



7º CONGRESO FORESTAL ESPAÑOL

**Gestión del monte: servicios
ambientales y bioeconomía**

26 - 30 junio 2017 | Plasencia
Cáceres, Extremadura

7CFE01-150

Edita: Sociedad Española de Ciencias Forestales
Plasencia. Cáceres, Extremadura. 26-30 junio 2017
ISBN 978-84-941695-2-6

© Sociedad Española de Ciencias Forestales

Delineación de copas, identificación de especies y estimación de la producción de piña a partir de tecnología LiDAR.

BLÁZQUEZ-CASADO, A.^{1,5}, VALBUENA, M,², CALAMA, R³, PÉREZ-CABELLO, F⁴, RODRIGUEZ, F.^{1,5}

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

² Departamento de Educación Gobierno Vasco. IES Murgía BHI

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

⁴ Departamento de Geografía, Universidad de Zaragoza, C/ Pedro Cerbuna 12, 50009, Zaragoza, España

⁵ Departamento de Producción Vegetal y Recursos Forestales, EU de Ingenierías Agrarias. Universidad de Valladolid, Soria, España

Resumen

En las últimas décadas se ha producido una explosión de la aplicación de la información LiDAR en multitud de campos, muy especialmente como fuente de información base en el desarrollo del inventario forestal a diferentes escalas en España. La aplicación de esta tecnología en la estimación de variables dendrométricas presenta gran interés en la gestión de masas forestales abiertas o adhesadas donde tradicionalmente la gestión se ha desarrollado a nivel de árbol individual.

Por otra parte, el uso de imágenes de satélite en el ámbito forestal, cada día está más extendido y, sobre todo, los datos son más accesibles, más recurrentes y con mayor resolución espacial.

El presente trabajo detalla los resultados de un inventario con datos LiDAR del Plan Nacional de Ortofotografía Aérea (PNOA) mediante el método de árbol individual en masas puras y mixtas de *Pinus pinea* y *Pinus pinaster* en la meseta castellana. Este tipo de masas relativamente abiertas permite el uso de esta metodología de individualización, por la cual se consigue la delineación de las copas de cada individuo. A parte, se ha ensayado el uso de imágenes de muy alta resolución, combinado con la información de la delineación de copas, de cara a discriminar especies (piñonero vs pinaster vs pies menores). Finalmente, sólo sobre los pies identificados como piñonero, se aplica un modelo de producción de piña (Calama et al., 2016) desarrollado originalmente para su uso a nivel de árbol individual utilizando datos habituales de los inventarios domésticos, pero adaptado para su uso con información procedente de LiDAR.

Palabras clave

Inventario, individualización, modelización, satélite, piña.

1. Introducción

La evolución tecnológica de la teledetección a lo largo de las últimas décadas ha permitido el desarrollo de una serie de sistemas de seguimiento de los elementos de la superficie terrestre tanto a escala global como regional con una periodicidad cada vez mayor. Todo ello, unido al aumento paulatino de la resolución espacial, espectral y temporal que ofrecen los distintos sensores, posibilita la generación de un enorme volumen de información espacialmente georreferenciada y radiométricamente corregida -en mayor o menor grado- aplicable tanto en el ámbito urbano-socioeconómico como en el ambiental (Martínez-Cañadas et al., 1998). La naturaleza de esta información depende del sensor que la registre. En este sentido encontramos desde ortofotos a imágenes multiespectrales, hiperespectrales, LiDAR o RADAR, que nos permiten recibir información de regiones del espectro electromagnético para las que el ojo humano no está preparado, tales como la región ultravioleta, la infrarroja o la propia de las microondas empleadas en la tecnología RADAR (Chuvieco, 2008).

Para una adecuada gestión forestal, debemos conocer la situación de las masas forestales tanto en su aspecto cuantitativo como cualitativo. Generalmente, este conocimiento se ha basado en la realización de inventarios forestales por muestreo. Actualmente estos inventarios se realizan casi siempre de forma manual y su coste económico es muy elevado. En consecuencia, no siempre es posible asumir su costo y no se ejecutan con la frecuencia e intensidad necesaria para obtener la información con que apoyar la toma de decisiones para la adecuada gestión forestal (Rodríguez y Lizarralde, 2015). Como alternativa a la adquisición manual de datos, la aparición de sensores LiDAR ha supuesto una gran revolución al permitir reducir el coste de inventario a más de la mitad. La inclusión de esta información LiDAR en el Plan Nacional de Ortofotografía Aérea (PNOA) ha permitido el desarrollo de métodos automáticos de inventario y evaluación de recursos forestales de gran precisión y coste muy reducido. Estos métodos para la explotación de datos LiDAR en la evaluación de recursos forestales responden a dos estrategias bien diferenciadas: por un lado, los “métodos de masa” y por otro los de “árbol individual”.

El cálculo del Inventario Forestal por “métodos de masa” a partir de datos LIDAR, requiere de una componente de inventario clásico. Por un lado, se realizan parcelas de muestreo en campo para la obtención de datos de la masa forestal. Por otro lado, se obtienen variables estadísticas a partir de los datos LIDAR. En base a esta información se calculan los modelos de regresión de cada una de las variables del inventario forestal a nivel de parcela. Los modelos relacionan variables directas obtenidas a partir del procesado de datos LIDAR (variables explicativas), con información recogida en campo utilizando las técnicas convencionales de inventario forestal (variables respuesta) (Nilsson, 1996; Latifi et al., 2010; Magnussen et al., 2010).

