidar for forest and carbon monitoring: Tree growth, biomass dynamics, and carbon flux. *Remote Sensing of Environment*, 204, 883–897. <https://doi.org/10.1016/J.RSE.2017.09.007>



## Appendix: papers metrics



The following table summarize the publication references that constitute the PhD Thesis body, the impact factor of the journals (IF in the year of the paper, *Journal Citation Reports* 2019), and the task performed by the authors in each study. The PhD student, Darío Domingo, is the first author and responsible for each one of the published articles. This research works will not be part of any other PhD Thesis by paper compendium.

| Domingo, D., Lamelas, M.T., Montealegre, A.L., de la Riva, J. 2017. Comparison of regression models to estimate biomass losses and CO <sub>2</sub> emissions using low-density airborne laser scanning data in a burnt Aleppo pine forest. <i>European Journal of Remote Sensing</i> , 50 (1), 384-396. doi: 10.1080/22797254.2017.1336067   |            |                |              |          |
|--|------------|----------------|--------------|----------|
| In this work the PhD student, Darío Domingo, is responsible for most of the work, having performed the statistical and spatial analysis and being the main responsible for writing contents, discussion and conclusions. Dra. María Teresa Lamelas, Dr. Antonio Luis Montealegre and Dr. Juan de la Riva listed as co-authors are responsible of methodological development, fieldwork campaign and reviewing the analyses performed and manuscript writing. |            |                |              |          |
| IF   | IF 5 years | JCR® category  | Order number | Quartile |
| 1.122  | 1.383      | Remote Sensing | 24 of 30     | Q4       |

| Domingo, D., Lamelas, M.T., Montealegre, A.L., García-Martín, A., de la Riva, J. 2018. Estimation of total biomass in Aleppo pine forest stands applying parametric and nonparametric methods to low-density airborne laser scanning data. <i>Forests</i> , 9,158-175. doi: 10.3390/f9030158   |            |               |              |          |
|--|------------|---------------|--------------|----------|
| In this work the PhD student, Darío Domingo, is responsible for most of the work, having performed the statistical and spatial analysis and being the main responsible for writing contents, discussion and conclusions. Dra. María Teresa Lamelas, Dr. Antonio Luis Montealegre, Dr. Alberto García-Martín, and Dr. Juan de la Riva listed as co-authors are responsible of methodological development, fieldwork campaign and reviewing the analyses performed and manuscript writing. |            |               |              |          |
| IF   | IF 5 years | JCR® category | Order number | Quartile |
| 2.116  | 2.453      | Forestry      | 17 of 67     | Q2       |

| Domingo, D., Alonso, R, Lamelas, M.T., Montealegre, A.L., Rodríguez, F, de la Riva, J. 2019. Temporal Transferability of Pine Forest Attributes Modeling Using Low-Density Airborne Laser Scanning Data. <i>Remote Sensing</i> , 11(3), 261. doi:10.3390/rs11030261.   |            |                |              |          |
|--|------------|----------------|--------------|----------|
| In this work the PhD student, Darío Domingo, is responsible for most of the work, having performed the statistical and spatial analysis and being the main responsible for writing contents, discussion and conclusions. Dr. Rafael Alonso helped developing the statistical analysis as well as collaborate in reviewing the manuscript and fieldwork campaign. Dra. María Teresa Lamelas, Dr. Antonio Luis Montealegre, Dr. Francisco Rodríguez, and Dr. Juan de la Riva listed as co-authors are responsible of methodological development, fieldwork campaign and reviewing the analyses performed and manuscript writing. |            |                |              |          |
| IF   | IF 5 years | JCR® category  | Order number | Quartile |
| 4.118  | 4.740      | Remote Sensing | 7 of 30      | Q1       |



Domingo, D., Montealegre, A.L., Lamelas, M.T., García-Martín, A., de la Riva, J., Rodríguez, F, Alonso, R. 2019. Quantifying forest residual biomass in *Pinus halepensis* Miller stands using Airborne Laser Scanning data. *GIScience and Remote Sensing*, 56 (8), 1210-1232. doi: 10.1080/15481603.2019.1641653

In this work the PhD student, Darío Domingo, is responsible for most of the work, having performed the statistical and spatial analysis and being the main responsible for writing contents, discussion and conclusions. Dr. Antonio Luis Montealegre, Dra. María Teresa Lamelas, Dr. Alberto García-Martín, Dr. Juan de la Riva listed as co-authors are responsible of methodological development, fieldwork campaign and reviewing the analyses performed and manuscript writing. Dr. Francisco Rodríguez, and Dr. Rafael Alonso listed as co-authors are responsible of reviewing the manuscript and fieldwork campaign.

| IF    | IF 5 years | JCR® category          | Order number | Quartile |
|-------|------------|------------------------|--------------|----------|
| 3.588 | 3.186      | Remote Sensing         | 9 of 30      | Q2       |
| 3.588 | 3.186      | Geography,<br>Physical | 16 of 50     | Q2       |