Aunque la gestión forestal no se realice a nivel de árbol, sino de rodal, en ocasiones es necesario conocer características concretas de árboles individuales. Es aquí en donde entra en juego el método de “árbol individual”, que representa un inventario pie a pie, pero realizado a partir de la información LiDAR y en el que sobre cada árbol medimos su altura y estimamos el resto de sus características. Hay que tener en cuenta que la individualización requiere la localización y delineación de la copa de cada árbol, independientemente de su forma y estado. Cuando la densidad de pies descende, como ocurre en dehesas, sabinares, masas de piñonero para obtención de piñón o masas de pinaster para resinación, la eficacia de los métodos de individualización crece y es posible realizarlos con la información LiDAR del PNOA (Hyyppä et al., 2001; Popescu et al., 2003; Heurich et al., 2004).

Por otra parte, la identificación de especies, tanto en masas puras como sobre todo en masas mixtas es uno de los grandes retos forestales que tiene la teledetección. La gestión forestal tiene cada vez más a valorar las masas mixtas (Koenig & Höfle, 2016; Zhang et al., 2016; Brandtberg, 2007; Yao et al., 2012) y las masas con cobertura completa, promoviendo una gestión próxima a la naturaleza (Mason et al., 1999; Davies et al., 2008). Es en estos casos en donde más nos interesa ser capaces de identificar especies a partir de técnicas basadas en sensores remotos.

Finalmente, en el ambiente mediterráneo en donde se desarrolla este trabajo, y, sobre todo, sobre las especies estudiadas (pino piñonero y pino resinero), el mayor valor económico de su aprovechamiento no responde a la madera si no a los denominados productos no maderables. Por tanto, resulta imprescindible poder obtener modelos precisos sobre la producción de éstos.

2. Objetivos

En el presente estudio se pretende alcanzar tres objetivos, todos ellos gracias al uso de sensores remotos; (i) delinear copas de pino piñonero y de pino resinero (en masas pura y mixtas) con datos de baja densidad LiDAR procedentes del PNOA, (ii) identificar especies a partir de dicha delineación y del uso de la información procedentes de imágenes de satélite, gracias a técnicas “machine learning” para su clasificación, y (iii) implementar en los anteriores resultados un modelo de producción de pía, sólo para los pies identificados como *Pinus pinea*.

3. Metodología

3.1. Área de estudio y datos de campo

El área de estudio se ubica en la Meseta Norte de España, sobre masas boscosas de *Pinus pinaster* y *Pinus pinea* donde el aprovechamiento de los productos forestales no maderables como el piñón o la resina han aumentado su valor considerablemente y actualmente se presentan como el principal motor económico de estas zonas despobladas. Concretamente, la zona de estudio se divide en 3 partes: (i) Una primera zona donde encontramos masas puras de pino piñonero localizada en los pinares del sur de la provincia de Valladolid. (ii) Una segunda zona ubicada sobre masas puras de pinaster en resinación en la provincia de Segovia, y (iii) una última zona situada sobre masas mixtas de pino piñonero y pinaster, cada vez más frecuentes en la meseta castellana.

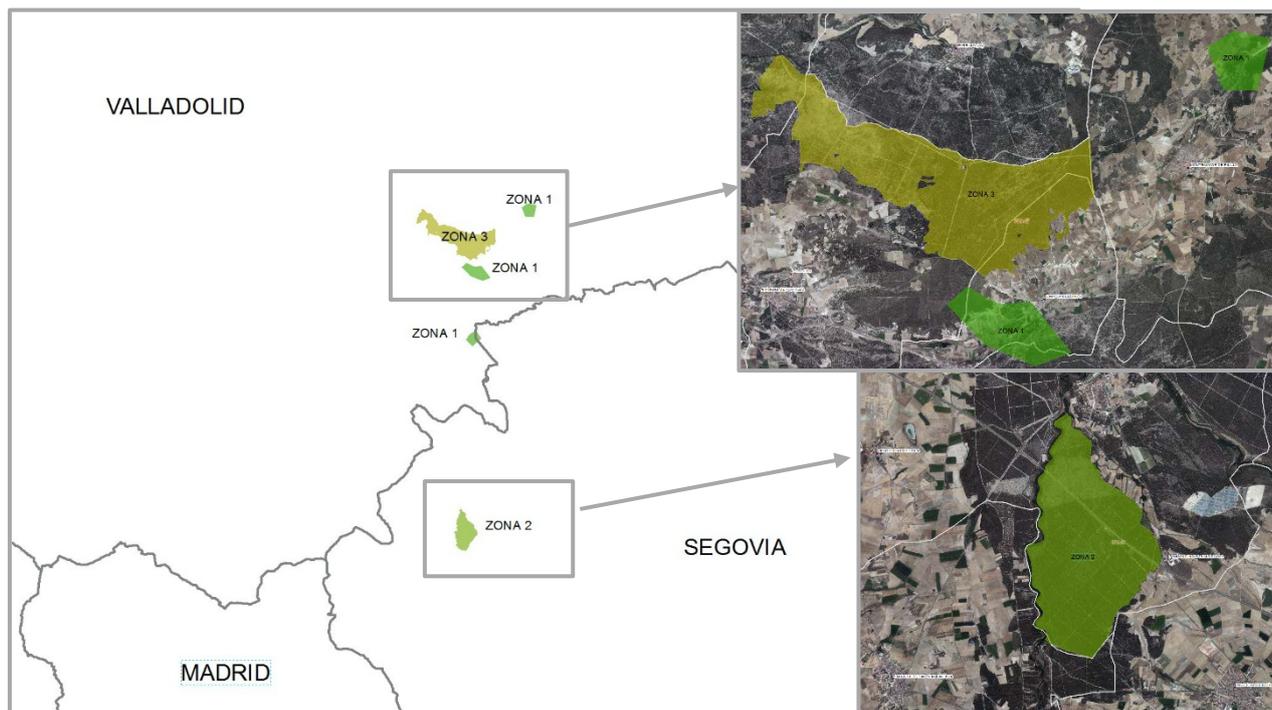


Figura 1. Áreas de estudio. Zona 1: Masa pura de *Pinus pinea*, Zona 2; Masa pura de *Pinus pinaster*, Zona 3: Masa Mixta de *Pinus pinea* y *Pinus pinaster*.

Se tomaron datos de campo sobre parcelas de radio 15 m y en cada una de ellas se tomó la coordenada del centro de la parcela y se recogió la especie de cada uno de los pies dentro del área establecida y se midió su rumbo y distancia para la posterior geolocalización exacta de los mismos sobre el plano. El número total de árboles tomados como muestra de entrenamiento en cada zona se especifica en la siguiente tabla.

Tabla 1. Datos de Campo

Número de pies de la Muestra de entrenamiento				
	Pies menores	P. pinea	P. pinaster	Total
Zona 1	497	0	239	736
Zona 2	195	365	0	560
Zona 3	43	104	138	285
Total	735	469	377	

3.2. Fuentes de datos procedentes de sensores remotos

Para las tres zonas de estudio se emplearon los datos LiDAR del PNOA. El objetivo del proyecto PNOA es cubrir todo el territorio de España mediante nubes de puntos con coordenadas X,Y,Z obtenidas mediante sensores LiDAR aerotransportados, con una densidad de 0,5 puntos/m². La precisión altimétrica obtenida es mejor de 20 cm RMSE Z. Estas coberturas se han realizado mediante colaboración y cofinanciación entre el Ministerio de Fomento (por medio del IGN y el CNIG), el Ministerio de Agricultura, Alimentación y Medio Ambiente (por medio de la Dirección General del Agua, las Confederaciones Hidrográficas y el FEGA) y el Ministerio de Hacienda y Administraciones Públicas (por medio de la Dirección General del Catastro), además de las Comunidades Autónomas. Los datos se distribuyen a través del Centro de Descarga del CNIG en ficheros digitales de 2x2 km de extensión. El formato de descarga es LAZ (formato de compresión de ficheros LAS). Las nubes de puntos han sido capturadas mediante sensores LiDAR y posteriormente clasificadas de forma automática y coloreadas en RGB a partir de las ortofotos del PNOA con tamaño de pixel de 25 o 50cm. En nuestro caso, el sistema geodésico de referencia es ETRS89 y proyección UTM en el huso 30N. Las coordenadas Z son altitudes ortométricas.

Las imágenes de satélite empleadas en este proyecto corresponden a imágenes de muy alta resolución, cedidas por el Plan Nacional de Teledetección (PNT) para este trabajo. El PNT es un proyecto nacional integrado dentro del Plan Nacional de Observación del Territorio (PNOT) que arrancó en el año 2004 con el objetivo de centralizar, a nivel de todas las administraciones, la adquisición de imágenes de satélite para toda la cobertura del país. El PNT ha puesto a disposición de este trabajo tres imágenes Pleiades de los meses junio y julio de 2014 para cubrir la zona de estudio. La misión Pleiades consta de los satélites Pleiades 1A y Pleiades 1B, los cuales nacen con el objetivo de complementar las capacidades de los satélites SPOT, los cuales tienen un campo de visión más amplio, pero de menor resolución espacial, pudiendo así trabajar conjuntamente y obtener una imagen de cualquier lugar del mundo en menos de 24 horas. La constelación Pleiades está diseñada para cubrir las necesidades de los usuarios tanto militar como civil a muy alta resolución, siendo utilizado por ejemplo en la monitorización de volcanes activos. Estas imágenes envuelven una cobertura superficial de 400 km², ofreciendo una resolución espacial de 0,5 metros en la banda pancromática, y 2 metros en las 4 bandas multiespectrales (visible e infrarrojo cercano). El nivel de procesamiento en el que se han proporcionado las imágenes ha sido en el nivel LO correspondiente a datos en bruto.

3.3. Delineación de copas

La metodología para la delineación individual de copas se basa en analizar la nube de puntos directamente en busca de máximos relativos. Una vez se localiza un máximo, se busca el siguiente retorno LiDAR, más bajo que él, y se calcula a qué distancia está de ese máximo, y a partir de unas reglas en función de los datos de campo, se determina si esa distancia implica que ese punto es un nuevo ápice o, sin embargo, si forma parte de la copa en donde se había encontrado ese primer ápice. Así sucesivamente con cada uno de los retornos LiDAR (Valbuena et al., 2016). Una vez realizada la delineación de copas, se calcularon una serie de variables LiDAR que serán empleadas como variables independientes en el modelo de clasificación. Dichas variables se muestran en la siguiente tabla:

Tabla 2. Estadísticas Descriptivas de las variables LiDAR

	Mínimo			Máximo			Media			Desviación estándar		
	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3
SUPERFICIE COPA (m2)	15.80	9.56	7.02	237.84	287.08	122.31	43.37	24.78	39.88	33.71	28.40	25.51
ALTURA_MAX (m)	2.54	4.00	1.03	18.34	18.25	17.70	9.25	7.59	10.80	4.50	4.62	3.62
ALTURA_MIN (m)	2.50	4.00	1.00	14.30	12.68	9.15	5.99	4.99	3.79	2.81	1.48	2.46
ALTURA_COP (m)	0.00	0.00	0.00	15.25	13.11	15.01	3.25	2.60	7.02	2.72	3.75	3.70
COPA_VIVA (%)	0.00	0.00	0.00	85.72	76.16	93.64	29.36	19.95	61.68	18.74	25.26	22.95
DENSIDAD (pies/ha)	19.00	13.00	17.00	278.00	162.00	370.00	101.69	74.66	118.13	48.99	26.96	49.21

El esquema seguido en la delineación de copas es el siguiente:

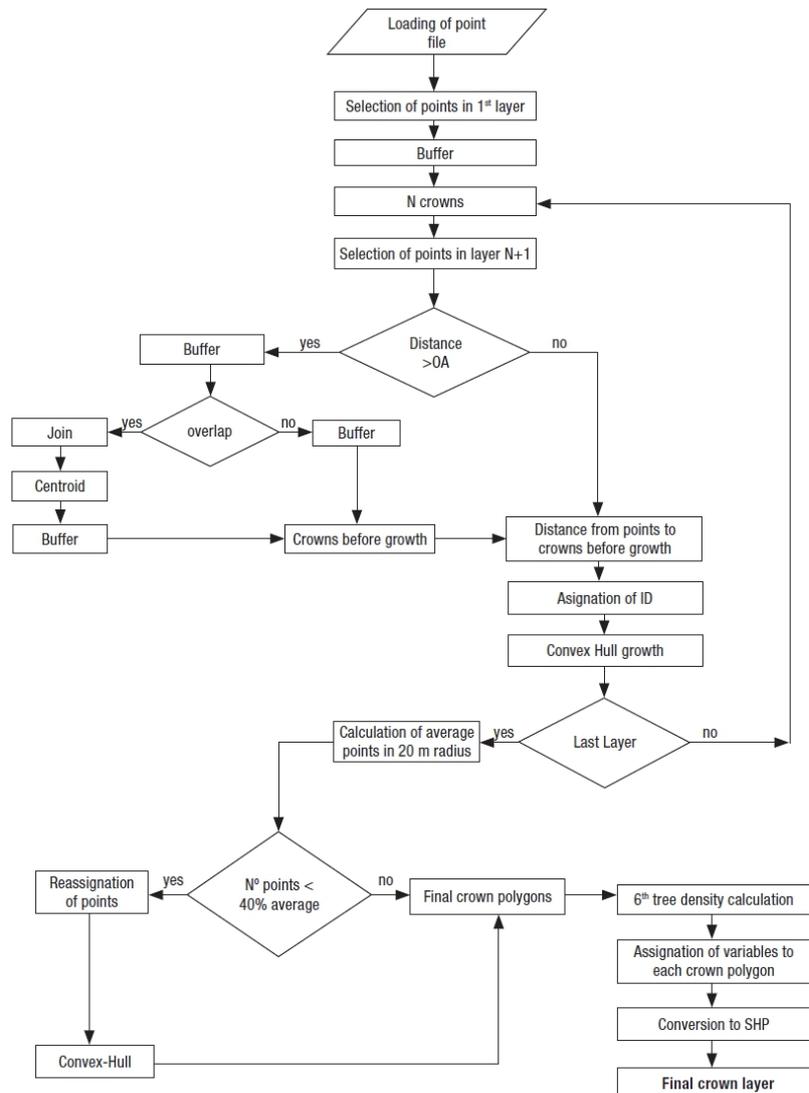


Figura 2. Esquema del algoritmo de delineación de copas

Por otro lado, a partir de los datos tomados en campo se ha realizado un modelo Diámetro-Altura con el objeto de aplicarlo a las copas delineadas y así asignar a cada árbol un diámetro de tronco. Ya con el diámetro y con la altura se puede disponer de las variables dasométricas requeridas para la aplicación del modelo de producción de piña. El resultado ha sido un modelo potencial que ha obtenido un coeficiente de determinación corregido $R^2 = 0.8654$ siendo la formulación del modelo la siguiente (en donde D es el diámetro normal en cm y H es la altura total en metros):

$$D = 0.489730519189084 * H^{1.73532}$$

3.4. Correcciones radiométricas y geométricas sobre Pleiades y cálculo de índices de vegetación

Debido a que las imágenes proporcionadas vienen definidas en el nivel de procesamiento correspondiente a datos brutos, ha sido necesario aplicar correcciones radiométricas y geométricas con el objetivo de llegar al nivel de procesamiento denominado como “sensor perfecto”, cuyo objetivo es corregir las distorsiones propias del sensor y del terreno, pudiendo utilizar las imágenes sin tener en cuenta la complejidad de la geometría del sensor real, y manteniendo al mismo tiempo su precisión original (Cantou et al., 2006; Fernández de Sevilla, 2011). En primer lugar, se realizó la corrección radiométrica utilizando para ello una combinación de parámetros relativos al sensor y a las condiciones atmosféricas locales presentes en el momento de la toma de las imágenes. Después, se realizó la corrección geométrica de la imagen multiespectral y pancromática con el objetivo de geoposicionar con total exactitud ambas imágenes y así realizar la fusión o *phansarpened* entre las mismas. Todas las correcciones fueron realizadas mediante el software ENVI.

El último paso en la preparación de los datos, ha sido el cálculo de los índices de vegetación a partir de las imágenes pleiades. El empleo de índices espectrales, calculados a partir de la reflectividad en diferentes bandas, que nos indiquen la abundancia y estado de la vegetación; se basan en el comportamiento reflectivo peculiar de la vegetación. Permiten identificar la presencia de vegetación verde en la superficie y caracterizar su distribución espacial, así como, su estado fotosintético. La signatura espectral característica de la vegetación sana muestra un fuerte contraste entre la baja reflectividad en el rojo (0.6 y 0.7 nm) y la alta reflectividad en el infrarrojo de longitud de onda más corta (0.7 y 1.1nm). Esta diferencia es tanto mayor cuanto mayor es la densidad de la vegetación y mejor su estado fitosanitario. Con el objetivo de caracterizar cada uno de los árboles de forma independiente se han incluido como variables independientes las variables espectrales que se muestran en la tabla 4, incluyendo los índices espectrales de vegetación calculados según la tabla 3.

Tabla 3. Índices Espectrales de Vegetación calculados

Nombre	Fórmula
Enhanced Vegetation Index	$EVI = 2.5 * (NIR - RED) / (NIR + 6 * RED - 7.5 * BLUE + 1)$
Chlorophyll Index Green	$Cig = NIR / GREEN - 1$
Ratio	$R = NIR / RED$
Soil Ajusted Vegetation Index	$SAVI = (NIR - RED) / (NIR + RED + L) (1 + L)$
Normalized Difference Vegetation Index	$NDVI = (NIR - Red) / (NIR + Red)$

Tabla 4. Estadísticas Descriptivas de las variables Espectrales

	Zona 1: Masa pura P. pinea				Zona 2: Masa pura P. pinaster				Zona 3: Masa mixta			
	Mínimo	Máximo	Media	DS	Mínimo	Máximo	Media	DS	Mínimo	Máximo	Media	DS
BANDA 1	96.10	1509.41	347.16	172.47	187.56	2076.56	472.46	148.33	34.14	1143.69	435.34	179.02
BANDA 2	257.94	1932.24	559.55	205.64	304.03	2552.23	655.48	176.37	83.39	1524.58	661.85	205.63
BANDA 3	119.05	2207.60	507.33	258.58	82.44	2806.95	612.17	242.96	118.71	1736.39	659.12	291.27
BANDA 4	1423.06	3376.52	2261.83	249.43	1511.70	3432.54	2068.44	237.53	1155.61	2920.95	2169.55	269.97
NDVI	0.20	0.90	0.64	0.13	0.10	0.92	0.55	0.12	0.25	1.00	0.55	0.15
EVI	0.47	3.32	1.67	0.37	0.33	4.58	1.76	0.53	0.60	3.69	1.40	0.49
Cig	0.73	6.65	3.40	1.17	0.28	5.59	2.29	0.67	0.88	22.08	2.61	1.52
RATIO	1.51	19.62	5.33	2.27	1.22	24.30	3.82	1.58	1.97	37.76	3.98	6.52
SAVI	0.31	1.35	0.97	0.19	0.15	1.38	0.83	0.17	0.37	1.70	0.83	0.23

3.5. Identificación de especies a partir de métodos “Machine Learning”

La identificación de especies es un conocimiento imprescindible para el forestal, dadas sus diferencias en gestión, uso y aprovechamiento. En nuestro caso, el aprovechamiento principal del

pino piñonero es la piña, mientras que el del pino pinaster es la resina; por tanto, resulta imprescindible cuantificar de forma precisa el número de individuos de cada una de las especies.

Por otra parte, el aprendizaje automático (del inglés, "Machine Learning") es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender. Esos algoritmos son ampliamente empleados en el ámbito de la teledetección para realizar clasificaciones supervisadas (Naidoo et al., 2012; Immitzer et al., 2012; Li et al., 2013).

El siguiente paso de la metodología seguida fue la toma de "verdad terreno" a partir de la cual se va a entrenar el sistema de clasificación. Así, se tomó como "verdad terreno" las copas de los árboles delineados con LiDAR para cada una de las especies, creando así tres categorías diferentes de clasificación; copas de piñonero, copas de pinaster y copas de pies menores de ambas especies.

Una vez preparadas las copas de la muestra de entrenamiento de las diferentes zonas del área de estudio, se llevó a cabo una clasificación de cada una de las copas identificadas utilizando dos técnicas diferentes: (i) Técnica de Árboles de Clasificación o CRT, y (ii) Random Forest.

3.5.1 CRT

La técnica de los árboles de clasificación o CRT es un método de clasificación basado en árboles de decisión que clasifica casos en grupos o pronostica una variable dependiente basada en valores de variables independientes. Son especialmente útiles para grandes bases de datos donde no está clara la relación de las variables explicativas pudiendo crear complejos entramados de relaciones entre variables con el objetivo de discriminar las distintas clases establecidas (Mika et al., 2012). La construcción de los árboles de decisión se ha realizado mediante el software estadístico SPSS.

3.5.2 Random Forest

Dentro de las técnicas de clasificación supervisada, el método random forest (RF) tiende a mejorar la precisión obtenida por los CRT, ya que es una combinación de un número muy elevado de árboles predictores, pero con un componente de aleatoriedad en la construcción de cada uno de los mismos. Esta aleatorización se introduce tanto en la partición del espacio (selección de variables a testar en cada nodo), como en la sub-selección de la muestra de entrenamiento. En definitiva, RF se considera uno de los algoritmos de clasificación más precisos actualmente, que funciona bien con grandes bases de datos, es no paramétrico y permite el uso de gran cantidad de variables predictivas, incluso correlacionadas, sin necesidad de excluir ninguna. Para su construcción se ha empleado el software estadístico R.

3.6. Aplicación de un modelo de producción de piña

Calama et al. (2016) desarrollan un modelo que permite predecir la producción de piña en un árbol en un año determinado, valor que posteriormente puede ser extrapolado a nivel de rodal. El modelo queda formulado como un modelo bietápico, donde en una primera fase se evalúa la probabilidad de que el árbol produzca fruto en ese año determinado, mientras que en una segunda fase se predice la cantidad de piña producida en el árbol (en peso) condicional a que el árbol tenga fruto. El modelo utiliza como variables explicativas las habitualmente medidas en los inventarios dasocráticos: diámetro normal de los árboles, diámetro medio cuadrático, e índice de espesura de Reineke, así como la clasificación del rodal de acuerdo a su localización en las distintas unidades naturales definidas en la provincia de Valladolid. El modelo incluye asimismo variables climáticas, principalmente la precipitación en momentos clave durante el proceso trienal de maduración del fruto.

Para aplicar el modelo a los datos LIDAR ha sido necesario desarrollar previamente un modelo que permita determinar el diámetro normal de los árboles a partir de variables individuales obtenidas con el LIDAR, habiendo identificado como mejor variable explicativa la altura individual. A partir de

este modelo preliminar se ha determinado el diámetro normal de cada pie, así como el diámetro medio cuadrático del rodal y el índice de Reineke. Las variables climáticas entran en el modelo de manera estandarizada, por lo que en el cálculo para un año promedio – como es el objeto del caso de estudio – se considera un valor cero para las mismas.

4. Resultados

4.1. Delineación de copas

La calibración del algoritmo para piñonero se llevó a cabo con las parcelas medidas en la zona 1, mientras que la calibración para pinaster se llevó a cabo con las parcelas de la zona 2. En ambos casos los errores obtenidos fueron menores del 10 por ciento, En la siguiente imagen se muestra un ejemplo de la individualización de copas LiDAR realizada en piñonero en la zona 1, a partir de datos procedentes del PNOA a una resolución de 0.5 ptos/m².



Figura 3. Delineación de copas individuales en el monte "Arenas"

Con el objetivo de estimar la bondad de uso de la individualización de copas, pero esta vez sobre una muestra independiente y además sobre una masa mixta, se aplicó el algoritmo en una muestra independiente de 1219 pies (zona 3). La precisión obtenida en cuanto al número de pies totales medidos en campo y detectados mediante la individualización fue cercana al 80%, siendo el porcentaje de aciertos mayor en los pies mayores (80%) que en los pies menores (75%).

Tabla 5. Matriz de Confusión de Individualización

	Pies Medidos	Pies LiDAR	Porcentaje acierto
Nº pies totales	926	731	78.94%
Nº pies mayores	786	630	80.15%
Nº pies menores	140	105	75.00%

4.2. Identificación de especies

4.2.1 Resultados CRT

En la siguiente figura se muestran los resultados de la matriz de clasificación y la matriz de riesgo del árbol de clasificación obtenido. Como puede observarse, el porcentaje global de acierto es elevado, mayor del 95%, acompañado de unos datos de resustitución y validación cruzada bajos, lo que refleja una alta bondad de ajuste de la muestra.

Tabla 6. Matriz de Riesgo CRT

Estimación	Error estándar
0.032	0.005

Tabla 7. Matriz de Confusión CRT

Observado	Pies menores	P. pinea	P. pinaster	Porcentaje correcto
Pies menores	690	0	0	100.00%
P. pinea	1	343	21	94.00%
P. pinaster	0	19	217	91.90%
Porcentaje global	53.50%	28.00%	18.40%	96.80%

4.2.2 Resultados Random forest

Una vez construidos los modelos de clasificación, el paquete 'RandomForest' proporciona una validación cruzada, junto a un error global que denomina out-of-bag error (OOB-error). Para cada uno de los árboles utilizados para construir el modelo de clasificación, 'RandomForest' utiliza de forma automática 2/3 de la muestra seleccionada aleatoriamente como muestra de entrenamiento, y el 1/3 restante como muestra independiente de validación. El conjunto de esas validaciones individuales es lo que se utiliza para generar esa matriz de confusión global que, siempre que la muestra de entrenamiento sea realmente representativa de la población, dará como resultado unas precisiones muy próximas a las que tendrá el modelo una vez aplicado a la población.

Tabla 8. Matriz de Confusión Random Forest

Observado	Pies menores	P. pinea	P. pinaster	Error
Pies menores	690	0	0	0.000
P. pinea	1	359	5	0.016
P. pinaster	1	6	229	0.030
OOB estimate error				1.01%

Por otra parte, las variables seleccionadas por los algoritmos de clasificación, ordenadas en función de la importancia de las mismas, se muestran en la siguiente figura. Destaca la Banda 1 de Pleiades y la Altura de los árboles (Altura_Max).

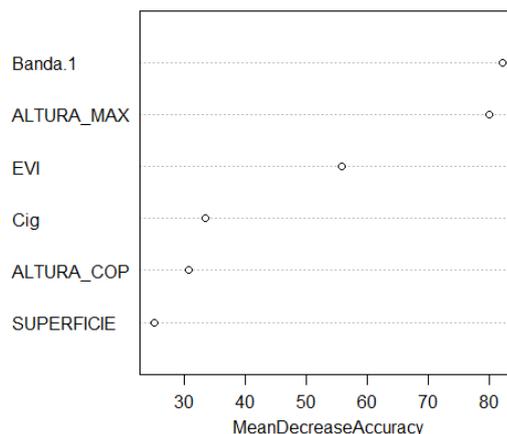


Figura 4. Variables escogidas por los algoritmos de clasificación de Random Forest.

4.3. Aplicación del modelo de producción de piña

Dado que el modelo de clasificación obtenido mediante los algoritmos de clasificación de RF ofrece mejores porcentajes de acierto que los obtenidos mediante el CRT, estos han sido aplicados sobre la masa mixta. Solamente sobre los árboles clasificados como piñonero, se ha aplicado el modelo de producción de piña para cada árbol individualmente. En la siguiente tabla se muestran los resultados de producción de piña en Kg/ha clasificados en función de la densidad de la masa y del porcentaje de mezcla presente en las parcelas. Se ha considerado densidad baja por debajo de 250 pies/ y alta por encima de 350 pies/ha, mientras que el porcentaje de mezcla ha sido clasificado con alto por encima del 75% de presencia de piñonero frente a pinaster. La producción media estimada ha sido de 185 kg/ha, que corresponde a un valor medio de 2,469 Kg/árbol.

Tabla 9. Producción media de piña en Kg/ha/año

Densidad/Porcentaje de mezcla	Menor de 75%	Mayor de 75%
Baja (<250p/ha)	166.95	128.55
Media (250<x<350p/ha)	78.69	106.66
Alta (>350p/ha)	316.31	176.80

En la siguiente imagen se representa una masa mixta sobre la que se ha aplicado la delineación de copas y posteriormente el modelo de clasificación entre las especies objeto de estudio: *Pinus pinea* (color rojo), *Pinus pinaster* (color amarillo) y regeneración (color azul).

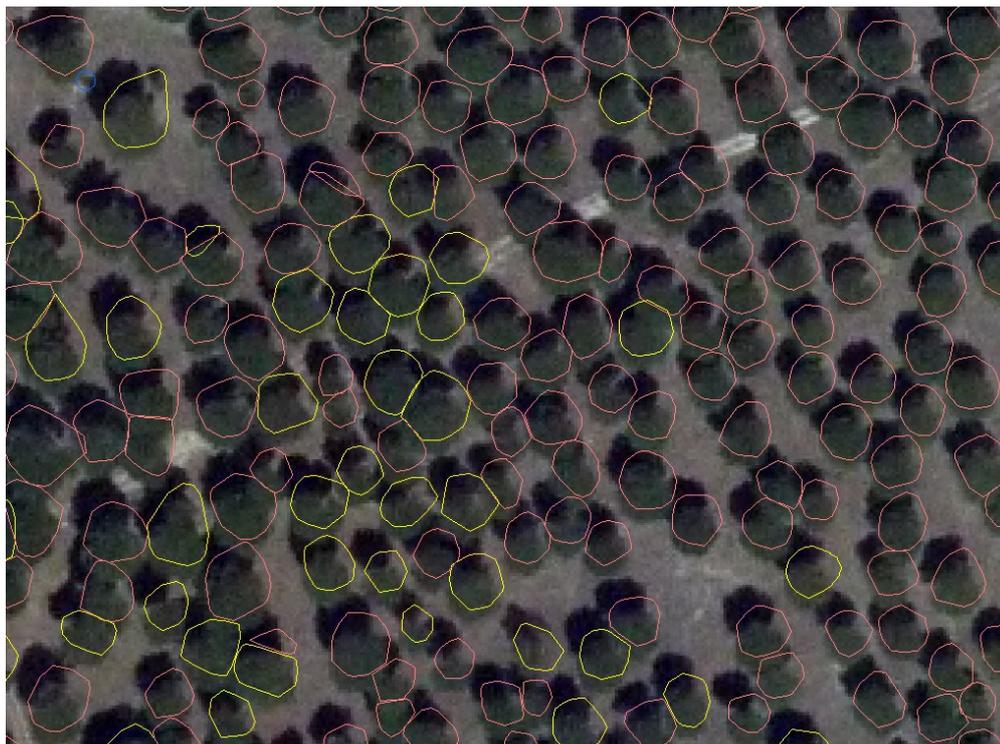


Figura 5. Modelo de clasificación aplicado sobre masa mixta.

5. Discusión

No hemos encontrado referencias de trabajos realizados de clasificación de especies en masas naturales basadas en individualización de copas mediante datos LiDAR de baja resolución combinado con imágenes satelitales de alta resolución. Normalmente, los métodos de árbol individual son utilizados para la estimación de variables dendrométricas, principalmente diámetros, a partir de variables propias de los datos LiDAR como es la altura (Valbuena et al., 2016). Por otro lado, estos trabajos parecen estar siempre localizados sobre masas puras. Sin embargo, existen algunas referencias similares en cultivos agroforestales discriminando olivos, algarrobos y almendros siendo sus resultados de discriminación en torno al 92% combinando datos LiDAR de baja resolución e imágenes DMC de alta resolución (Viñas et al., 2009). Por tanto, se ha obtenido un resultado de clasificación satisfactorio entre las diferentes especies objeto de estudio, localizadas en masas naturales combinando datos LiDAR e imágenes de satélite de muy alta resolución respectivamente.

Por otro lado, el modelo de producción de piña predice una producción media por hectárea de 185 kg (error estándar $\pm 20,4$ kg), y una producción media por árbol de 2,469 kg ($\pm 0,203$ kg). Ambos valores son muy cercanos a la producción media por hectárea obtenida para la provincia de Valladolid, y algo inferior a los observados para la Unidad Natural "Páramo calizo", donde se encuentra la zona de estudio, y en la que la producción media por árbol es de 3,367 kg (Calama et al. 2012). En cualquier caso, los valores obtenidos resultan lógicos para una masa con el porcentaje de mezcla observado.

6. Conclusiones

El LiDAR PNOA es una fuente de información válida para poder delinear copas en masas abiertas o con cobertura incompleta. El algoritmo usado (Valbuena et al., 2016) ya había mostrado su bondad de uso en diferentes especies y con diferentes densidades de puntos (Valbuena et al., 2016),

pero nunca sobre masas mixtas y usando datos PNOA. Este algoritmo es útil en este tipo de masas abiertas.

La identificación de especies a escala árbol es un producto muy innovador y útil para este tipo de masas mixtas en donde el aprovechamiento principal es un producto no maderable. En este trabajo se ha realizado de forma satisfactoria, combinando la información LiDAR procedente de la delimitación de copas y usando imágenes de muy alta resolución.

La adaptación e implementación de un modelo de producción de piña, diseñado para su uso con datos propios de los inventarios, resulta muy útil cuando se han delimitado copas y se han discriminado especies con la fusión de datos LiDAR e imágenes de satélite de muy alta resolución.

7. Agradecimientos

Este trabajo ha sido posible gracias a (i) Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+i del Ministerio de Economía y Competitividad por la beca de Doctorado Industrial (DI-14-06953) asignada a A.B.C., (ii) al Plan Nacional de Teledetección (PNT), por la aportación de las imágenes utilizadas en el estudio, (iii) al Plan Nacional de Ortófotografía Aérea (PNOA) por la cesión de los datos LiDAR, y (iv) al Proyecto RTA-2013-00011-C2.1, “Vulnerabilidad y Adaptación de los pinares de *Pinus pinea* al Cambio Climático”

8. Bibliografía

BRANDTBERG, T.; 2007. Classifying individual tree species under leaf-off and leaf-on conditions using airborne LiDAR. ISPRS J. Photogramm. Remote Sens., 61, 325–340.

CALAMA, R.; MUTKE, S.; PARDOS, M.; MORALES, L.; GORDO, F.J.; MONTERO, G.; FINAT, L.; 2012. La producción de piña y piñón de *Pinus pinea* L. en la Meseta Norte. En: La regeneración natural de los pinares en los arenales de la meseta castellana (eds. FJ Gordo, R Calama, M Pardos, F Bravo, G Montero) IUIGFS-Uva-INIA: 81-98.

CALAMA, R.; GORDO, J.; MADRIGAL, G.; MUTKE, S.; CONDE, M.; MONTERO, G.; PARDOS, M.; 2016. Enhanced tools for pre-dicting annual stone pine (*Pinus pinea* L.) cone production at tree and forest scale in Inner Spain. Forest Systems, 25(3), e079.

CANTOU, J. P.; MAILLET, G.; FLAMANC, D.; BUISSART, H.; 2006. Preparing the use of Pleiades images for mapping purposes: preliminary assessments at IGN-France. International Archives of Photogrammetry and Remote Sensing, 36.

DAVIES, O.; HAUFE, J.; POMMERENING, A.; 2008. Silvicultural Principles of Continuous Cover Forestry a Guide to best Practice. Forestry Commission Wales. 111 p.

FERNANDEZ DE SEVILLA, F.; 2011 La constelación de satélites Pleiades-HR1 y 2: De la teledetección a la fotogrametría. XIV Congreso de la Asociación Española de Teledetección, Mieres, Oviedo.

HEURICH, M.; PERSSON, A.; HOLMGREN, J.; KENNEL, E.; 2004. Detection and measuring individual trees with laser scanning in mixed mountain forest of central Europe using an algorithm developed for Swedish boreal forest conditions. In: Proceedings of the international Conference. Laser-Scanners for Forest and Landscape Assessment - Instruments, Processing Methods and Applications. Freiburg im Breisgau. Germany.

HYPPA, J.; KELLE, O.; LEHIKONEN, M.; INKINEN, M.; 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. IEEE T Geosci. Remote 2001, 39(5), 969-975.

IMMITZER, M.; ATZBERGER, C.; KOUKAL, T.; 2012. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. Remote Sensing, 4(9), 2661-2693.

KOENING, K.; HOFLE, B.; 2016. Full-Waveform Airborne Laser Scanning in Vegetation Studies—A Review of Point Cloud and Waveform Features for Tree Species Classification. Forests, 7(9), 198.

LATIFI, H.; NOTHDURFT, A.; KOCH, B.; 2010. Non-parametric prediction and mapping of standing wood volume and biomass in a temperate forest: application of multiple optical/LiDAR-derived predictors. *Forestry*, 83(4), 395-407.

LI, M.; IM, J.; BEIER, C.; 2013. Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntington Wildlife Forest. *GIScience & Remote Sensing*, 50(4), 361-384.

MAGNUSSEN, S.; NAESSET, E.; GOBAKKEN, T.; 2010. Reliability of LiDAR derived predictors of forest inventory attributes: A case study with Norway spruce. *Remote Sens. Environ.* 114(4), 700-712.

MASON, B.; KERR, G.; SIMPSON, J.; 1999. What is continuous cover forestry? *Forestry Commission Information Note 29*. Forestry Commission, Edinburgh. O'Hara K., Gersonde R. 2004. Stocking Control Concepts in Uneven-Aged Silviculture. *Forestry* 7(2), 131-143.

NAIDOO, L.; CHO, M. A.; MATHIEU, R.; ASNER, G.; 2012. Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 69, 167-179.

NILSSON, M.; 1996 Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sens Environ.* 56, 1-7.

PECK, M.; MARISCAL, A.; PADBURY, M.; CANE, T.; KNIVETON, D.; CHINCHERO, M. A.; 2012. Identifying tropical Ecuadorian Andean trees from inter-crown pixel distributions in hyperspatial aerial imagery. *Applied Vegetation Science*, 15(4), 548-559.

POPESCU, S. C.; WYNNE, R. H.; NELSON, R. F.; 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Can. J. Remote Sens.* 29(5), 564-577.

RODRIGUEZ, F.; LIZARRALDE, I.; 2015. Inventario forestal con tecnología LiDAR, ¡Más barato y más preciso!. *Revista Forestalis*. 24; 33-34.

VALBUENA, M.; SANTAMARIA, J.; SANZ, F.; 2016. Estimation of diameter and height of individual trees for *Pinus sylvestris* L. based on the individualising of crowns using airborne LiDAR and the National Forest Inventory data. *Forest Systems*, Volume 25, Issue 1, e046. <http://dx.doi.org/10.5424/fs/2016251-05790>.

YAO, W.; KRZYSZEK, P.; HEURICH, M.; 2012. Tree species classification and estimation of stem volume and DBH based on single tree extraction by exploiting airborne full-waveform LiDAR data. *Remote Sens Environ* 123:368-380.

ZHANG, Z.; KAZAKOVA, A.; MOSKAL, L. M.; STYERS, D. M.; 2016. Object-Based Tree Species Classification in Urban Ecosystems Using LiDAR and Hyperspectral Data. *Forests*, 2016, 7(6), 122.